Bioprocess Intensification of Cell-free Enzymatic Cascades for the Synthesis of CDP-glycerol, CMP-Neu5AC, GDP-fucose and LNFP III

Dissertation

zur Erlangung des akademischen Grades

Doktor Ingenieur (Dr.-Ing)

von M.Sc. Edgar Alberto Alcalá Orozco

geb. am 15. August 1994 in Mexiko Stadt, Mexiko.

genehmigt durch die Fakultät für Verfahrens- und Systemtechnik der Otto-von-Guericke-Universität Magdeburg

Promotionskommission:

Prof. Dr. Jan von Langermann (Vorsitz)

Prof. Dr.-Ing. Udo Reichl (Gutachter)

Dr. Steffen Klamt (Gutachter)

Prof. Dr. Sabine Flitsch (Gutachter)

Promotionskolloquium am 5. November 2025

Abstract

Sugars represent one of the most prevalent and significant classes of chemicals on our planet. They can be found as conjugated glycans within various proteins and other macromolecules essential to biology, which has facilitated technologies like the development of glycoconjugate vaccines targeting bacterial pathogens. They also serve as essential sources of energy and nutrition in various forms within our food, spanning from glucose to Human Milk Oligosaccharides (HMOs). Notably, in infant nutrition, the incorporation of HMOs, alongside human-like protein and fat sources, represents a pivotal advancement in developing healthier alternatives to cow milk-based infant formulas. HMOs are synthesised in nature from nucleotide sugars, including uridine diphosphate glucose (UDP-Glc), uridine diphosphate galactose (UDP-gal), guanosine diphosphate fucose (GDP-fucose), and cytidine monophosphate *N*-acetyl neuraminic acid (CMP-Neu5Ac), making it important to develop processes to produce them.

Since the discovery and isolation of Leloir-glycosyltransferases in the 1950s, carbohydrate scientists worldwide have been working on using these enzymes for synthesising many oligosaccharides under mild conditions. These enzymes transfer a sugar moiety from a donor (nucleotide sugar) to an acceptor molecule with high specificity. Enzymatic methods can be performed at room temperature and have the potential to achieve perfect stereoselectivity. In contrast, the chemical synthesis of oligosaccharides is challenging, requires several reaction steps, and often results in low stereospecificity. However, two key challenges remain for the widespread application of Leloir-glycosyltransferases: the availability of enzymes and the accessibility of nucleotide sugars. This work addresses the latter and takes steps toward the economically viable sourcing of nucleotide sugars. To this end, nucleotide sugars were produced in One-Pot Multi-Enzyme (OPME) cell-free enzymatic cascades, which were improved in a series of process intensification steps, such as the co-expression of enzymes and the optimisation of reaction parameters.

This work presents the synthesis of cytidine diphosphate glycerol (CDP-glycerol), CMP-Neu5Ac, and GDP-fucose, with a final chapter dedicated to a use-case scenario that extends the GDP-fucose enzymatic cascade to produce the HMO Lacto-N-fucopentaose III (LNFP III). The first target molecule, CDP-glycerol, is the nucleotide-activated form of glycerol that has recently been utilised in the development of novel glycoconjugate vaccines, particularly those that include synthetic teichoic acid fragments. A new enzymatic cascade capable of synthesising CDP-glycerol was developed. After optimisation using a Design of Experiments (DoE) approach, it achieved a substrate conversion yield of 89 % and a product titer of 31.2 mM at a 200 µl scale. The second target molecule, CMP-Neu5Ac, is an acidic nucleotide sugar that can act as a substrate in the enzymatic synthesis of HMOs and the in vitro glycoengineering of therapeutic proteins. After completing the process intensification steps for enzyme co-expression and DoE optimisation, the reaction could be scaled to produce 6.2 g of product with a 98 % substrate conversion yield. The third target molecule was GDP-fucose, one of the costliest nucleotide sugars available and among the most interesting to produce at large scales. The process intensification steps of kinetic modelling, enzyme co-expression, DoE optimisation, and scaling to gram-scale production were completed. Using the optimised method, a total of 1.76 g could be produced with a 97.8 % substrate conversion yield in a one-pot batch reaction. Finally, the GDP-fucose producing enzymatic cascade was extended to an enzymatic cascade for the synthesis of LNFP III. The HMO could be synthesised to a product titer of 6.8

mM in a 200 µL scale, showcasing the potential of nucleotide sugar synthesis cascades to be used modularly in producing HMOs.

This study closes the gap between academic research and commercial application by utilising a variety of process intensification and reaction engineering methodologies aimed at improving established, yet underdeveloped, enzymatic cascades for the synthesis of nucleotide sugars.

Kurzfassung

Zucker sind eine der am weitesten verbreiteten und wichtigsten Klassen von Chemikalien auf unserem Planeten. Sie sind als konjugierte Glykane in verschiedenen Proteinen und anderen Makromolekülen zu finden, die für die Biologie wichtig sind, was die Entwicklung von Glykokonjugat-Impfstoffen gegen bakterielle Krankheitserreger erleichtert hat. Sie dienen auch als wesentliche Energie- und Nährstoffquellen in verschiedenen Formen in unserer Ernährung, von Glukose bis hin zu humanen Milcholigosacchariden (HMOs). Insbesondere in der Säuglingsernährung stellen HMOs neben menschenähnlichen Protein- und Fettquellen einen entscheidenden Fortschritt bei der Entwicklung gesünderer Alternativen zu Säuglingsnahrung auf Kuhmilchbasis dar. HMOs werden in der Natur aus Nukleotidzuckern wie Uridindiphosphat-Glukose (UDP-Glc), Uridindiphosphat-Galaktose (UDP-Gal), Guanosindiphosphat-Fukose (GDP-Fukose) und Cytidinmonophosphat-N-Acetyl-Neuraminsäure (CMP-Neu5Ac) synthetisiert, weshalb es wichtig ist, Verfahren zu ihrer Herstellung zu entwickeln.

Seit der Entdeckung und Isolierung der Leloir-Glykosyltransferasen in den 1950er Jahren haben Forscher weltweit daran gearbeitet, diese Enzyme für die Synthese zahlreicher Oligosaccharide unter milden Bedingungen einzusetzen. Diese Enzyme übertragen mit hoher Spezifität ein Zuckermolekül von einem Donor (Nukleotidzucker) auf ein Akzeptormolekül. Enzymatische Methoden können bei Raumtemperatur durchgeführt werden und können eine perfekte Stereoselektivität erreichen. Im Gegensatz dazu ist die chemische Synthese von Oligosacchariden schwierig, erfordert mehrere Reaktionsschritte und führt häufig zu einer geringen Stereospezifität. Für die breite Anwendung von Leloir-Glykosyltransferasen gibt es jedoch noch zwei zentrale Herausforderungen: die Verfügbarkeit der Enzyme und die Verfügbarkeit der Nukleotidzucker. Die vorliegende Arbeit befasst sich mit letzterem und unternimmt Schritte in Richtung einer wirtschaftlich tragfähigen Beschaffung von Nukleotidzuckern. Zu diesem Zweck wurden Nukleotidzucker in zellfreien One-Pot-Multi-Enzyme (OPME)-Enzymkaskaden hergestellt, die durch verschiedene Prozessintensivierungsschritten, wie die Koexpression von Enzymen und die Optimierung der Reaktionsparameter, verbessert wurden.

In dieser Arbeit wird die Synthese von Cytidindiphosphatglyzerin (CDP-Glyzerin), CMP-Neu5Ac und GDP-fucose vorgestellt, wobei das letzte Kapitel einem Anwendungsszenario gewidmet ist, das die GDP-fucose-Enzymkaskade zur Herstellung der HMO Lakto-N-Fucopentaose III (LNFP III) erweitert. Das erste Zielmolekül, CDP-Glyzerin, ist die nukleotidaktivierte Form von Glyzerin, die in jüngster Zeit bei der Entwicklung neuartiger Glykokonjugat-Impfstoffe, insbesondere solcher, die synthetische Teichonsäurefragmente enthalten, eingesetzt wurde. Für die Synthese von CDP-Glyzerin wurde eine neue enzymatische Kaskade zur Synthese von CDP-Glyzerin. Nach der Optimierung durch einen DoE-Ansatz (Design of Experiments) wurde eine Substratumwandlungsausbeute von 89 % und ein Produkttiter von 31,2 mM im 200-µl-Maßstab erreicht. Das zweite Zielmolekül, CMP-Neu5Ac, ist ein saurer Nukleotidzucker, der als Substrat für die enzymatische Synthese von HMOs und dem In-vitro-Glycoengineering von therapeutischen Proteinen dienen kann. Nach Abschluss der Prozessintensivierungsschritte zur Enzymkoexpression und die DoE-Optimierung konnte die Reaktion so skaliert werden, dass 6,2 g Produkt mit einer Substratumwandlungsausbeute von 98 % erhalten wurden. Das dritte Zielmolekül war GDP-fucose, einer der teuersten verfügbaren Nukleotidzucker und einer der interessantesten für die Herstellung in großem Maßstab. Die Prozessintensivierungsschritte der kinetischen Modellierung,

der Koexpression des Enzyms, der DoE-Optimierung und der Skalierung auf die Produktion im Gramm-Maßstab wurden abgeschlossen. Mit der optimierten Methode konnten insgesamt 1,76 g mit einer Substratumwandlungsausbeute von 97,8 % in einer One-Pot-Batch-Reaktion hergestellt werden. Schließlich wurde die GDP-fucose produzierende Enzymkaskade um die Synthese von LNFP III erweitert. Das HMO konnte in einem Maßstab von 200 µL bis zu einem Produkttiter von 6,8 mM synthetisiert werden, was das Potenzial von Nukleotidzuckersynthesekaskaden für den modularen Einsatz bei der Herstellung von HMOs verdeutlicht.

Diese Arbeit schließt die Lücke zwischen akademischer Forschung und kommerzieller Anwendung, indem sie eine Reihe von Methoden zur Prozessintensivierung und Reaktionstechnik eingesetzt werden, um etablierte, aber unterentwickelte enzymatische Kaskaden für die Synthese von Nukleotidzuckern zu verbessern.

*Translation made with DeepL

Table of Contents

Abbreviations

2'-FL	2'-fucosyllactose
3/4-FT	α1-3/4-fucosyltransferase
3'-FL	3-fucosyllactose
3'-SL	3'-sialyllactose
6'-SL	6-sialyllactose
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
AATP	Adenosine tetraposphate
BSA	Bovine serum albumin
CCO	Central composite orthogonal
CDP	Cytidine diphosphate
CDP-glycerol	Cytidine diphosphate glycerol
CMP	Cytidine monophosphate
CMP-Neu5Ac	Cytidine monophosphate *N*-acetylneuraminic acid
CTP	Cytidine triphosphate
Da	Dalton
DFL	Difucosyl lactose
DoE	Design of Experiments
DSNLT	Disialyl lacto-*N*-Tetraose
E. coli	*Escherichia coli*
FDA	Food and Drug Administration
Fuc-1P	Fucose1-phosphate
GalNAc-1P	*N*-acetyl galactosamine-1-phosphate
GDP	Guanosine diphosphate
GDP-fucose	Guanosine diphosphate L-fucose
GDP-Man	Guanosine diphosphate mannose
Glc-1P	Glucose-1-phosphate
Gly-3P	Glycerol-3-phosphate
Glc-6P	Glucose-6-phosphate
GlcNAc-1P	*N*-acetyl glucosamine-1-phosphate
GMO	Genetically modified organism
GMP	Guanosine monophosphate
GRAS	Generally regarded as safe
GTP	Guanosine triphosphate
His-tag	Histidine tag
HMOs	Human milk oligosaccharides
HPAEC	High-performance anion-exchange chromatography
HPLC	High-performance liquid chromatography
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KEGG	Kyoto Encyclopedia of Genes and Genomes
Lac	Lactose

LacNAc	*N*-acetyl lactosamine
LB	Lysogeny broth
LIFT	Laser-induced fragmentation and transfer
LNFP III	Lacto-*N*-fucopentaose III
LNFP V	Fucosyl lacto-*N*-neotetraose V
LNnT	Lacto-*N*-neotetraose
LNT	Lacto-*N*-tetraose
LNnDFH II	Lacto-*N*-neodifucohexaose II
LSTc	Sialylacto-*N*-tetraose c
LTA	Lipoteichoic acids
m/z	Mass-to-charge ratio
mAb	Monoclonal antibody
MALDI-TOF	Matrix-assisted laser desorption/ionization time of flight
ManNAc	*N*-acetylmannosamine
mAU	Milli absorbance unit
MCS	Multiple cloning site
MOPS	3-(*N*-morpholino)propane sulfonic acid
MPI	Max Planck Institute
MS	Mass spectrometry
MLR	Multiple linear regression
MTP	Multi-titer plate
MWCO	Molecular weight cut-off
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
nC	Nanocoulombs
NMR	Nuclear magnetic resonance
OPME	One-pot multi-enzyme
PAD	Pulsed amperometric detector
Pi	Monophosphate
PolyP	Polyphosphate
PPi	Pyrophosphate
PG	Peptidoglycan
PLS	Partial least squares
pET-Duet	Plasmid T7 expression system, origin of replication ColE1, dual MCS
pRSF-Duet	Plasmid with RSF1030 origin of replication, dual MCS
pACYC-Duet	Plasmid with p15A origin of replication, dual MCS
pCDF-Duet	Plasmid with CloDF13 origin of replication, dual MCS
pColA-Duet	Plasmid with Colicin A origin of replication, dual MCS
Q^2	Internal measure of consistency
R^2	Measure of model fit to the original data
SDS-PAGE	Sodium dodecylsulphate–polyacrylamide gel electrophoresis
SNFG	Standard nomenclature for glycans
SRES	Stochastic Ranking Evolution Strategy
SOC	Super optimal broth with catabolite repression
TB	Terrific broth

TON	Turn over number
TLC	Thin layer chromatography
UV	Ultraviolet
WTA	Wall teichoic acid

Full Name	Abbreviation	Standard nomenclature for glycans (SNFG) symbol
Fucose	Fuc	
N-acetylneuraminic acid	Neu5Ac	
Galactose	Gal	
Glucose	Glc	
Mannose	Man	
Xylose	Xyl	
N-acetylgalactosamine	GalNac	
N-acetylglucosamine	GlcNac	
Glucoronic acid	GlcA	

Nomenclature

Km	Michaelis-Menten Constant	[mmol/L]
Yp/s	Substrate Conversion Yield	[$g_{product}/g_{substrate}$]
Kcat	Enzyme turnover number	[s^{-1}]
Vmax	Maximal reaction rate velocity	[mol/min]
K_i	Inhibition constant	[mmol/L]
DoE Response[1]	HPLC-UV average peak area	[mAU*min]
DoE Response[2]	HPLC-UV calculated molarity	[mM]
DoE Response[3]	Substrate conversion yield	[$g_{product}/g_{substrate}$]

1 Introduction

Glycans are molecules characterized by glycosidic bonds, mainly composed of sugars such as polysaccharides and other carbohydrates. These glycans hold immense biological significance as they participate in processes including cell-to-cell signaling, tissue structure maintenance, and energy storage, among others[1]. An enzyme class known as Leloir Glycosyltransferases is of particular interest for glycan biosynthesis. These enzymes can form carbon-carbon bonds (glycosidic linkages) by transferring a sugar moiety from a donor nucleotide sugar molecule to an acceptor[2], which is often a glycoprotein or an oligosaccharide. Naturally, Leloir glycosyltransferases have the potential to be used in an industrial setting for the production of speciality oligosaccharides. These enzymes can potentially change the way we produce all kinds of glycans, such as Human Milk Oligosaccharides (HMO), cell surface antigens, and even therapeutic oligosaccharides[3], because enzymatic biocatalysis offers a variety of advantages over chemical or fermentative methods. These advantages include synthesizing under mild reaction conditions and benefiting from established biocatalyst production methods, made possible by recombinant DNA technology. However, there is a barrier to the widespread use of glycosyltransferases, namely the limited availability of their donor substrates: nucleotide sugars. As of now, no supplier can provide reasonably priced nucleotide sugars in multi-gram quantities. Most vendors only offer small amounts for research purposes or as analytical standards. The largest amount of GDP-fucose that one can buy from Sigma Aldrich® is 5 mg, and it costs about 1,360 € (summer 2024); and Biosynth®, a low-cost alternative for carbohydrates, offers a maximum of 10 mg for 748 €. In contrast, the global milk formula production in 2017 was estimated at 2.7 billion tons (reported by the market research firm Girafood in 2018). Assuming that adding 5 g/L of HMO content is desired in modern baby formula compositions[4] (roughly 3% of the solid baby formula powder), it would be required to source enough nucleotide sugars to sustain an HMO production of roughly 80 million tons each year. It has become evident that a reliable and cost-effective method for producing nucleotide sugars is necessary to fully exploit the potential of glycosyltransferases. To this end, a synthesis approach using cell-free enzymatic cascades was developed. This method can already produce nucleotide sugars in gram quantities and beyond. However, there is a significant distance between lab-scale synthesis and commercial-scale production. This dissertation focuses on bridging that gap by undertaking the initial process intensification steps necessary for producing nucleotide sugars on a gram scale.

There are four major ways to produce nucleotide sugars and oligosaccharides: organic chemistry methods[5], fermentation[6], whole-cell catalysis[7] and cell-free biocatalysis[8]. Furthermore, each of these technologies can be applied with a variety of design choices that make it challenging to assess which one is the right tool for the job. For example, fermentation can be performed not only in batch mode[9], but also as a fed-batch[10] or in continuous mode[11]; and cell-free biocatalysis can be done using purified enzymes[12] or with cell lysates[13]. With this in mind, the choice of technology is arguably the most critical step towards achieving an efficient synthesis and, later, its production in large scale. It is natural to choose what is familiar, but that may not always be the best design choice for a particular application. After all, give a kid a hammer and everything will look like a nail, teach

fermentation to an engineer and everything will look like bioethanol. History has shown that wildly different technologies and strategies are used to produce the molecules that we, as humanity, know and love. From the German chemists who gave us the indigo dye at the beginning of the 20th century [14] to the production of insulin in bacteria using human recombinant DNA technology[15], the development of new technologies opens the possibility to manufacture not only new molecules but also old ones in a more efficient and environmentally friendly manner. As resourceful as it might be, the now well-established organic chemist has found its limitations and challenges, mainly when dealing with carbon coupling reactions and synthesising enantiomerically pure chiral compounds[16, 17]. In this context, biocatalysis has started gaining traction in the scientific and industrial communities. This thesis revolves around carbohydrate chemistry, particularly the synthesis of Human Milk Oligosaccharides (HMOs) and nucleotide sugars, an application in which biocatalysis happens to excel. The reason is that synthesising oligosaccharides requires the formation of carbon-carbon bonds and enantiomerically pure reactions, as all sugars are chiral compounds[18], which is a challenging task with traditional organic chemistry methods. To further justify the choice of catalysts, one can use Jacobsen´s five criteria, described in 1994 and later commented on by Bommarius in 2004[19]. 1. enantioselectivity of the product, 2. amount of product obtained per amount of catalyst used, 3. availability and costs of the catalyst, 4. substrate specificity, and 5. comparison of the method with alternative strategies. These five criteria are a good starting point for evaluating the choice of catalyst and for having an early overview of the strengths and possible downsides of developing a production process based on the catalyst of choice. From pondering these criteria, it becomes clear that (1) the synthesis of HMOs and nucleotide sugars is already a natural reaction that has evolved to produce a particular enantiomer; (2) enzymes have the potential to be reused or immobilised to dramatically reduce the amount of enzyme needed; (3) the cost of catalyst is very affordable with modern recombinant DNA technology, especially when using *E. coli* as expression platform; (4) glycosyltransferases are specific to certain substrates and linkages, but substrate promiscuity or change in selectivity can be introduced with enzyme engineering[20]; and, (5) enzymatic methods seem to outperform chemical and in some cases whole cell methods in economical and efficiency metrics for the synthesis of nucleotide sugars and complex HMOs.

HMOs can be synthesised using glycosyltransferases that are found in many kinds of organisms, from bacteria to humans[21], and nucleotide sugars are produced in nature by two metabolic pathways, known as de novo and salvage pathways[22]. However, it has taken decades for the technology to mature and it is recently that efficient methods emerged. As an example, the enzymatic synthesis of the nucleotide sugar GDP-fucose has been around since the 70s[23-25]. Still, it was heavily limited by the unavailability of enzymes, which had to be extracted directly from the pig (in one case, at least) or from microorganisms, with very low yields. Furthermore, the latest publication for chemically synthesising GDP-fucose dates back to 1992[26], and later, after the first enzymatic methods using recombinant DNA technology were published, every single research effort towards the synthesis of this nucleotide sugar would utilise enzymes in one way or another[10, 27-33]. The rapid spread of enzymatic biocatalysis can be attributed to certain characteristics, such as each enzyme being specifically tailored to their substrate[34], which is an advantage when aiming to avoid side products. Conversely, the total product yield

per unit of catalyst varies significantly among different enzymes, as it relies on the enzyme's activity and stability. Thus, successful development hinges on selecting the appropriate enzyme. Based on recent research[21, 35-37], we believe that the enzymatic synthesis of HMOs and nucleotide sugars will address these challenges and be ready for industrial implementation soon. Since there are nine main nucleotide sugars[38] and over 200 HMOs[39], efficient, large-scale production of all of them is a task that will require time and the efforts of many people. In this context, we selected CDP-glycerol and the two nucleotide sugars CMP-Neu5Ac and GDP-fucose, along with the HMO LNFP III as the target molecules for this study. The availability of enzymes has been identified as a significant bottleneck that can be addressed through the utilization of DNA recombinant technology. Nevertheless, challenges remain regarding the availability of nucleotide sugars that necessitate solutions if we aspire to produce nucleotide sugars and human milk oligosaccharides (HMOs) at a commercial scale. This research addresses these challenges by identifying optimal operational conditions (pH, temperature, cofactor concentrations) through a design of experiments (DoE) setup; optimising relative enzyme loads via mathematical kinetic modelling; and minimizing the time and resources required through an enzyme co-expression strategy. In conjunction with various proof-of-concept and exploratory experiments, this research successfully achieves the synthesis of nucleotide sugars at a gram scale, while considering process design choices that will enable its advancement in the future.

2 Theoretical Background

2.1 Nucleotide Sugars and Human Milk Oligosaccharides

Nucleotide sugars were first discovered in the 50´s by Luis Federico Leloir and his research group at the Campomar Foundation in Argentina; this led to a variety of carbohydrate metabolism discoveries, like the elucidation of the pathways for sucrose and glycogen biosynthesis in plants. Leloir was awarded the Nobel Prize in chemistry for “his discovery of sugar-nucleotides and their role in the biosynthesis of carbohydrates”[22]. Nucleotide sugars are formed by a monosaccharide and a nucleoside phosphate moiety, and they are the building blocks for the biosynthesis of carbohydrate glycans and conjugates in all known organisms [2]. This biosynthesis occurs through two known metabolic pathways: the de novo pathway, which starts from glucose (it requires more enzymes than the salvage pathway, as well as an energy-rich NADPH cofactor), and the salvage pathway, which uptakes a species of monosaccharide to form the nucleotide sugar directly[40, 41]. According to the KEGG database, more than 60 nucleotide sugars exist in humans[42, 43], and these are precursors to various glycoconjugate molecules such as proteoglycans and glycosphingolipids[44]. There are nine nucleotide sugars that stand out by their abundance in the human cell and serve as donors for the following monosaccharides: D-glucose (Glc), D-mannose (Man), D-galactose (Gal), D-glucuronic acid (GlcA), L-fucose (Fuc), *N*-acetyl-D-glucosamine (GlcNAc), *N*-acetyl-D-galactosamine (GalNAc), D-Xylose (Xyl) and *N*-acetyl-D-neuraminic acid (Neu5Ac)[45], Figure 1. Since natural metabolic pathways utilize specific enzymes to synthesise each nucleotide sugar, an independent production process is required for each. In this thesis, we focus on the production processes of the nucleotide sugars CMP-Neu5Ac, GDP-fucose, and the nucleotide sugar-like molecule CDP-glycerol. These high-value molecules have nutritional and therapeutic potential that remains unexplored due to a lack of bulk supply.

One of the most promising uses of nucleotide sugars in the nutrition industry is the enzymatic synthesis of HMOs. Almost a century ago (1930), a carbohydrate fraction called gynolactose was identified in human milk. Through observation, it was discovered that the faecal bacterial composition of breastfed and bottle-fed babies differed, and the effect was attributed to this fraction, which was named the gynolactose fraction. This discovery ignited a research journey that continues to this day, aiming to characterise every human milk oligosaccharide in structure and function[46].

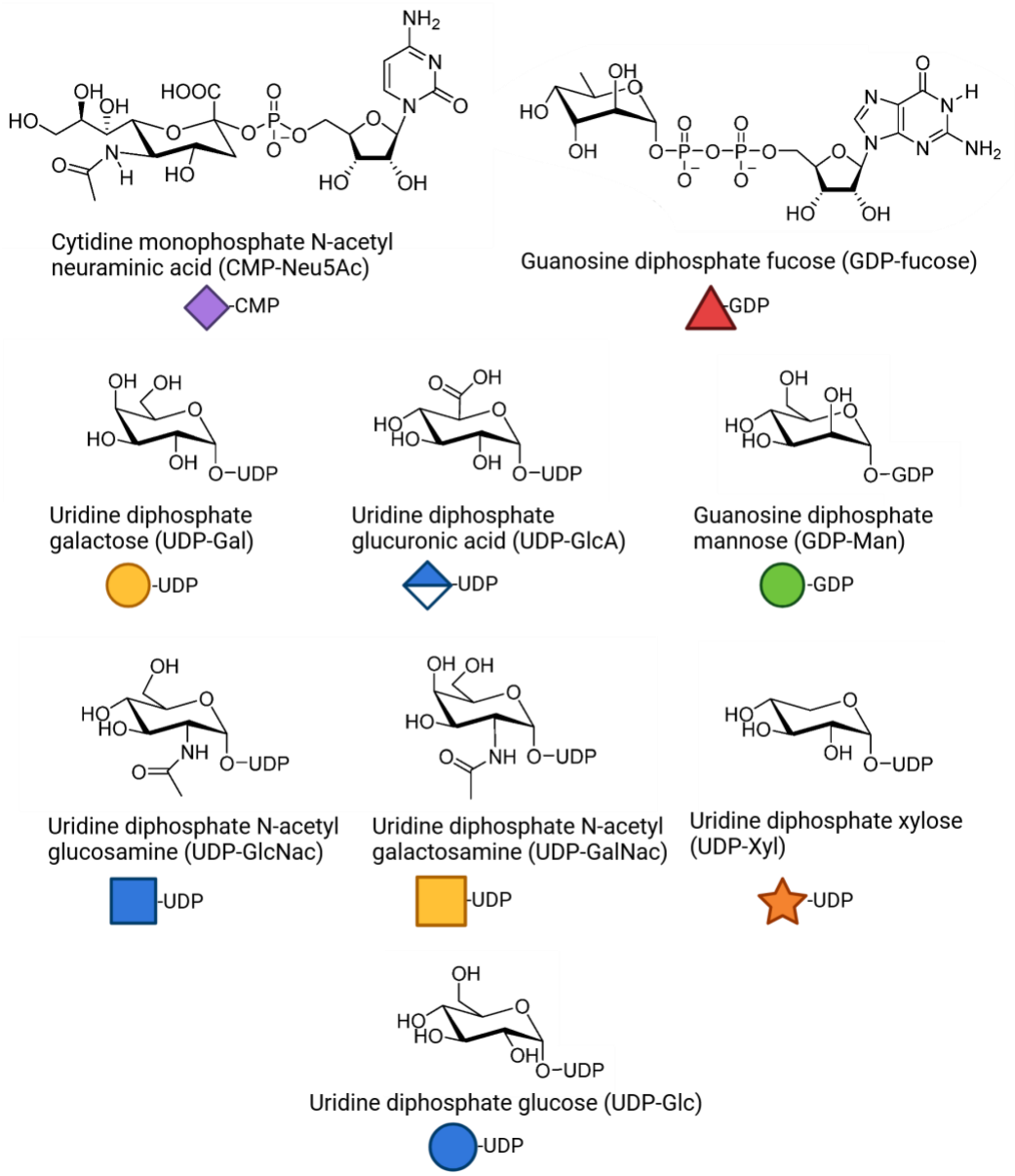


Figure 1 - Most abundant nucleotide sugars in the human cell. Chemical structures of D-glucose (Glc), D-mannose (Man), D-galactose (Gal), D-glucuronic acid (GlcA), L-fucose (Fuc), *N*-acetyl-D-glucosamine (GlcNAc), *N*-acetyl-D-galactosamine (GalNAc), D-Xylose (Xyl) and *N*-acetyl-D-neuraminic acid (Neu5Ac). SNFG symbols are also illustrated below each structure.

More than 200 HMOs have been identified today, and their structure consists of a lactose core structure that can be elongated or branched and can contain Neu5Ac, fucose, GlcNac, glucose or galactose[6]. The HMO content in human milk varies widely, depending on geographical, temporal, nutritional and genetic circumstances. In addition, the HMO content also changes during the various stages of lactation, making it challenging to define general values for HMO content. One of the most significant factors determining HMO composition is a variation in the FUT2 gene, which determines in a major way whether or not a woman will have fucosylated HMOs in her milk[47]. The tremendous variation in contents and variety of HMOs in women makes research complex, but big steps have been made. Nowadays, a significant amount of work has been done to understand not only

each HMO but also their structural interactions and mechanisms of action regarding human health[48]. An abundance of information on HMOs associated with human health is available to us now. For example, we now know that HMOs are essential for their substantial prebiotic effects that promote the growth of beneficial bacteria in the infant gut[39], and they also prevent pathogens from attaching to the mucosa of the digestive tract tissue[49], effectively reducing the risk for infants to contract infections. Interestingly, research suggests that certain HMOs can regulate inflammation and complex immune system interactions[50], opening a whole set of possibilities for research in HMOs and early life infant development. Sialylated HMOs are believed to support healthy cognitive development through specific gene regulation in early life[51]; nevertheless, evidence for this in human studies is sparse, and the mechanisms behind it remain elusive. In a more particular case, it was found that the presence of a single HMO in the diet of infants, disialylacto-N-tetraose (DSNLT), was correlated with the prevention of necrotising enterocolitis, a devastating and often fatal disease[52].

With all the data pointing towards how important it is for infants to have HMOs in their diet, the goal of having modern milk formula with HMOs as an ingredient emerges. At the time of writing, there are a few baby formulas on the market that contain some simple HMOs. With Nestle adding as many as six of them (2'-fucosyllactose (2´FL), difucosyl lactose (DFL), 3 3'-fucosyllactose (3´FL), 6´-sialyllactose (6´SL), lacto-N-neotetraose (LNnT) and 3´-sialyllactose (3´SL)). To this end, much progress has been made in producing HMOs through fermentation[6]. The fermentative approach has been developed successfully into a commercial process by the company Jennewein (now Chr. Hansen), and various research groups worldwide are making exciting progress using metabolic engineering tools and other bioprocess optimisation strategies[29, 36, 53]. Unfortunately, a major limitation has been found in producing HMOs with fermentation processes. When producing more complex HMOs, cellular export becomes inefficient and significantly reduces product yield compared to simple HMO production[54]. Although lower in abundance, complex HMOs might have significant health functions, but little is known because recent HMO research is heavily focused on simpler HMOs with three to four monosaccharide residues[55]. For a more in-depth analysis of the recent data on HMO content, I refer the reader to this review[56] by Soyyilmaz and colleagues, up to date until 2021. A rough estimate based on the available pooled globalized data for the composition of mature milk is shown in Figure 2 to illustrate how abundant different types of HMOs are to each other. The most abundant HMOs tend to be simple fucosylated HMOs like 2´FL (with an obvious exception in non-excretory mothers) and the content of complex HMOs, considered as containing four or more monosaccharides, is estimated to take around 60% of the total HMO content.

This work utilises a cell-free enzymatic cascade methodology to facilitate the efficient production of human milk oligosaccharides (HMOs). By removing the cellular membranes and the burdens of central metabolism, this approach will become the industry standard for producing complex HMOs in the future.

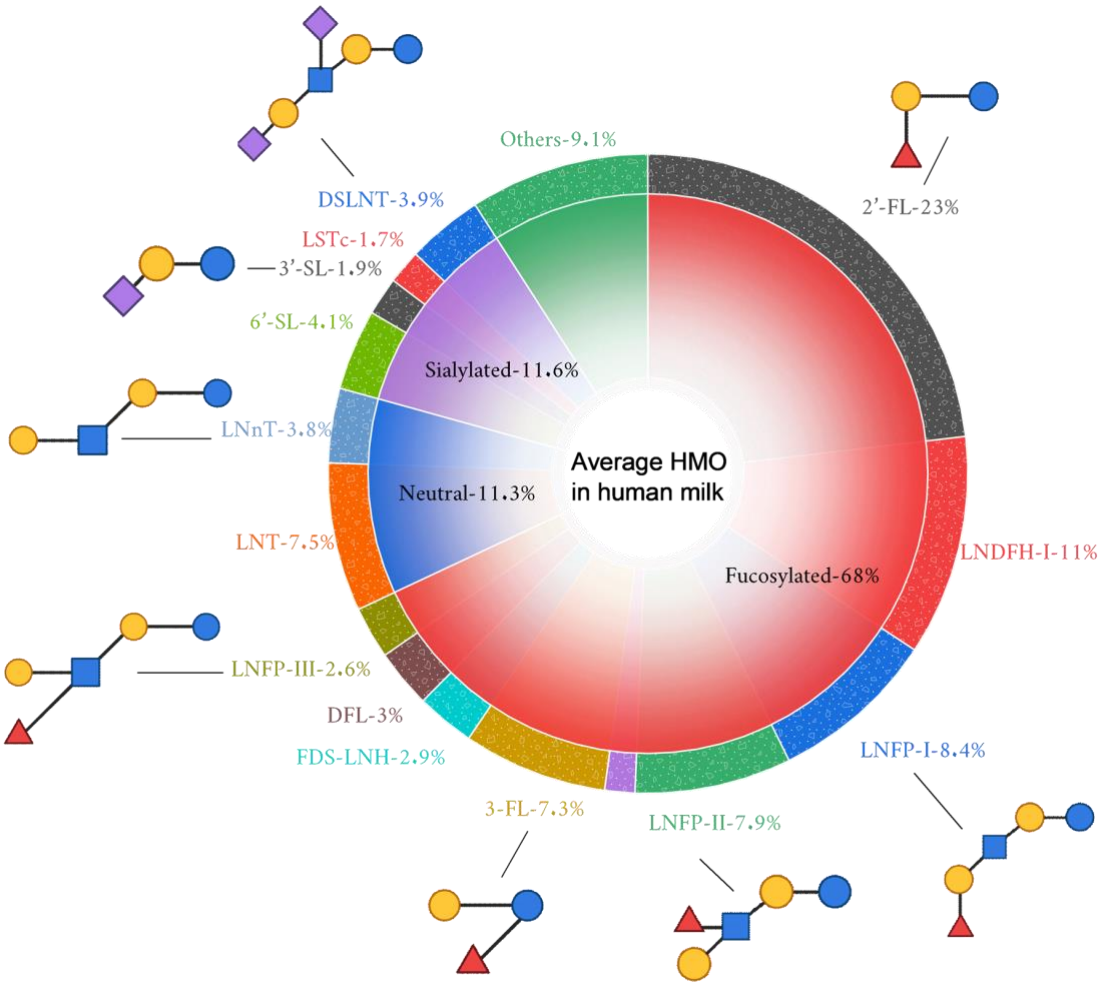


Figure 2 – Content and structure of HMOs in human milk. A core structure made from lactose is elongated or branched with Glc, Gal or GlcNac residues to form neutral HMOs, which account for approximately 11% of all HMO content. These core structures can be fucosylated (68% of total HMOs) or sialylated to form acidic HMOs (11% of HMO content). This figure was constructed based on the averaged pooled global data of mature human milk from available literature up to 2021[56].

2.2 Cell-free Enzymatic Cascades

As the industrial production of specialty chemicals strides towards sustainable manufacturing, the development of bio-based solutions is increasingly taking the spotlight. Fermentation and metabolic engineering technology have been developed to such a degree that very complex products can be produced with microorganisms nowadays. However, many of these processes have been improved or replaced in some cases by in vitro enzymatic processes, notably in the production of food additives, pharmaceutical active compounds and polymers like plastics and carbohydrates[57, 58]. Arguably, the closest technology to cell-free synthesis is whole-cell biocatalysis, where the cells that produce the enzymes are kept in the reaction. This is done not because of fermentation, but for other possible

beneficial effects such as natural buffering, co-factor recycling, and cheaper production of some pathway components. On the other hand, cell-free biocatalysis offers many advantages like high volumetric productivity, no concern about toxicity of cellular intermediates, no competition with the central metabolism, and — possibly the biggest advantage— a high control and freedom of design compared to whole cell systems[59] (see Figure 3). For a complete discussion on whole-cell and cell-free biocatalysis, I recommend the review article by Claassens and colleagues[60]. There, it is highlighted that the system choice will be dictated on a case-by-case basis by the specific application and product to be produced. There is a growing trend of using cell-free systems to produce sugar nucleotides and specialty carbohydrates like HMOs. This trend in cell-free systems is made possible by the steadfast development of technologies like high throughput enzyme mining/discovery, protein engineering, DNA synthesis, and bioinformatics. Cell-free biocatalysis is becoming an accessible technology with increasingly promising potential[61]. There are examples of cell-free biocatalysis being used for the production of commodities like ethanol, hydrogen, and butanol that have substantial potential[62]. Still, a significant number of state-of-the-art examples of applied biocatalysis are of late-stage methyl functionalisation for the production of pharmaceuticals. One recent example by Lowell and coworkers is the hydroxylation of the natural product M-4365 G1A (a natural macrolide antibiotic produced by *Micromonospora capillata sp.)* by a P450 monooxygenase to form the antibiotic juvenimicin B1[63]. Interestingly, recent examples of C-C bond formation[64] showcase the potential of biocatalysis to tackle these challenging reactions and, more importantly, carbohydrate chemistry. Coming back to Leloir glycosyltransferases, One-pot multienzyme (OPME) synthesis has been thoroughly adopted in recent years to provide glycosyltransferase-catalyzed reactions with nucleotide sugars. Cell-free OPME systems have clear advantages for this particular application: tighter control over the reaction conditions, no transport limitations (no cell membrane), no competition of resources with the cell's central metabolism, and reduced incidence of side products are some of these[12].

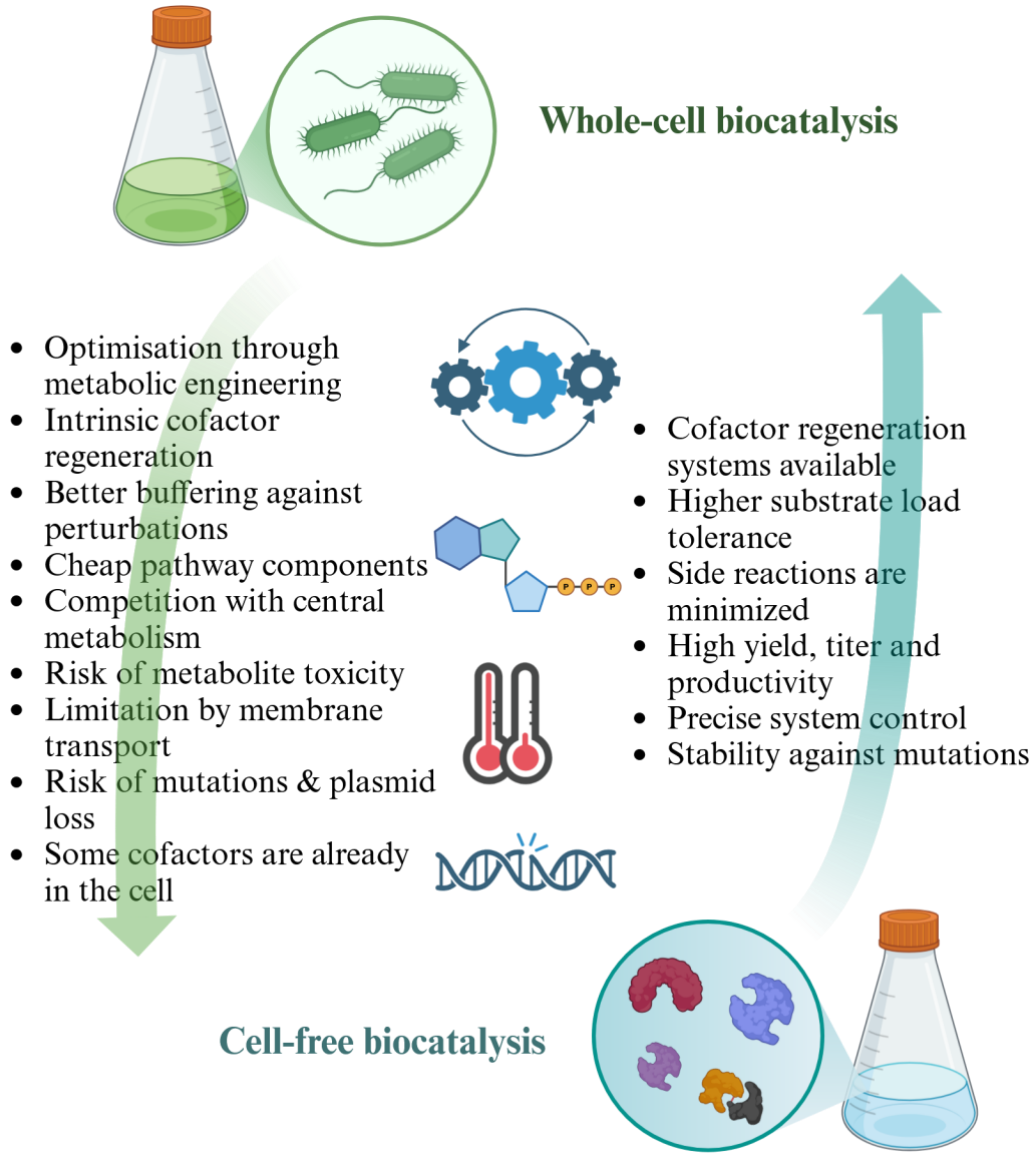


Figure 3 - Cell-free vs whole-cell biocatalysis. The main advantages of whole-cell systems are the cheaper substrates required to produce metabolic intermediates and *free* cofactors. In contrast, cell-free systems excel in processes that need to be tightly controlled and have the potential to achieve higher product yields.

The cell-free approach was first developed to synthesise CMP-sialic acid derivatives[65], and it was inspired by the natural biosynthesis routes of nucleotide sugar synthesis. These are referred to as enzymatic cascades and have proven to be highly flexible due to their modular nature. A monosaccharide phosphorylation step performed by glucokinases is often coupled to a suitable nucleotide transferase[8], and advances in nucleotide triphosphate (NTP) regeneration techniques allow, for example, minimal use of expensive adenosine triphosphate (ATP) as a cofactor[66]. To this day, OPME systems have been used successfully for the synthesis of a variety of nucleotide sugars, including CMP-Neu5Ac[67], UDP-Glc, UDP-Gal[68], UDP-GalNAc[69], UDP-GlcNAc[70, 71], UDP-Xyl[72], GDP-Man[73, 74], GDP-fucose[32] and even non-natural sugars[75-79].

2.3 State of the Art

Although various methods have been used to synthesise nucleotide sugars in the past, they have rarely been produced at gram scales or above. This section discusses the historical background and prior art regarding production methods.

2.3.1 CDP-glycerol

CDP-glycerol is a molecule structurally similar to nucleotide sugars, with applications in glycoconjugate vaccine development. Glycoconjugate vaccine technology has been in development since the 1980s, and it has already found its way into widespread medical use[80]. Utilising vaccines as a preventive measure against bacterial infections presents an attractive alternative to antibiotic treatment. Initial bacterial vaccines consisted of attenuated pathogens[81]. Later, surface carbohydrate structures on bacteria emerged as promising vaccine targets. The first carbohydrate vaccine, Pneumo-Vax, released commercially in 1983 by Merck and Co., involved capsular polysaccharides isolated from 14 *Pneumonia* serotypes[82]. However, to overcome the inability of some polysaccharide vaccines to elicit a T-cell-dependent immune response, glycoconjugate vaccines were developed[83]. These new vaccines integrate a polysaccharide antigen, typically derived from pathogen cultures, with a highly immunogenic carrier protein. Several Gram-negative bacteria produce capsule polymers, such as poly (glycosylglycerol phosphate), which can serve as antigens for novel glycoconjugate vaccines targeting pathogens like *Actinobacillus pleuropneumoniae*, *Neisseria meningitidis*, and *Bibersteinia trehalose*[84]. Structurally analogous polymers are found in Gram-positive wall teichoic acids (Figure 4), which have diverse biological functions that make them interesting vaccine targets[85]. Recent advancements have introduced synthetically produced polysaccharides, which can surpass their natural counterparts regarding immunisation efficacy and accessibility while maintaining safety[86, 87]. Although these polymers can be synthesised in vitro through enzymatic reactions, the process requires CDP-glycerol as a substrate for the wall teichoic acid polymerising enzymes[88], and CDP-glycerol is not commercially available, hindering research into glycerol-phosphate-containing polymer synthesis. While the chemical synthesis of CDP-glycerol is known, it is laborious and yields only low substrate conversion[89]. An enzymatic one-step reaction synthesis was proposed[88, 90], utilising glycerol phosphate cytidylyltransferase (tagD) and glycerol kinase (glpK). However, this enzymatic synthesis relies on expensive substrates—CTP and glycerol-3-phosphate (Gly-3P).

The scale and efficiency, in the form of total product synthesised and substrate conversion yield (Yp/s), are useful for comparing different synthesis attempts and understanding their potential to become a fully developed production process. Table 1 summarises the most relevant literature on CDP-glycerol synthesis for various methods and production scales, most of them in the order of milligrams. The lack of gram-scale production processes in the literature attests to the underdeveloped research for the production of this molecule. In time, advancements in the glycoconjugate vaccine field will increase the demand for CDP-glycerol, justifying further development into its large-scale production.

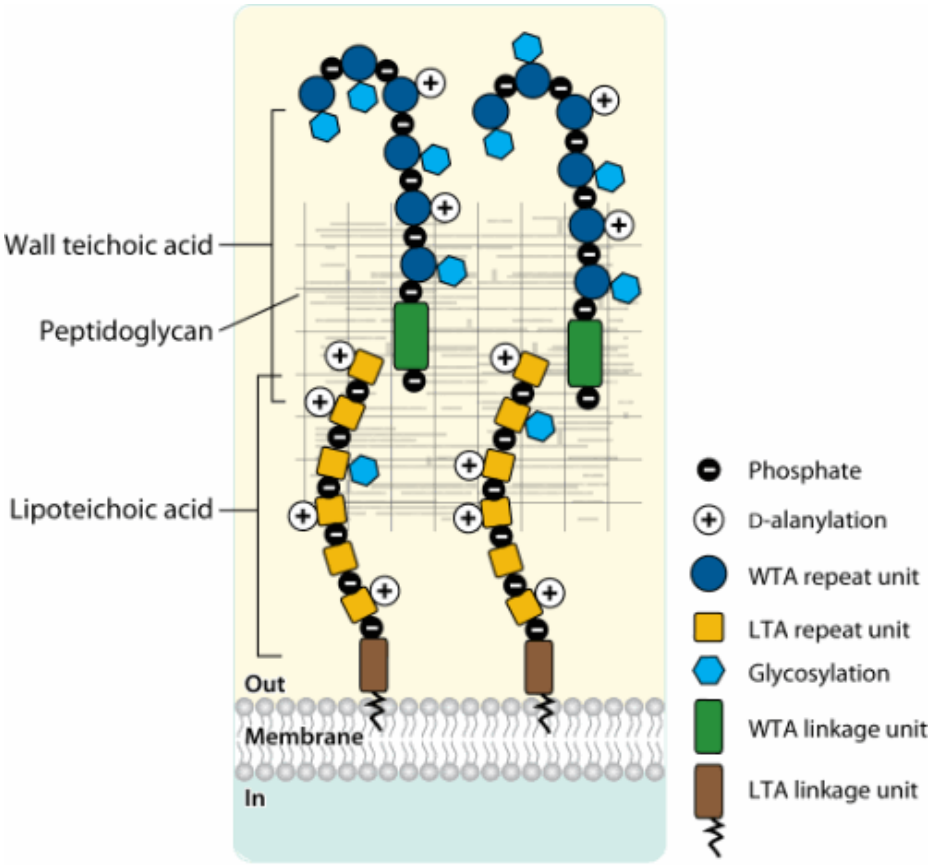


Figure 4 - Structure of wall teichoic acids from Gram-positive bacteria. Lipoteichoic acids (LTAs) are coordinated to the cell membrane, and wall teichoic acids (WTAs) are covalently linked to peptidoglycan (PG). The WTA repeating unit consists of polyol phosphates whose biosynthesis requires CDP-glycerol as a building block. This figure was adapted from Brown et al. 2013[91], under the Annual Reviews license agreement no. 1544743-1.

This thesis introduces an enzymatic cascade for CDP-glycerol synthesis using the cost-effective substrates cytidine and glycerol. The cascade design is similar to previously established enzymatic cascades for sugar nucleotide synthesis[32, 68, 71, 73]. A DoE approach was used to screen and identify optimal operation conditions for temperature, pH and cofactor concentration. These optimised conditions signify an essential step towards applying the cascade in an industrial context in the future.

Table 1 - Summary table of CDP-glycerol syntheses reported in the literature.

Method	Total Amount	Form	Substrate / Yield	Reference
Chemical	600 mg	Ca^{+2}, Li^{+} salt	Cytidine / 45 %	Baddiley, 1958 [89]
Chemical	0.5 mg	Purified	Cytidine phosphoromorpholidate / 65 %	Roseman, 1961 [92]

Method	Total Amount	Form	Substrate / Yield	Reference
Enzymatic	--	Unpurified	CTP / detected in TLC	Shaw, 1962 [93]
Enzymatic	0.62 µg	Unpurified	CTP / 97 %	Badurina, 2003 [90]
Enzymatic	--	Intermediate	CTP / detected in HPLC	Litschko, 2021 [88]
OPME	3 mg	Unpurified	Cytidine / 89 %	Alcalá, 2023 [94]

*CTP – cytidine triphosphate; TLC – thin layer chromatography; HPLC – high performance liquid chromatography; OPME – one pot multienzyme.

2.3.2 CMP-Neu5Ac

As one of the most abundant sugar nucleotides, CMP-Neu5Ac has a variety of biological roles in living organisms. It is known to be a donor of Neu5Ac (the most common form of sialic acid) for the biosynthesis of many polysaccharides and their conjugates, which have prominent functions in cellular recognition and modulation events, particularly in immune system regulation, cell adhesion and communication[95]. The most relevant example of such events occurs in the human immune system, where cell signalling effectors are mostly sialylated glycoproteins on the surface of immune cells. The presence of sialic acid (and its derivatives) in the surface glycoproteins of immune cells has effects on self/non-self-discrimination, macrophage mediates phagocytosis, lymphocyte-mediated cytotoxicity, maturation/activation of immune cells and adhesion phenomena, among other mechanisms that are yet to be elucidated[96]. In recent literature, sialylated molecules and their associations with inflammatory diseases are being studied[97]. The importance of sialylated molecules in the human immune system makes them important targets for novel therapeutics. For example, it was recently found that abnormal regulation of the enzymes responsible for the sialylation of low-density lipoprotein receptors and blood cells is associated with atherosclerosis; therefore, making inhibitors for such enzymes is now a potential opportunity for treatment[98]. In another example, regarding a different mechanism, the ability of sialic acid to mask galactose moieties in glycoproteins can be used to increase the circulatory half-life of therapeutic proteins in the organism[99], making the glycoengineering of proteins like monoclonal antibodies desirable as a way to enhance their pharmacokinetic attributes. Lastly, CMP-Neu5Ac is highly relevant in the context of cancer research. It has been observed that there is an abnormal overexpression of sialylated glycoproteins in the surface of some tumours, which has been associated with characteristics like tumour growth, escape from apoptosis, metastasis formation, and resistance to therapy[100, 101]. Striding away from therapeutic prospects, and looking at nutrition and human development. In this front, sialic acids have an important biological role in the biosynthesis of HMOs; around 13% of HMOs are sialylated on average[102], not counting the complex HMO

structures. It has been found that acidic HMOs have a role in supporting resistance against pathogens, gut maturation, and cognitive development in humans, particularly in infants[103]. In modern day, an impressive amount of research has been done into the effects of sialylated HMOs in humans and their production methods. In a recent review by Zhu and colleagues, the sialylated HMOs 3´-SL and 6´-SL are looked at in detail as key HMOs that are now recognised as GRAS (Generally Regarded As Safe) by the FDA (American Food and Drug Association) and are starting to be added to baby milk formula[104]. The large-scale production strategies that have prevailed for sialylated HMOs are fermentative methods with genetically engineered strains. These methods have been successfully used for simple HMOs like 3'-SL and 6'-SL[105-107]. These systems bypass the need for an expensive substrate and utilise only glucose or glycerol as the starting raw material, effectively compensating for the relatively low titers by making the starting material very cheap. Even though success has been achieved with simple sialylated HMOs, not all are available commercially, which implies other challenges. For instance, modular approaches such as cell-free platforms may be needed to synthesise more complex HMOs, as whole-cell systems are prone to producing side-products. The availability of better sialyltransferases will also be necessary to improve such processes further. Advances in protein engineering will soon enable the creation of better enzymes, which may withstand unnatural conditions, like high temperatures, and promote faster enzyme kinetics, but these may be incompatible with fermentation systems. Lastly, more efficient methods of providing the crucial intermediate CMP-Neu5Ac will be necessary, whether the final product (HMO or therapeutic protein) is produced in vivo or in vitro.

The efforts to effectively synthesise CMP-Neu5Ac are divided between whole cell and enzymatic methods; see Table 2 for an overview of selected synthesis attempts. Although the first CMP-Neu5Ac enzymatic synthesis was reported in 1993 by Liu and colleagues[108], reports of gram-scale production setups are scarce. Although various processes have been reported, but no significant progress has been made in making this intermediate more available. At the time of writing, the price of CMP-Neu5Ac is more than 1000 Euros for only 25 mg of the disodium salt (summer 2024).

In this work, we take the OPME system for the synthesis of CMP-Neu5Ac and take several process intensification steps to make it widely available in the near future. These steps include optimising operation conditions with a DoE strategy, co-expressing the cascade reaction enzymes, and a preparative multi-gram scale synthesis setup.

Table 2 - Summary table of previous CMP-Neu5Ac syntheses reported in literature.

Method	Total Amount	Form	Substrate / Yield	Reference
OPME	1.1 g	Purified, freeze-dried	CMP/ManNac / 98%	Liu, 1993 [108]
Whole-cell	510 mg	Unpurified	Neu5Ac / 48%	Endo, 2000 [109]

Method	Total Amount	Form	Substrate / Yield	Reference
OPME	n.a.	Unpurified	CMP/Neu5Ac / 67%	Ishige, 2001 [110]
Whole-cell	1.1 mg	Unpurified	CMP/ManNac / 90%	Lee, 2002 [111]
Whole-cell	406 mg	Purified	CTP/ManNac / 84 %	Song, 2003 [112]
Immobilized Enzyme	6 mg	Unpurified	CMP/ManNac / 99%	Nahálka, 2004 [113]
Chemo-enzymatic	100mg	Purified	CTP/ManNac / 99 %	Yu, 2004 [65]
Whole-cell	108 mg	Unpurified	CMP/GlcNac / 60%	Hamamoto, 2005 [114]
Enzymatic, Whole-cell	340 mg	Unpurified	CMP/ManNac / 73%	Nahálka, 2009 [115]
OPME	5.56 mg	Unpurified	CMP/Neu5Ac / 90%	Mahour, 2022 [116]
OPME	1.16 mg	Unpurified	Cyt/Neu5Ac / 95%	Mahour, 2022 [116]
OPME	3.02 mg	Unpurified	Cyt/GlcNac / 77%	Mahour, 2022 [116]
Co-expressed OPME	6.4 g	Unpurified	CMP/GlcNac / 99%	This work

*OPME – one pot multienzyme; CMP – cytidine monophosphate; ManNac – N-acetylmannosamne; Neu5Ac – N-acetylneuraminic acid; GlcNac – N-acetylglucosamine; Cyt – cytidine.

2.3.3 GDP-fucose

Fucosyltransferases are enzymes present in many kinds of living organisms, and their function is to perform the last step in synthesising fucosylated products. These fucosylated products can be proteins or oligosaccharides that, for example, have functions in cell adhesion and lymphocyte recirculation phenomena in humans[117]. Fucosylation is an abundant and important oligosaccharide modification involved in inflammation processes and even cancer; recent glycobiology research has revealed that fucosylation can be used as a biomarker for detecting several carcinomas. This had such an impact that fucosylated alpha-fetoprotein is now widely used for the diagnosis of hepatocellular carcinoma[118]. Regarding the potential for therapeutics, fucosylation does not fall behind in relevance. Considering the monoclonal antibody industry, glycosylation of these therapeutics is now considered a critical quality characteristic. Core fucosylation, for instance, weakens the complement-dependent cytotoxicity mechanism, effectively reducing the efficacy of monoclonal antibodies to recruit immune effector cells in cancer

therapies[119]. Developing non-fucosylated or low-fucosylated antibodies is highly desirable to enhance many already effective therapeutics[120-122], especially when these use the recruitment of immune effectors to kill diseased cells as the primary mechanism of action. As a direct consequence, most of the research on fucosylation for therapeutics revolves around removing it, not adding it. Large-scale production of GDP-fucose sparks little interest in the area of therapeutics. However, in recent years, another research trend has made the large-scale production of GDP-fucose very appealing. This trend is the interactions of fucose with prokaryotic organisms, particularly in the human microbiota. In prokaryotic organisms, fucosylation is less common than in eukaryotes. Still, it has a significant involvement in several cellular processes like molecular mimicry, adhesion, colonisation, and modulating the host immune response[123]. These effects, especially the ones regarding bacterial colonization, are why fucosylation events in bacteria are closely related to one fascinating example of human fucosylated molecules, fucosylated HMOs. As shown, most of the HMOs found in human milk comprise fucosylated HMOs, with the most abundant being 2´-FL, followed by slightly more complex ones like LNFP I and LNFP III. These HMOs have been correlated to the development of a healthy gut microbiota in infants and supporting their immature early immune system[124]. The mechanisms behind this have been studied in the past years. Evidence points towards a double effect of preventing pathogen adhesion while serving as nutrients to beneficial bacteria. Some pathogenic bacteria like *Pseudomonas aeruginosa*, *Campylobacter jejuni* and *Streptococcus agalactiae* recognise fucosylated glycans as part of their adhesion mechanisms. Fucosylated oligosaccharides in the gut act as a decoy to prevent these bacteria from attaching[125]. At the same time, it has been observed that beneficial bacteria from the genus *Bifidobacterium* thrive on and consume the HMOs 2′-FL, DFL and 3´-FL when anaerobically cultured with these[126], clearly showing that fucosylated HMOs have a stellar role in the development of healthy gut microbiota.

As promising as all the applications for fucosylated biomolecules may seem, and even when the different fucosyltransferases are capable of adding fucose to a wide variety of acceptors, all fucosyltransferases have one thing in common: they require GDP-fucose as a fucose donor molecule. GDP-fucose is among the most expensive nucleotide sugars available, and no large-scale provider exists for this nucleotide sugar; this significantly hinders the development of new applications for fucosylated conjugates in therapeutics or nutrition. There has been much interest in developing a synthesis process that is efficient enough to bring the scale up and the cost down enough to justify using GDP-fucose as a raw material for the synthesis of HMOs and fucosylated therapeutic glycoproteins. GDP-fucose is synthesised in nature via two different pathways: the de novo pathway and the salvage pathway[127]. The de novo pathway produces most of the GDP-fucose in mammals, making it from GDP-mannose[128]. In contrast, the salvage pathway produces GDP-fucose by *activating* (phosphorylating) exogenous fucose through ATP-dependent enzymes[129]. This has inspired attempts to synthesise GDP-fucose enzymatically in vivo and in vitro systems. In 1975, Prohaska and colleagues reported an GDP-fucose enzymatic synthesis with a substrate conversion yield of up to 81%[23]. At the time, recombinant enzyme technology as we understand it today was not yet available; therefore, the extraction of enzymes was conducted directly from porcine organs. This represented a bottleneck for the enzymatic synthesis strategies for at least another two decades. Up until the 90s, chemical

synthesis processes were created and reached up to 300 mg[5, 26] of product. However, these processes require many chemical reaction steps and could take days or even weeks to complete. As fermentation and recombinant DNA technology developments facilitated the availability of enzymes for biocatalysis, the syntheses of GDP-fucose reported in literature became heavily dominated by enzymatic, fermentative or whole-cell bioconversion strategies. An overview of selected syntheses reported in the literature is shown in Table 3. As far as the author is aware, the only gram-scale syntheses of GDP-fucose reported in the literature to date are the OPME reaction reported by Li and colleagues[130], where they could synthesise 1.6 g of the product and the repeated batch OPME setup reported recently by Frohnmayer, where they also managed to produce 1.6 g of GDP-fucose, but in this case, in a repeated batch system that allowed them to re-use the biocatalyst and save resources[33].

While clear progress is being made, the current costs and poor availability of GDP-fucose suggest that there is still much work to be done before these processes are mature enough to be brought to a large scale. In this work, the focus was on intensifying the already established OPME strategy for synthesising GDP-fucose to bring it closer to industrial implementation. The process intensification strategies applied to this enzymatic cascade include enzyme co-expression, kinetic mathematical modelling, reaction conditions optimisation through the DoE methodologies and preparative gram-scale production of GDP-fucose in a batch reaction setup.

Table 3 - Summary table of previous GDP-fucose syntheses reported in the literature.

Method	Total Amount	Form	Substrate - Yield	Reference
Enzymatic - hog extracts	16.4 mg	Purified, freeze-dried	Fucose – 81 %	Prohaska, 1975 [23]
Chemical	300 mg	Purified, freeze-dried	Guanosine 5'-monophosphomorphalidate – 50 %	Ichikawa, 1992 [26]
Whole-cell	17 mg	Purified, freeze-dried	GMP – 11 %	Koizumi, 2000 [131]
Enzymatic – two steps	78 mg	Purified, 2Na salt	GDP-α-d-mannose – 78%	Albermann, 2000 [132]
Enzymatic	30-50 mg	Isolated	L-fucose & derivatives – >75%	Wang, 2009 [133]
Enzymatic	150 mg	Purified, freeze-dried	GTP – 50% L-fucose – 99%	Zhao, 2010 [27]
Whole-cell	0.13 mg	Unpurified	L-fucose – 0.2 %	Liu, 2011 [134]
Fermentation	33.7 mg	Unpurified	Glucose – n.a.	Zhai, 2015 [10]

Method	Total Amount	Form	Substrate - Yield	Reference
OPME – three steps	n.a.	Unpurified	Mannose – 14 %	Wang, 2019 [135]
OPME	1.6 g	Purified, freeze-dried	L-fucose – 68%	Li, 2021 [130]
OPME	8.25 mg 8.95 mg	Unpurified	L-fucose – 70% Mannose – 72 %	Mahour, 2021 [32]
OPME, – repeated batch	1.6 g	Unpurified	L-fucose – 31%	Frohnmeyer, 2022 [33]
Co-expressed OPME	1.76 g	Unpurified	L-fucose – 98%	This work

*GMP – guanosine monophosphate; GDP – guanosine diphosphate; GTP – guanosine triphosphate; OPME – one pot multienzyme.

2.3.4 HMOs – LNFP III

A variety of production approaches have emerged to make HMOs available as nutritional ingredients, from extracting them directly from human milk or through the enrichment of bovine milk to the many synthesis routes such as chemical, enzymatic, chemo-enzymatic, bacterial fermentation and cell culture[136]. The structure of LNFP III is shown in Figure 5.

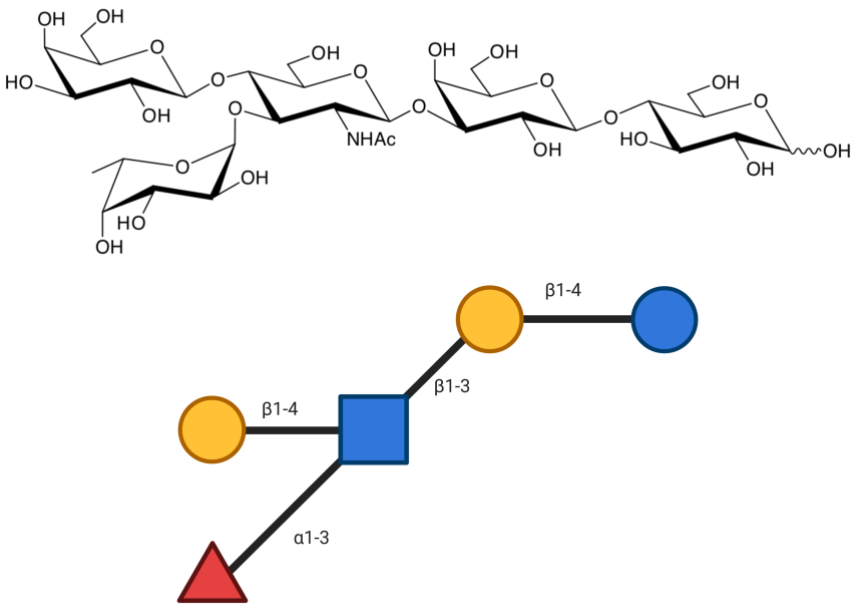


Figure 5 - Structure of LNFP III. Molecular structure (top) and SNFG nomenclature (bottom).

In the specific case of fucosylated HMOs, enzymatic biocatalysis has been the most studied in scientific literature. Various fucosylated HMOs have already been synthesised using cell-free enzymatic cascades[137]. On the other hand,

fermentative and whole-cell approaches have been reported to synthesise small HMOs like 2´-FL[138] at a commercial scale.

A challenge for fermentation is the formation of side products, which makes it more difficult to apply the production of more complex HMOs. For example, if the desired metabolic pathway uses lactose as a starting substrate for the in vivo production of LNnT and further fucosylation into LNFP III, the fucosyltransferases responsible for this last step could also take lactose directly as a substrate, and the fermentation will accumulate 2´FL instead. In a recent publication, however, Sugita and colleagues from the Kirin Central Research Institute in Japan reported a fermentation process that yielded 11.5 g of LNFP III in a 3 L reactor[139]. This process was made possible by a novel α1,3-fucosyltransferase from *Parabacteroides goldsteinii,* which does not utilise lactose as a substrate, significantly reducing the number of side products and, for the first time, efficiently producing a complex HMO through fermentation. A summary of the reported synthesis of LNFP III is shown in Table 4.

Table 4 - Summary table of previous LNFPIII synthesis reported in literature.

Method	Total Amount	Form	Substrate / Yield	Reference
Enzymatic (chicken serum)	2.1 mg	Partially purified	LNnT/GDP-fucose / 20%	Totani, 2002 [140]
Fermentation	1.5 g	Partially purified	Glucose/Lactose / 30%.	Dumon, 2004 [141]
Enzymatic	1.7 mg	Unpurified	3′-FL/LNnT / 21%	Saumonneau, 2016 [142]
OPME	Detected on HPLC	Unpurified	LNnT/GTP/fucose / 88%	Yu, 2017 [30]
Enzymatic	Detected on HPLC	Unpurified	3′-FL/LNnT / 11%	Zeuner, 2018 [143]
Enzymatic	Detected with ^{1}H NMR	Unpurified	LNnT/GDP-fucose / 52%	Huang, 2019 [144]
Enzymatic	Detected by 2D NMR	Unpurified	LNnT/GDP-fucose / n.a.	Bai, 2019 [145]
Fermentation	11.5 g	Unpurified	Glucose/Lactose / n.a.	Sugita, 2023 [139]
OPME	1 mg	Unpurified	LNnT/GMP/fucose / 68%	This work

*LNnt – Lacto-N-neotetraose; GDP-fucose – guanosine diphosphate fucose; 3´FL – 3' fucosyllactose; OPME – one pot multienzyme; HPLC – high performance liquid chromatography; GTP – guanosine triphosphate; NMR – nuclear magnetic resonance.

Similarly, the critical bottleneck of side-product formation is being addressed for fermentative processes; this work tackles OPME systems challenges of nucleotide sugar availability and costly recombinant enzyme production by implementing a nucleotide sugar regeneration system and an enzyme co-expression platform. Currently, no large-scale process for synthesising LNFPIII is being used commercially, and cell-free OPME systems hold great potential for developing such methods.

2.4 Bioprocess Intensification

Transferring an enzymatic reaction from the test tube to the industrial-scale bioreactor is not a trivial task and is often carried out by multidisciplinary teams. Bioprocess intensification is a significant step towards successfully implementing a process, which requires attaining specific benchmarks in terms of productivity, cost and environmental impact to be industrially competitive. Intensification refers to an increase in bioproduct output relative to either cell concentration, time, reactor volume or cost[146]; an increase that results from strategic changes in technology, setup or optimisation of them. The latter refers to incremental improvement of a process by, for example, changes in operational conditions. Optimisation is often included as a step of the bioprocess intensification process[147]. In general, bioprocess intensification of cascade reactions is done in layers, starting from the cascade design that will dictate the cascade performance fundamentally and sometimes draws inspiration from nature; there are informatic tools built for this very purpose[148]. The process design is also vital to overcome the reaction limitations; for example, an enzyme used in a reaction might have a high reaction rate but is inhibited at high substrate loads. In this case, a fed-batch strategy can be beneficial to bypass such a limitation; as an example, this scenario has been recently described for the enzymatic hydrolysis of sugar cane molasses[149]. In other layer, selecting the suitable catalyst is critical for cascade performance since this is the heart of the whole process. Enzymes can be found from natural sources through enzyme mining[150] in databases like Uniprot[151] and BRENDA[152], which make their information available for the research community. Enzymes can also be further improved thanks to groundbreaking developments like directed evolution[153], which earned Frances Arnold the Nobel prize in 2018, and other rational design strategies that have been recently described[154-157]. A variety of clever process intensification strategies are found in the literature, one of which is ultimately applied in this work to reduce the number of fermentation and purification steps to save costs and resources in the overall process. This strategy is the co-expression of recombinant enzymes in a single host strain using the pDuet vectors from Novagen[158].

Much of modern basic science in biocatalysis focuses on discovering and characterising novel enzymatic reactions. Still, few of these make their way to become utilised in an industrial setting. Upon establishing the enzymatic cascade, improving it and optimising each step is imperative to attain a process of industrial relevance. In biocatalysis, particularly in biocatalysis with enzymatic cascades, the strategies for optimising processes have been explored significantly in recent research [159]. However, a single generally suitable method for optimising enzymatic cascades does not exist, and it must be chosen on a case-by-case basis since every enzyme has different optimal conditions in terms of temperature, pH, component concentrations and solvents. Identifying a set of conditions that will be favourable for all the reactions happening in the cascade is a very challenging task.

The most common practice to identify such conditions is a screening of cascade reactions that often involves a one-at-a-time variation of parameters. Although it has been used successfully in the past[160], this approach is prolonged, resource-consuming (typically high throughput technology is required) and neglects the complexity that characterises enzymatic cascades. This method is unsuitable for identifying interactions between factors, which makes it very limited when used in complex systems; nevertheless, it is still widely used for applications like enzyme characterisation, where it is standard for identifying pH, temperature and cofactor optimums[161-163]. Alternatively, a DoE approach can be used. This screening method is based on statistical tools and has been used for years in applications predominantly encompassing mechanical and materials engineering, but is increasingly being used in all branches of science, including biotechnology[164]. Another approach is using enzyme kinetic models developed to predict optimal inputs for biocatalytic processes. This approach is particularly well-suited for optimising biocatalyst and substrate loads when enzyme kinetic parameters are well characterized[165]. However, in the context of more intricate cascade reaction setups, incorporating additional factors such as pH value and temperature for optimisation renders the selection of kinetics and appropriate parametrisation notably challenging[166]. As no single tool can suit all optimisation problems, this thesis explores the flexible use of the DoE and kinetic modelling optimisation, complemented with traditional exploratory experimentation, to identify the best conditions for operating the three different enzymatic cascades.

2.4.1 Design of Experiments

The concept of DoE was introduced by Sir Ronald Aylmer Fisher, a Cambridge-educated mathematician, in the early 20th century. Fisher was hired to determine the effects of a 70-year-long agricultural experiment in Rothamsted Estate in Harpenden, UK. The experiment consisted of growing crops with different kinds of organic and inorganic fertilisers while recording yield and rainfall data every year to identify the best-performing treatment for each crop. As there was no formal structure for experimentation at the time, the treatments were not randomised, replicated, or blocked. When Fisher looked at the data, he concluded that the effects of treatment and rainfall were confounded with each other; that is, the impact of one could not be separated from the other when analysed, leaving Rothhamstead with 70 years of useless data to show. This experience led Fisher to focus on analysing data statistically and proposing a methodology for *how* this data should be generated in the first place[167]. In time, Fisher published his two works, *Statistical Methods for Research Workers*[168] and *The Design of experiments*[169], which set the ground for the experimental designs still used today. DoE has been used historically more widely in mechanical engineering and agriculture because these fields often deal with random factors that are hard to control –such as the weather in agriculture—or complex phenomena that cannot be easily predicted. DoE is a tool to determine the amount and kind of experimentation needed to discern the effects between treatments. Other fields, like chemistry and physics, traditionally perform experimentation in more controlled environments where traditional experimentation and observation can suffice. This resulted in a much slower adoption of DoE methods in such fields.

Adopting DoE approaches in biocatalysis holds great promise as systems become increasingly sophisticated, especially when targeting scale-up for industrial production[170]. While this approach has yet to be employed for enzymatic

cascades synthesising nucleotide sugars, numerous examples exist where DoE has successfully improved biocatalytic reactions involving a variety of enzyme chemistries[171]. For instance, the improvement of an enzymatic process for Ambrox production (a common ingredient in cough medicine) consisted of a DoE screening for temperature, pH value, and substrate concentrations; this resulted in a titer increase from approximately 20 to >30 g/L[171]. Another recent application of DoE screening in biocatalysis facilitated selecting a suitable enzyme and reaction conditions for the reductive amination of cyclohexanone with cyclopropylamine[172], a helpful building block for synthesising pharmaceutical compounds. Furthermore, DoE has been used for the optomising of a variety of enzymatic cascades. Examples of these include enzymatic cascades for the synthesis of ε-coprolactone from cyclohexanone[173], and the production of dihydrogen from xylose[174]. This work describes the optimisation of OPME cascade reactions for the synthesis of nucleotide sugars by using DoE methods for the first time.

2.4.2 Kinetic Modelling

Some argue that the field of enzyme kinetics was born with Michaelis and Menten´s original publication on the kinetics of an invertase-mediated reaction, released more than a century ago, in 1913[175]. Their breakthrough was, of course, standing on the shoulders of other significant advances of the time, such as the law of mass action[176], previous research on the invertase[177] and the general notions in chemical kinetics[178]. Michaelis and Menten´s contribution lies in analysing time courses in terms of initial rates; this approach makes it possible to analyse a reaction´s kinetics without interfering with factors like enzyme denaturation, substrate depletion, and pH changes over time[179]. The general applicability of the method and the widespread standardisation of the Michaelis and Menten´s equation for enzyme kinetics analysis have resulted in many scientific contributions encompassing all classes of enzymes. As technology develops and computational power increases, enzymatic modelling has gone a long way since the times of Michaelis and Menten, introducing not only time course analysis but also sophisticated computational tools for modelling substrate docking, conformational changes, metabolic networks[180], and even predicting protein tertiary structures with the rise of AlphaFold[181]. Enzyme kinetic modelling based on the traditional Michaelis and Menten´s equation led the way for other developments, with contributions to accommodate for deviations from the classical behaviours such as feedback inhibition[182], two-substrate reactions[183] and substrate specificity[184].

Nowadays, mechanism-based fitting through computer simulation allows researchers to create kinetic models that fit experimental data[185]. Kinetic modelling has been used for cell metabolism modelling for some time now. This modelling is handy for analysing and rebuilding metabolic networks and designing fermentation processes[186]. Metabolic modelling has been implemented in complex processes like photosynthesis[187] and whole-cell metabolisms[188]. However, examples of its use in cell-free enzymatic cascades are still quite limited[189-192]. The author is confident that enzyme kinetic modelling will continue to develop, and all kinds of enzymatic cascades will benefit from it. Latest advancements in sophisticated computational methods[166] may one day make the rational design of complex cascade reactions possible. In this context, this thesis introduces the first enzyme kinetic modelling of a cell-free OPME cascade for synthesising nucleotide sugars, and it presents a new methodology for working under parameter uncertainty.

3 Materials and Methods

The following sections detail all the experimental and analytical methods used in this work. Three main categories arise, each corresponding to a process development stage: biocatalyst production, reaction engineering, and analytics.

3.1 Biocatalyst Production

3.1.1 Vector Design for Recombinant Enzyme Production

The T7 expression system[193] was generally utilised for recombinant enzyme expression. The *E. coli* BL21(DE3) strain and its variants were obtained from New England Biolabs GmbH. All enzyme gene sequences were optimised for efficient expression in *E. coli* through codon optimisation. Furthermore, these sequences were modified to include a 6x His-tag at the *N*-terminus (unless stated otherwise) and were subsequently incorporated into a bacterial expression vector by BioCat GmbH. A complete list of all bacterial strains created for this thesis is shown in Table 13 in the appendix.

The enzymes were selected from a database screen (Uniprot, BRENDA), taking into account the following characteristics as criteria for selection: bacterial origin to facilitate recombinant expression in *E. coli*, highly annotated and, if available, experimentally validated entries and compatible reaction conditions and cofactors. A comprehensive list of all enzymes used in this thesis is shown in Table 5.

Table 5 - Enzymes used in this thesis. All protein sequences were retrieved from the online database Uniprot [151]. Bacterial organisms were preferred for the selection of genes.

Gene	Enzyme	EC #	Organism	Reaction	Ref.
tagD	Glycerol-3-phosphate cytidylyltransferase	2.7.7.39	*Bacillus subtilis*	Gly3P + CTP → CDP-glycerol + PPi	[194]
UDK	Uridine kinase	2.7.1.48	*Escherichia coli*	Cytidine + ATP → CMP + ADP	[195]
URA6	UMP-CMP kinase 3	2.7.4.14	*Arabidopsis thaliana*	CMP + ATP → CDP + ADP	[196]
PPK3	Polyphosphate kinase 3	2.7.4.1	*Ruegeria pomeyori*	NDP + PolyP(n) → NTP + PolyP(n-1)	[197]
glpK	Glycerol kinase	2.7.1.30	*Thermococcus kodakarensis*	glycerol + ATP → Gly3P + ADP	[198]
PPA	Inorganic pyrophosphatase	3.6.1.1	*Pasteurella multocida*	PPi → 2Pi	

Gene	Enzyme	EC #	Organism	Reaction	Ref.
AGE	*N*-acylglucosamine 2-epimerase	5.1.3.8	*Anabaena variabilis*	*N*-acyl-D-glucosamine ⟶ *N*-acyl-D-mannosamine	[199]
NANA	*N*-acetylneuraminate lyase	4.1.3.3	*Pasteurella multocida*	*N*-acetylneuraminate ⟶*N*-acetyl-D-mannosamine + pyruvate	[200]
CSS	*N*-acylneuraminate cytidylyltransferase	2.7.7.43	*Neisseria meningitidis*	*N*-acylneuraminate + CTP ⟶ CMP-*N*-acyl-β-neuraminate + pyrophosphate	[201]
FKP	Bifunctional L-fucokinase/L-fucose-1-P guanylyltransferase	2.7.1.52 2.7.7.30	*Bacteroides fragilis*	ATP + L-fucose ⟶ADP + β -L-fucose-1-phosphate GTP+ beta-L-fucose-1-phosphate ⟶pyrophosphate + GDP-L-fucose	[202]
GMK	Guanylate kinase	2.7.4.8	*Escherichia coli*	ATP + GMP ⟶ADP + GDP	[203]
Hp3/4FT	α1-3/4 fucosyltransferase	2.4.1.152	*Helicobacter pylori*	GDP-fucose + ROH acceptor⟶GDP + Fucα1-3/4OR	[30]

EC# - enzyme commission number; Gly3P – glycerol-3-phosphate; CTP – cytidine triphosphate, CDP – cytidine diphosphate; PPi – pyrophosphate; ATP – adenosine triphosphate; CMP – cytosine monophosphate; ADP – adenosine diphosphate; NTP – nucleotide triphosphate; NDP – nucleotide diphosphate; PolyP – polyphosphate; Pi – pyrophosphate.

3.1.2 pDuet Vector Expression System

For the co-expression of multiple genes in a single bacterial host, the pDuet vector expression system from Novagen was used[158]. This system comprises several vectors with compatible replication origins and different antibiotic resistance genes for selection. The T7 expression system is used, making them compatible with many *E. coli* strains in the market. Two genes can be inserted per vector, and 4 vectors can be simultaneously transformed in one host strain for up to 8 co-expressed recombinant proteins per strain. The pDuet vectors used in this thesis were pDuet-CDF, pDuet-CYC, pDuet-pET and pDuet-RSF. A detailed list of the vector constructs used in this work can be found Table 13 in the Appendix.

3.1.3 Bacterial Transformation

The high-efficiency transformation protocol employed was based on the New England Biolabs recommended protocol[204]. It consisted of the following steps: 1) Thaw a tube of competent *E. coli* cells on ice for 10 min. 2) Add 1-5 µl containing 1 pg to 100 ng of plasmid DNA to the cell mixture. 3) Carefully flick the tube 4-5 times to mix cells and DNA. 4) Place the mixture on ice for 30 min. 5) Heat shocks the cells at 42 °C for 10 sec. 6) Place on ice for 5 minutes. 7) Add 950 µl of room temperature super optimal broth (SOC) medium into the mixture. 8) Place at 37 °C for 60 min with shaking at 500 rpm in a Thermoblock. 9) Warm selection plates to 37 °C. 10) Spread 100 µL of the mixture into a selection plate. The concentration of antibiotics used for single plasmid transformations were as follows: 34 µg/mL for chloramphenicol-resistant plasmids, 50 µg/mL for spectinomycin and kanamycin-resistant plasmids and 100 µg/mL for ampicillin-resistant plasmids. 11) Centrifuge at 6,000 x g for 20 min, discard the supernatant, resuspend in 100 mL Ultrapure water and spread on a second selection plate. 12) Incubate overnight at 37 °C. The same protocol was employed for the bacterial transformation of pDuet vectors, but with two modifications based on Novagen´s user protocol. The amount of antibiotic used was halved to transform multiple vectors, and the vectors were introduced to the host either simultaneously or sequentially, with an intermediate step to create competent cells based on standard techniques[205].

To produce cryo-cultures, a single colony from the selection plate was used to innoculate 10 mL of LB media and incubated at 37 °C until OD_{600} reached a value of 0.8-1; it was then centrifugated at 6,000 x g and 4 °C for 15 min, the supernatant was discarded, and the cell pellet was resuspended in 2 mL Slant media:glycerol 50:50. This was enough to produce 4 x 0.5 mL cryo-stock aliquots that were then stored at -80 °C for later use. The master cell bank consisted of these four cryocultures, from which working cell banks were created.

3.1.4 Fermentation

An inoculation loop was used to pick biomass from a cryo-stock and used to inoculate 25 mL of TB media (Terrific Broth: yeast extract, 24 g/L; tryptone, 12 g/L; glycerol, 4 g/L; KH_2PO_4, 2.3 g/L; K_2HPO_4, 16.4 g/L) in a 200 mL baffled shake flask supplemented with the appropriate antibiotic(s), as described in Table 13. Only isolated colonies were selected when picking biomass from a freshly grown transformation agar plate. When picking biomass from a cryo-stock, the stock quickly returned to -80 °C storage before it could thaw. A cryo-stock was discarded if it stayed outside storage long enough to melt. The pre-culture was incubated at 37 °C and 150 rpm until OD600 reached a value of 0.8 – 1. At this point, 10 mL of

culture could be taken to produce new cryo-cultures of the strain, and the rest could be used to inoculate a main culture.

The main culture consisted of 200 mL TB media in a 1 L baffled shake flask, supplemented with the appropriate antibiotic(s) and inoculated with 2.5% v/v from a preculture grown overnight or until an OD600 between 0.8–1 was observed. The main culture was incubated at 37 °C and 150 rpm until an optical density (OD600) of 0.8–1.0 was observed. Protein expression was induced at this point by adding 0.4 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by incubation at 16 °C for 16–18 h. Biomass was harvested by centrifugation at 6,000 x g for 20 min, the supernatant was discarded, and the cell pellet was either processed immediately or stored at -20 °C.

The micro-fermentation experiments were carried out using a Biolector system from Beckman Coulter coupled with the RoboLector for liquid handling tasks. Parallel fermentations were done with a working volume of 1 mL in 48-well plates. The Biomass filter Ex 620 nm (bandpass: 10 nm / scattered light measurement) was used to monitor biomass growth.

3.1.5 Downstream processing and formulation

A cell pellet was resuspended in ~10 mL lysis/equilibration buffer (MOPS (pH 7.5), 50 mM; NaCl, 500 mM; imidazole, 10 mM; glycerol, 5% v/v) per gram of wet biomass. Cell lysis was achieved by high-pressure homogenisation using a Maximator® HPL6 device. The biomass was continuously run through the device at 1,000 bar and 4 °C for 5-7 min or until the mixture changed from a thick pale white to a more transparent shade. After cell lysis, cell debris was separated by centrifugation at 12,000 x g and 4 °C for 30 min. The supernatant was recovered and filtered through 5 µm and 0.45 µm filters in preparation for Ion Metal Affinity Chromatography (IMAC).

IMAC was performed with an Äkta start® device and 5 mL His-Trap HP columns from Cytiva®. Buffers used were a lysis/equilibration buffer or “buffer A” and an elution buffer or “buffer B” (MOPS (pH 7.5), 50 mM; NaCl, 500 mM; imidazole, 300 mM; $MgCl_2$, 10 mM; glycerol, 5% v/v). The chromatographic steps were as follows: column equilibration, 100% buffer A for 10 CV; sample application, 50 mL of cell lysate; wash, 10% buffer B for 10 CV; elution, 100% buffer B for 5 CV. Samples for SDS-PAGE analysis were collected at each step of the process using a carousel fractionator, and the elution step was recovered in 5 mL fractions, which were later pooled according to the UV detector peaks observed in the chromatogram. This general chromatographic method was used for all protein purifications performed for this thesis, except for a method developed later to enhance the purity of the enzyme Bifunctional L-fucokinase/L-fucose-1-P guanylyltransferase (FKP). This method consisted of the following steps: column equilibration, 100% buffer A for 10 CV; sample application, 50 to 100 mL of cell lysate; wash, 16% buffer B for 10 CV; elution, buffer B linear gradient from 16 to 50% during 5 CV followed by 5 CV at 100%.

After IMAC, the pooled elution fractions were buffer-exchanged to remove imidazole (exchange buffer: MOPS (pH 7.5), 50 mM; NaCl, 300 mM; $MgCl_2$, 10 mM). This was achieved by using Amicon® centrifugal filters with a MWCO value of 10 kDa and 6,000 x g for 20 min per centrifugation step. The final enzyme stocks were stored directly at 4 °C or diluted 1:1 with pure glycerol for long-term storage at -20 °C.

3.2 Reaction Engineering

As many factors can affect the performance of each enzymatic cascade, these were optimised using three independent strategies: DoE, kinetic modelling, and exploratory experimentation. This section describes the methodology utilised for each strategy.

3.2.1 Kinetic Modelling Under Uncertainty

In collaboration with the Analysis and Redesign of Biological Networks research group at the Max Planck Institute for the Dynamics of Complex Technical Systems in Magdeburg, Germany, Nicolas Huber performed the kinetic model building. The COPASI software version 4.35 (build 258)[206] was used for all simulation, parameter estimation and optimization tasks. A detailed description of the methods and algorithms used can be found in the original publication[207], and all the model files relevant to this work can be found in the following repository: https://github.com/klamt-lab/GDP-fucose_OptMet2023_Paper.

3.2.2 Design of Experiments

The software package MODDE® version 13 from Sartorius was used for all designs and analyses. First, the factors to be explored within the optimisation round were chosen. These varied from three factors, in the case of the design used for CDP-glycerol poduction, to up to nine factors, as in the design used for GDP-fucose production. The number of factors in the design directly affects its complexity, but also increases the chance of a more realistic depiction of the process. The number of factors and levels is directly related to the number of experiments required to obtain significant data, which is also referred to as the resolution of the design. The higher the resolution of the design, the more effectively one can identify the effects of individual factors and their interactions. Three experimental designs were chosen for this work: Central composite orthogonal (CCO), D-optimal and a fractional factorial scheme with resolution IV. The result of selecting particular levels, also called ranges, for each factor (i.e. temperature from 0 to 45 °C, pH value from 4 to 10…) is the design space. MODDE® constructs a regression model within this design space. After a DoE setup was created, the experiments detailed in the setup were performed at a working volume of 200 µL in either 1.5 mL microtubes or 96-well plates. All the reaction components were mixed, and the corresponding enzymes were added at the end to start the reaction. Incubation was done using an Eppendorf Thermomixer® for shaking and temperature control. Sampling was performed by removing 1-10 µL of the reaction and diluting it with ice-cold Milli-Q water to achieve a concentration of less than 200 µM in all UV-active components. HPLC-UV/PAD was used to measure the response as product generation, which was defined in terms of the UV absorbance for the CDP-glycerol design (lack of an analytic standard prevented product concentration calculations), product concentration for the CMP-Neu5Ac design and product conversion yield for the GDP-fucose design (substrate concentration was introduced as a variable). The results were then analysed using the statistical tools available within the MODDE® software, which consists of three steps: model fit, model evaluation and optimisation.

The model fit was performed using either PLS (Partial Least Squares) or MLR (Multiple Linear Regression) fitting methods. MODDE® uses MLR as the default setting, as it is often employed for more straightforward, interpretable models; PLS is preferred for handling models where multicollinearity and noisy data are expected. The model is then evaluated using four indicators: Q^2 (predictive ability), R^2 (coefficient of determination), model validity (or *lack-of-fit* test), and reproducibility (variation of the response).

Q^2 is a measure of the model's predictive power, computed through a cross-validation procedure. The value of Q^2 is always between 0 and 1, where 1 corresponds to the optimal predictive model. Q^2 tends to underestimate the model's goodness of fit. Inside MODDE®, Q^2 is calculated in the following way:

$$Q^2 = 1 - \frac{PRESS}{SStot} = 1 - \frac{PRESS}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Equation 1 – Calculation of predictive ability (Q^2) of a model within MODDE®.

Where:

- PRESS: prediction residual sum of squares (differs in MLR and PLS).
- SStot: the total sum of squares of y corrected by the mean.
- n: number of observations.
- y: the observed value of the response.
- $\bar{y}$: mean of the response.

R^2 is a measure of the goodness of fit and describes the fraction of the variation of the response explained by the model, its value is always between 0 and 1, where a value of 1 means that the model can fit the data very closely. It is of note that R^2 overestimates the goodness of fit. The value for R^2 is computed in the following way:

$$R^2 = 1 - \frac{SSrec}{SStot} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Equation 2 – Calculation of the coefficient of determination (R^2) of a model within MODDE®.

Where:

- SSrec: sum of squares of the residual corrected by the mean.
- SStot: the total sum of squares of y corrected by the mean.
- n: number of observations.
- y: the observed value of the response.
- $\bar{y}$: mean of the response.
- $\hat{y}$: predicted value of the response of the regression model.

Model validity is computed inside MODDE® as a *lack-of-fit* hypothesis test. A value above 0.25 means that the model has no lack-of-fit, while a value below 0.25 means that the model error is significantly larger than the pure error, which can be

due to significant model problems such as outliers, transformation problems or errors in the design and model fit. The model validity is calculated in the following way:

$$Validity = 1 + 0.57647 log_{10}(p_{lof})$$

Equation 3 – Calculation of validity of a model within MODDE®.

Where:

- 0.57647: Constant value so that $p_{lof} \geq 0.05$ will give a validity ≥ 0.25.
- p_{lof}: p-value for the *lack-of-fit* test.

Reproducibility is the last indicator and it evaluates the variation of the response at the same conditions. A value of 1 means that the model has perfect reproducibility. Inside MODDE® the reproducibility is calculated in the following way:

$$Reproducibility = 1 - MS_{pre}/MS_{tot}$$

Equation 4 – Calculation of reproducibility of a model within MODDE®.

Where:

- MS_{pre}: The mean square of the pure error.
- MS_{tot}: The total mean square of Y.

The model has to be adjusted and refined to satisfy these indicators before it can be used for prediction and optimisation. Outliers have to be excluded so that the variation of repeated experiments is less than the overall variation of the response. A transformation can be performed on the response if it has a non-normal distribution, a normal distribution will generally give better estimates. A log transformation is often preferable for positive skewness and a negative log transformation for a negative skewness of the distribution. Finally, the coefficient plot can be used to identify and exclude non-significant terms, in doing so, Q2 can be maximised.

Once the model has been adjusted and evaluated, it can be used for prediction and optimisation tasks. Inside MODDE®, optimisation is carried out using the downhill simplex method by Nelder and Mead[208].

3.3 Analytics

3.3.1 Protein Analytics

Protein purity and identification were determined with SDS-PAGE using the Bio-Rad protein electrophoresis devices (chambers and power supply) and pre-cast gels with 10-12.5% acrylamide. For molecular weight marking, the PageRuler® pre-stained protein ladder, 10-180 kDa, from Thermo Scientific was used. Electrophoresis running conditions were 200 V and 45min. The measurement of protein concentrations was performed using the Pierce™ BCA Protein Assay kits from Thermo Scientific and a Tecan infinite 200 plate reader device.

3.3.2 HPLC

High-Performance Liquid Chromatography with UV and Pulsed Amperometry Detection (HPLC UV/PAD) was used to measure compounds like nucleotide sugars and human milk oligosaccharides. A Dionex IC5000 system from Thermo Scientific was equipped with a CarboPac™ PA200 analytical (3 mm x 250 mm) and guard (3 mm x 50 mm) columns as the stationary phase, and a gradient of 1 M NaOAc was used as the mobile phase. An injection volume of 25 µL was chosen, and the eluent gradients were optimised depending on the components expected to be measured, see. In total, gradients for four different use cases were developed to measure the compounds involved in the synthesis of a) CDP-glycerol, b) CMP-Neu5Ac, c) GDP-fucose and d) LNFP III (Figure 6). Analytical standards containing each expected compound from a synthesis reaction were prepared in concentrations ranging from 2-200 µM to allow for the quantification of each component. Detection through UV absorbance at 260 nm could analyse nucleotide sugars, while HPLC-PAD was used for HMOs. The reason is that nucleotides show absorption at 260 nm, but HMOs do not. The workflow for Pulsed Amperometry Detection is the same as for UV detection, with the modifications that an electrochemical detector is used instead of a UV detector, and a different elution gradient was developed for the analysis of HMOs.

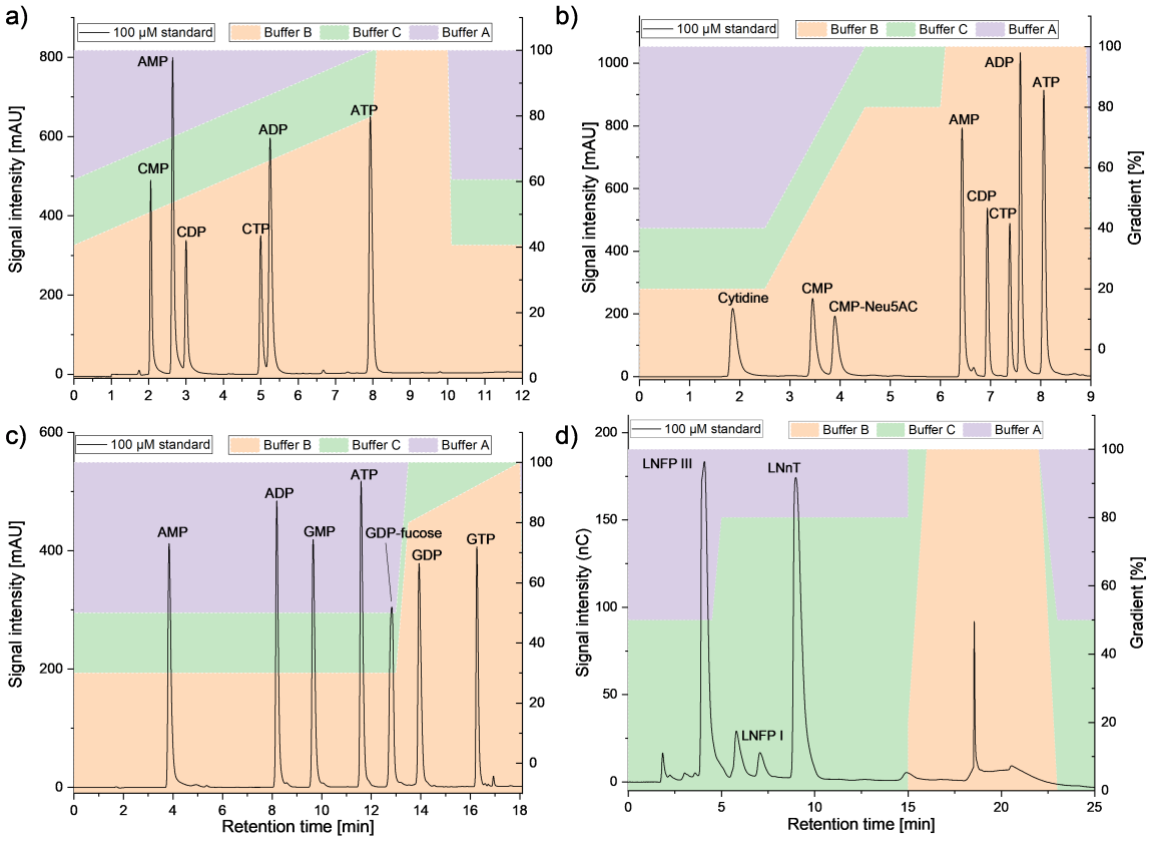


Figure 6 - HPLC-UV/PAD elution gradients. Four elution gradients were developed to analyse a) CDP-glycerol, b) CMP-Neu5Ac, c) GDP-fucose and d) HMOs. Buffer A (lilac) was Mili-Q Water, Buffer B (peach) was 1 M NaOAc and Buffer C (sap green) was 100 mM NaOH. Gradient D was measured using an electrochemical detector (ED-PAD) in contrast to gradients A, B and C, which were measured using a UV detector.

At the time of writing, CDP-glycerol was not available for purchase as an analytical standard from any chemical suppliers. For this reason, CDP-glycerol had to be identified by mass spectrometry, and its concentration in each sample was indirectly calculated from the mass balance of the other cytidine-containing components.

3.3.3 MALDI-TOF/TOF MS

Matrix-assisted laser Desorption/Ionization Time of Flight/Time of Flight Mass Spectrometry (MALDI-TOF/TOF MS) was carried out by the Bio/process analytics group at the Max Planck Institute for the Dynamics of Technical Complex Systems. The measurement was carried out using an utra-fleXtreme system from Brucker Daltonics. On an MTP AnchorChip 800/384 TF MALDI target, 1 µL 9-aminoacridine (9-AA, 10 mg/mL in acetone) was spotted, and then 1 µL of the sample was applied to the dried matrix layer. The samples were recrystallised by adding pure ethanol. A reflectron negative ion mode and a laser-induced fragmentation and transfer (LIFT) negative ion mode were used for the measurements.

4 Results and Discussion

4.1 CDP-glycerol

This chapter describes the development of a new enzymatic cascade for the synthesis of CDP-glycerol from glycerol, cytidine, ATP, and polyphosphate (PolyP). The cascade design is presented first, followed by the results of recombinant enzyme expression and proof-of-concept experiments. Identifying CDP-glycerol was necessary because obtaining an analytical standard was impossible at the time of writing. Once the synthesis of CDP-glycerol was achieved, further steps were taken to improve the cascade, such as evaluating the effect of adding inorganic phosphatase to the cascade and performing a DoE screen to identify optimal operating reaction conditions. Some of the content in this chapter has been published previously[94], in particular, figures 9-18 were reproduced from the original publication under the open-access license CC BY-NC-ND 4.0.

4.1.1 A Cell-free Enzymatic Cascade for the Synthesis of CDP-glycerol

A novel enzymatic cascade was developed with six enzymes and seven catalytic steps. The design of enzymatic cascades for synthesising sugar nucleotides was inspired by the nucleotide sugar salvage pathway in nature and supported by PolyP-driven ATP regeneration. The catalytic steps of the cascade are presented in Figure 7 and are as follows: Phosphorylation of cytidine to form CMP by an ATP-dependent cytidine kinase (UDK). Phosphorylation of CMP to form CDP by an ATP-dependent CMP kinase (URA6). In addition, phosphorylation of glycerol to Gly-3P by an ATP-dependent glycerol kinase (glpK). Phosphorylation of CDP to CTP by the PolyP kinase PPK3, which phosphorylates also ADP to ATP thanks to its innate substrate promiscuity. Transfer of the cytidylyl moiety to glycerol-3-phosphate (Gly-3P) by a Gly-3P cytidylyltransferase (tagD) to produce CDP-glycerol. The side-product pyrophosphate is degraded into monophosphates by an inorganic pyrophosphatase (PPA).

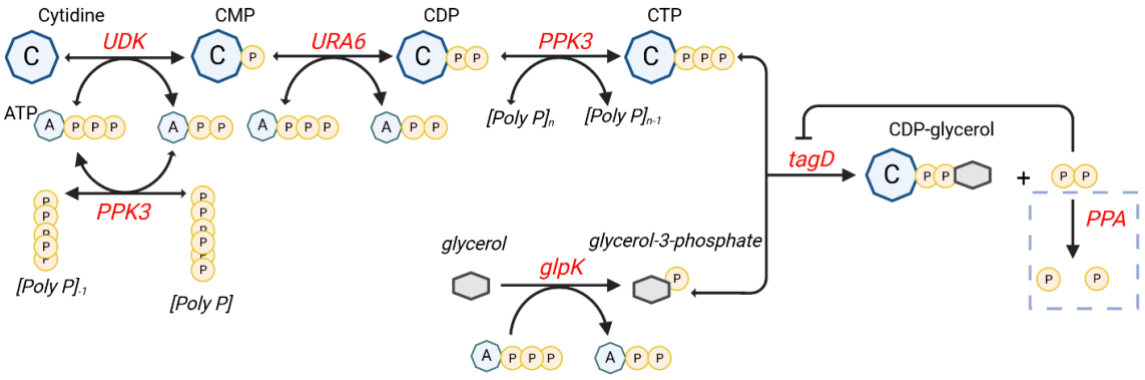


Figure 7 - Enzymatic cascade for the synthesis of CDP-glycerol. Enzymes are depicted in red and components in black; enzymatic reactions are shown in black arrows weighted by size for preferred reaction direction when reversible. A dashed box represents a non-essential part of the cascade.

In situ regeneration of ATP from PolyP using PolyP kinases has been established before[32, 68, 71, 73], a method which is also suited for this design.

Additionally, pyrophosphate, a co-product of the CDP-glycerol synthesis step, can potentially inhibit tagD. To counteract this inhibition, PPA can be introduced into the cascade reaction to degrade pyrophosphate into monophosphate, theoretically driving the cascade reaction toward the product side[194]. The cascade design originally comprised six enzymes, but it was reduced to five after finding out that the addition of PPA did not significantly improve the reaction, data discussed in section 4.1.2.4.

4.1.2 Results

The work required for the synthesis of CDP-glycerol began with the recombinant expression and purification (See section 3.1) of the six enzymes conforming to the cascade reaction. After the enzymes were successfully produced, a proof-of-concept experiment was conducted to demonstrate the synthesis of CDP glycerol with only two enzymes. CDP-glycerol was then identified through mass spectrometry. The successful CDP-glycerol synthesis using the proposed 6-enzymatic cascade required several exploratory experiments. Finally, this enzymatic cascade was improved using a DoE screening setup to determine the optimal reaction conditions.

4.1.2.1 Individually Purified Enzymes

Six different enzymes were produced to serve as biocatalysts for synthesising CDP-glycerol. Each enzyme (see Table 2) was produced using the same fermentation and purification procedures, and therefore, each was obtained in different concentrations and purities. The results of the evaluation of each enzyme are presented in this section.

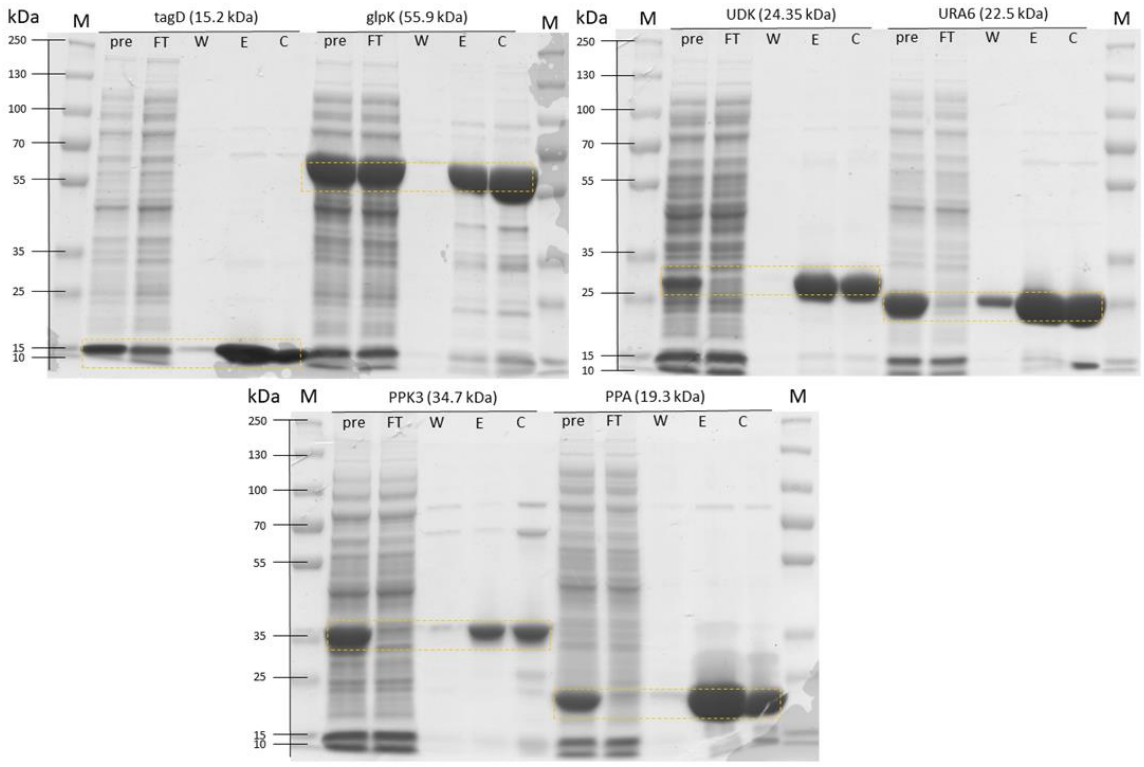


Figure 8 - SDS-PAGE showing all enzymes for the CDP-glycerol cascade. For each enzyme, metal affinity chromatography fractions are shown. kDa – kilodalton; M - Molecular weight

marker; tagD, glpK, UDK, URA6, PPK3 and PPA – enzymes; pre – cell lysate fraction; FT – flowthrough fraction; W – wash fraction; E – elution fraction; C – concentrate fraction (10 kDa MWCO Amicon centrifugal filtered). Between 1-2 µg of total protein was loaded into each 10% polyacrylamide gel well. The gel was stained with Coomassie blue and scanned using a tabletop Epson scanner.

After recombinant expression of each enzyme in *E. coli,* protein purification through IMAC, and concentration, the enzyme stocks were analysed by SDS-PAGE. A total amount of protein between 1-2 µg was loaded in each well. It can be observed in Figure 8 that each enzyme could be identified at its expected theoretical molecular weight. The molecular weight for each enzyme is as follows: tagD –15.27 kDa, glpK –55.9 kDa, PPK3 –34.74 kDa, URA6 –22.48 kDa, UDK – 24.35 kDa and PPA –19.31 kDa. It is of importance to remark that each enzyme carries a 6x His-tag which increases its molecular weight by 0.93 kDa.

After having verified the expression and purification of the enzymes, protein concentration was measured using a BCA assay with a BSA standard as a reference. The results of enzyme concentration in each of the enzymatic stocks are shown in Table 6, along with a comparison to values found in recent literature. PPA was the enzyme that could be obtained in the highest quantity per volume of bacterial culture. The enzyme with the lowest protein yield per bacterial culture volume was UDK.

Table 6 - Summary of purified enzyme stocks for CDP-glycerol cascade. The highest amount of total purified target protein obtained in this work is compared with that reported in other works in the literature.

Enzyme	Total target protein in this work [mg]	Bacterial culture volume [mL]	Total target protein previously reported [mg]	Bacterial culture volume [mL]
tagD	28	200	46.3 [209], 10-15 [210]	1,000
glpK	12.1	200	5, 9.2 [211, 212]	1,000
PPK3	14	200	57.85 [71]	500
URA6	31.5	200	42.84 [71]	500
UDK	11.2	200	0.6 [213]	10,000
PPA	61.7	200	53 [73]	500

*tagD – glycerol-3-phosphate cytidylyltransferase; glpK – glycerol kinase; PPK3 – polyphosphate kinase 3; URA6 – UMP-CMP kinase 3; UDK – uridine kinase; PPA – inorganic pyrophosphatase.

4.1.2.2 Proof-of-Concept

CDP-glycerol synthesis was first tested starting from CTP and glycerol and using only two enzymes. This experiment aimed to prove that the CDP-glycerol synthesis

part of the cascade was functional before adding the cytidine phosphorylation route and the ATP regeneration system. The reaction was set up with 15 mM ATP, 15 mM CTP, 1.5 M glycerol (calculated from the amount present in the enzyme stocks), 0.1 g/L tagD, 0.1 g/L glpK, 50 mM Tris/HCl (pH 8) and 30 mM $MgCl_2$ at 37 °C and 500 rpm for 18 h. An HPLC-UV chromatogram showing the compounds present at time zero (before adding enzymes) and after 18 h of reaction is shown in Figure 9.

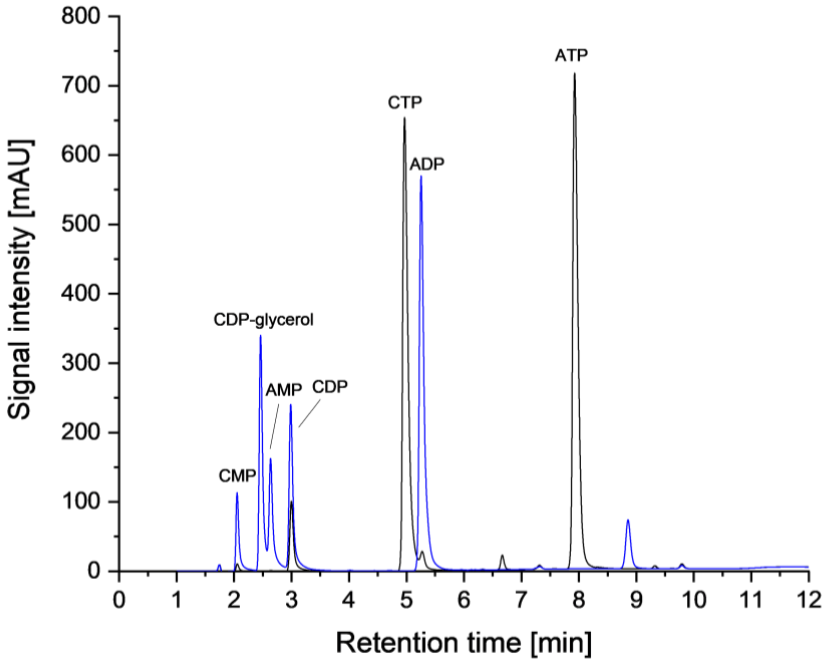


Figure 9 - Synthesis of CDP-glycerol from CTP and glycerol. HPAEC-UV chromatogram of the enzymatic cascade at start and after 18 h of reaction. In black: time zero, in blue: 18 h of incubation. The enzymes tagD and glpK were used as biocatalysts. Species were identified by their retention time in comparison to pure analytical standards.

According to the compounds observed in the chromatogram, the initial substrates, ATP and CTP, are fully consumed after 18 h of reaction. The resulting compounds could be identified as CMP, CDP-glycerol, AMP, CDP and ADP.

4.1.2.3 Identification of CDP-glycerol

To identify the synthesis of CDP-glycerol from the proof-of-concept experiments, MALDI-TOF MS/MS was used. The results from this analysis are shown in Figure 10 and in Figure 11. CDP-glycerol could be identified by its molecular weight at 476.05 m/z. Because the sample used for this analysis came from a reaction mixture, other species could also be identified, such as CMP, AMP, CDP, ADP, and, according to its m/z, what seems to be ADP-glycerol.

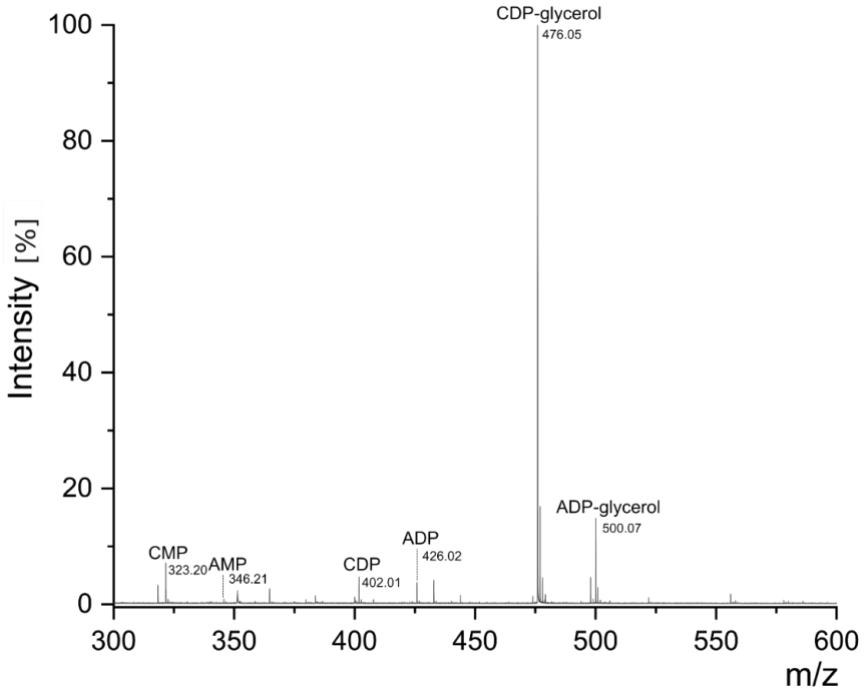


Figure 10 - MALDI-TOF-MS of the cascade reaction for the synthesis of CDP-glycerol. Sample taken from the cascade reaction performed using only the enzymes tagD and glpK; for the measurement, 1 μL of 9-aminoacridine (10 mg/L) was used as a matrix; [M-H](-) ions were detected in reflectron negative-ion mode.

Interestingly, a compound with an m/z of 500.07 was observed, likely to be the side product ADP-glycerol. Although no available literature reports that ADP-glycerol is a product of any of the enzymes used in this cascade, it is the author's hypothesis that the enzyme tagD could accept not only CTP but also ATP as a substrate and produce ADP-glycerol.

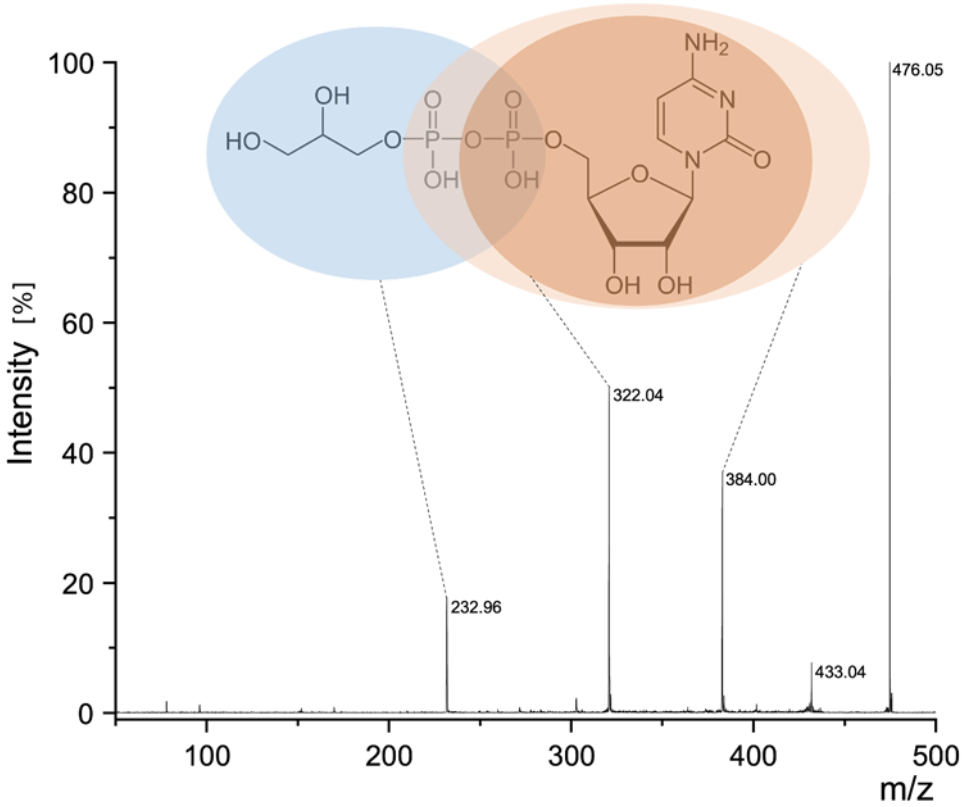


Figure 11 - MALDI-TOF/TOF-MS of CDP-glycerol. 9-AA was used as a matrix and [M-H](-) ions were detected in LIFT negative ion mode. The graphical representation of CDP-glycerol depicts the main fragment ions. Three distinct fractions were identified as components of the CDP-glycerol molecule: 1) in blue-glycerol diphosphate, 2) in orange-cytidine phosphate and 3) in light orange-cytidine diphosphate.

The precursor ion corresponding to CDP-glycerol was further analysed by MS/MS to confirm its identity through fragmentation. In Figure 11, three signals corresponding to each fragment were observed, resulting from a bond break at each of the molecule's phosphate sites.

4.1.2.4 Reaction Engineering

The synthesis of CDP-glycerol was successful during the proof-of-concept experiments using only the enzymes tagD and glpK; however, no CDP-glycerol could be observed when the cascade was extended with UDK, URA6, PPK3 and PPA (Figure 12b). Storage of enzyme stocks in 50% glycerol is a common practice for cryo-protection at temperatures of -20 °C. Because glycerol is a substrate for the enzyme glpK, it was hypothesised that an excess of it could be the reason behind the cascade reaction reaching only the CMP intermediate. Therefore, new enzyme stocks were freshly prepared and used as biocatalysts without adding glycerol for storage to test this. The conditions for the reaction were the same as before: 3 mM ATP, 10 mM cytidine, 4 mM PolyP, 10 mM glycerol, 0.02 g/L UDK, 0.05 g/L URA6, 0.016 g/L PPK3, 0.05 g/L tagD, 0.02 g/L gplK, 0.05 g/l PPA, 50 mM Tris/HCl (pH 8) and 30 mM $MgCl_2$. The chromatogram of the measurement at time zero and after 18 h of reaction is shown in Figure 12a. CDP-glycerol synthesis could be observed in the reaction using the enzyme stocks without added glycerol, along with what could be ADP-glycerol.

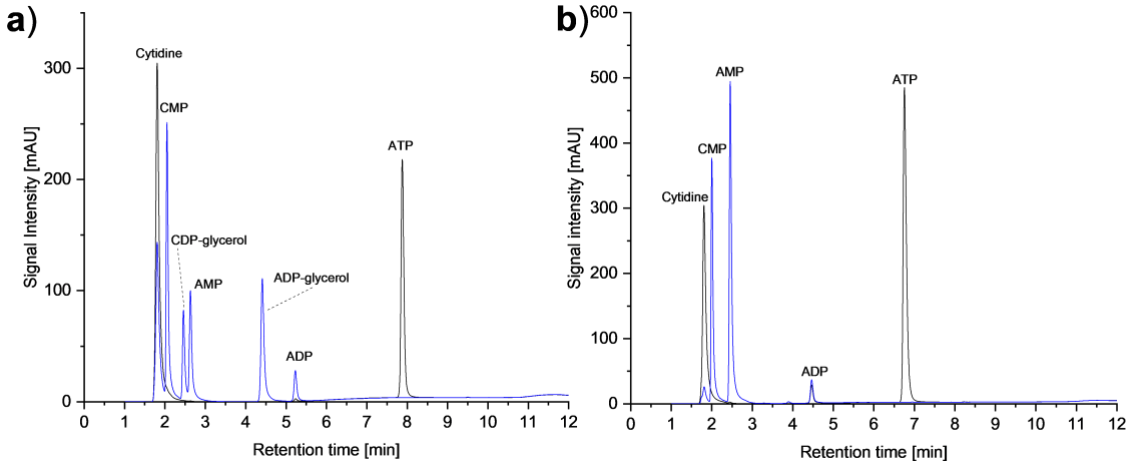


Figure 12 - Complete enzymatic cascade for CDP-glycerol synthesis. The reaction used tagD, glpK, UDK, URA6, PPK3, and PPA catalysts. Two reactions are shown in which the enzymes were stored: a) without glycerol and b) in 50% glycerol. In black, time zero; in blue, 18 h of incubation.

Another characteristic of the original cascade design tested was the addition of the enzyme PPA, which was intended to drive the cascade forward by degrading pyrophosphate, the co-product of tagD. To test this, a cascade reaction was set up with the following conditions: 7 mM ATP, 44 mM cytidine, 8 mM PolyP, 44 mM glycerol, 1 g/L tagD, 0.1 g/L glpK, 0.1 g/L UDK 0.1 g/L URA6, 0.5 g/L PPK3, 0.5 g/L PPA, 100 mM Tris/HCl (pH 8) and 45 mM $MgCl_2$.

The cascade reaction was monitored for 25 h, and the time course graphs of the components measured by HPLC-UV are shown in Figure 13. For the reaction set up without PPA, shown in a) and b), the concentration of CDP-glycerol reached ~40 mM, corresponding to a cytidine conversion yield of 70%. However, when PPA was added the concentration of CDP-glycerol reached 30 mM, equivalent to a 50% cytidine conversion yield. Not only was no improvement observed, but the contrary appeared to be the case. Although the rest of the components are remarkably consistent between both assays, it can be observed that there is a higher accumulation of CMP when PPA is added.

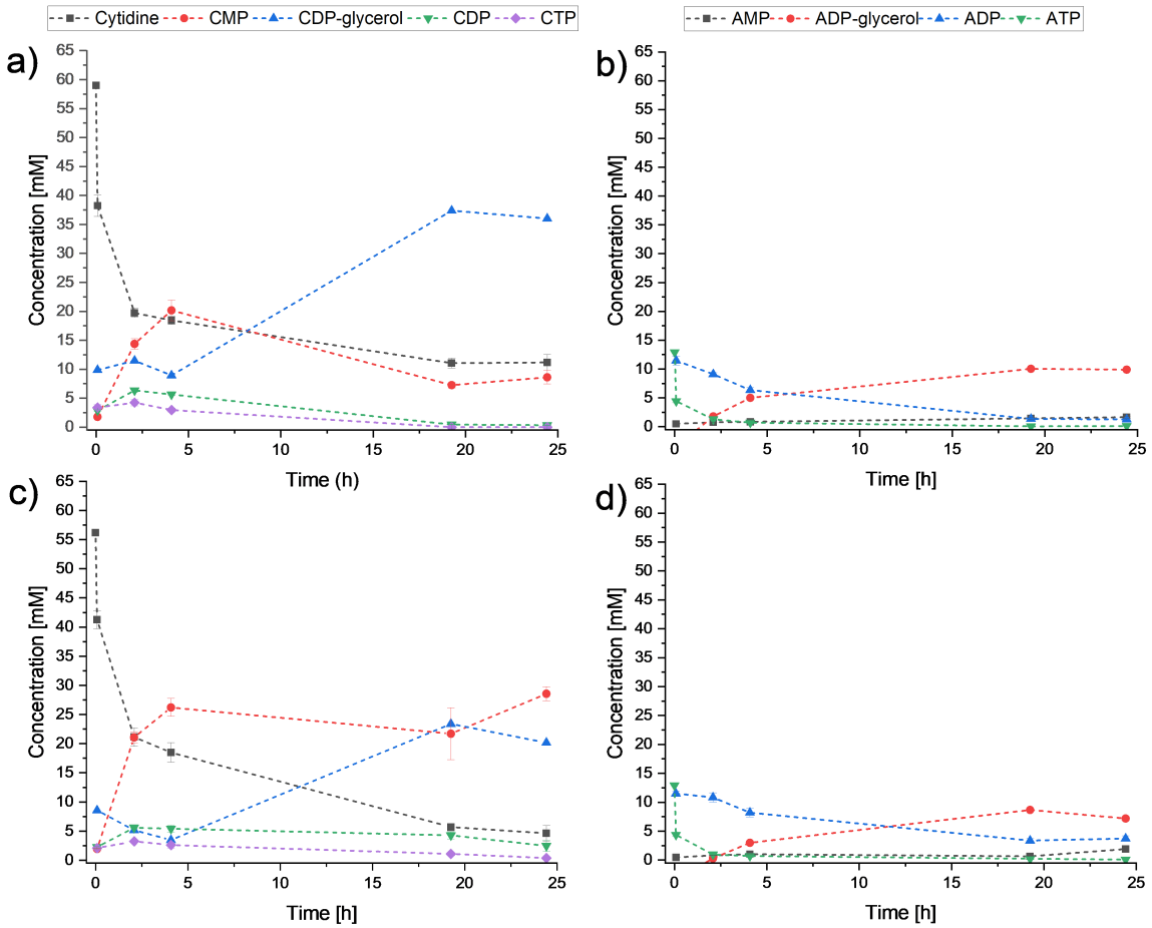


Figure 13 - Effect of PPA addition on the cascade for CDP-glycerol synthesis. Time course measurements using HPLC-UV. Cascade reaction without PPA: a) cytidine-containing compounds and b) adenosine-containing compounds. Cascade reaction with PPA: c) cytidine-containing compounds and d) adenosine-containing compounds. Cascade reactions were performed in biological triplicate; error bars show standard deviation.

4.1.2.5 Design of Experiments

A DoE screening was set up to identify the optimal temperature, pH and $MgCl_2$ concentrations for the enzymatic cascade. A central-composite fractional-factorial design was chosen to explore the experimental space with a temperature 16-49 °C, pH value 6.7-8.6 and $MgCl_2$ concentration 0-116 mM. 51 independent experimental runs were performed, including biological triplicates for each condition and nine replicates for the centre point. The average peak area of the CDP-glycerol peak measured by HPLC-UV [mAU*min] was chosen as a response. This was done to avoid errors when quantifying a concentration by mass balance. The specific conditions and responses for each run are listed in Table 14 of the Appendix. No transformation was done to the response distribution, and one outlier was found with an outlier threshold of 3 standard deviations (σ).

After performing the experimental design and measuring the responses, data was analysed using the software MODDE® to assess the model's quality and identify significant effects. A summary of this analysis is shown in Figure 14, a) represents the magnitude of each factor as green bars; these can be positive or

negative depending on whether the effect increases or decreases the response. The bars illustrate the significance of the response, and the quadratic terms are included to identify non-linear effects; that is, a significant quadratic effect indicates curvature in the response (suggesting an optimum rather than a constant linear relationship). The fit summary graph in b) shows four criteria that MODDE® defines as important: R^2, Q^2, model validity and reproducibility. The model was then adjusted by excluding outliers, and the criteria for a useful model were satisfied.

For this design, the statistical analysis shows that pH has the highest impact, followed by temperature. It also identified a significant non-linear effect of temperature. The software calculated the model fit criteria as R^2 of 0.51, Q^2 of 0.30, model validity of 0.299, and reproducibility of 0.49. With these values, the model fit can be accepted.

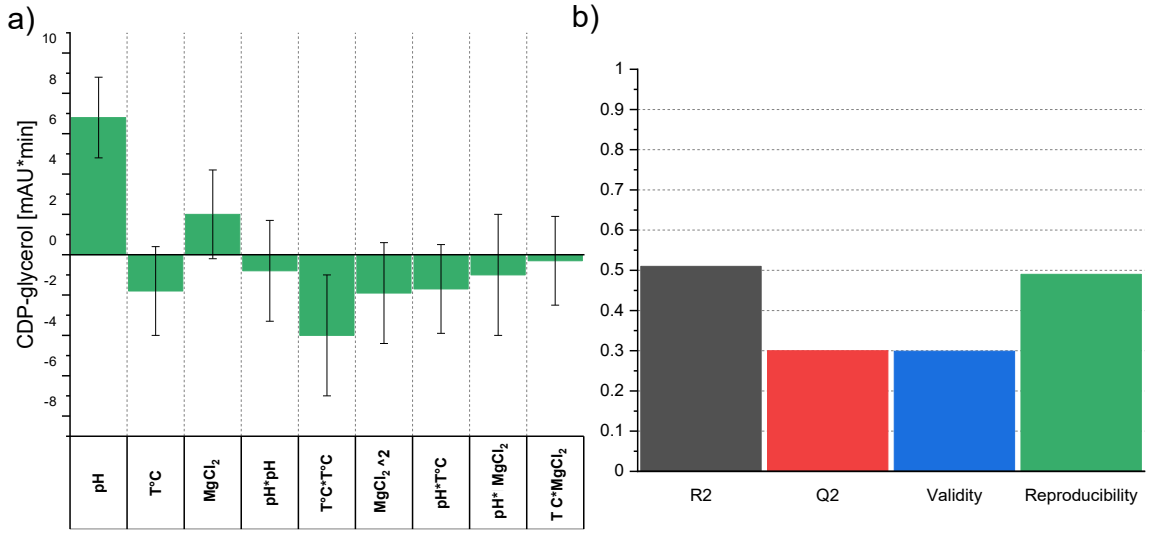


Figure 14 - DoE model evaluation for the CDP-glycerol enzymatic cascade. Sensitivity analysis and model evaluation. a) Effect graph for each factor. Green bars represent the magnitude of a positive or negative influence on the response, while the error bars represent the significance of the response. Single-factor terms, interaction, and quadratic terms are included for each factor. b) Summary of fit. R^2: goodness of fit; Q^2: prediction precision; model validity: a test for diverse model problems; reproducibility: variation within replicates. N=50, DF=40. Model fit and analysis were performed by the built-in tools in the software package MODDE® 13.

The built-in optimiser could then be used to estimate predictions and identify the optimal conditions for the reaction to yield a higher response of CDP-glycerol synthesis. The values found by the optimiser were a pH of 8.76, and a temperature of 29.2 °C and a $MgCl_2$ concentration of 58.52 mM. In Figure 15, a 4D contour heat map is displayed, illustrating the predicted response space. The optimal conditions predicted a peak area value of 29 mAU/min.

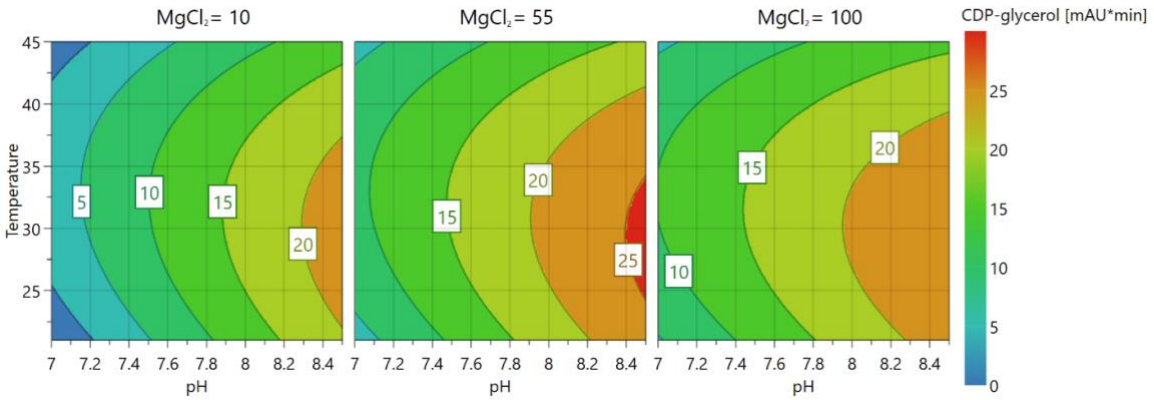


Figure 15 - Graphic visualisation of the predicted CDP-glycerol yield. Response 4D contour plot generated with MODDE®. CDP-glycerol synthesis is represented as a heat map with changing temperature, pH value and $MgCl_2$ concentrations.

The optimal conditions were experimentally validated with the following conditions: 100 mM Tris/HCl (pH 8.7), 58.5 mM $Mgcl_2$, 7 mM ATP, 35 mM cytidine, 8 mM PolyP, 35 mM glycerol, 1 g/L tagD, 0.1 g/L gplK, 0.1 g/l UDK, 0.1 g/L URA6, 0.5 g/L PPK3 at 29 °C with sampling over 24 h. The time course of the concentrations of the components is shown in Figure 16.

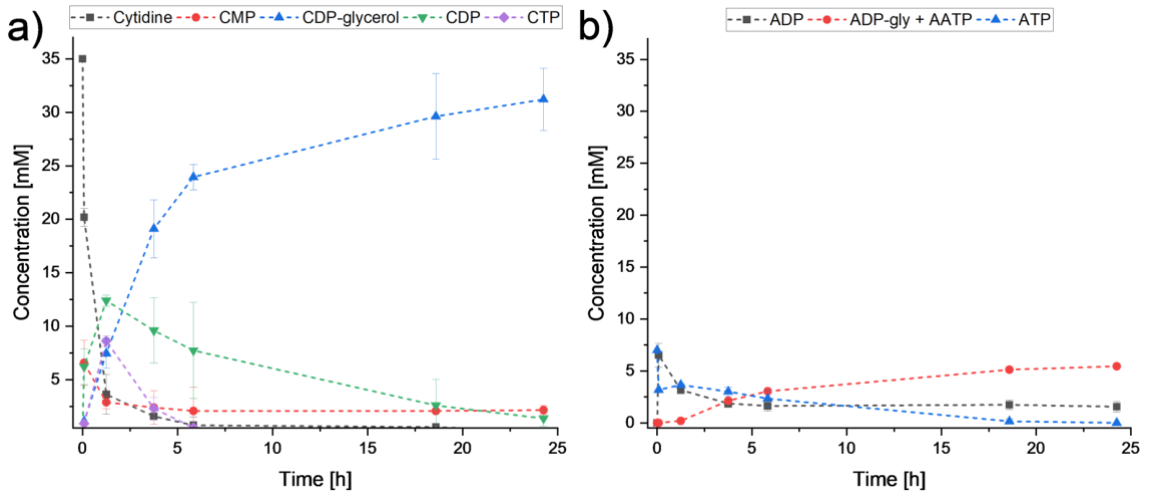


Figure 16 – Optimized cascade reaction for the synthesis of CDP-glycerol. a) time course of the cytidine-based components and b) adenosine-based components. Error bars represent the standard deviation of biological triplicates.

Notably, cytidine is rapidly consumed, and after 6 h, a concentration of 25 mM CDP-glycerol is obtained. This titer gradually increases to 31.2 mM (22.5 mAU/min) after 24 h, equivalent to a Yp/s of 89%.

For the adenosine-containing compounds (Figure 16b), ATP is slowly consumed, and side products are produced. Figure 17 shows an HPLC-UV chromatogram of the reaction after 24 h. The peaks detected at retention times 6 min and 10 min are most likely adenosine-containing components such as ADP-glycerol and AATP.

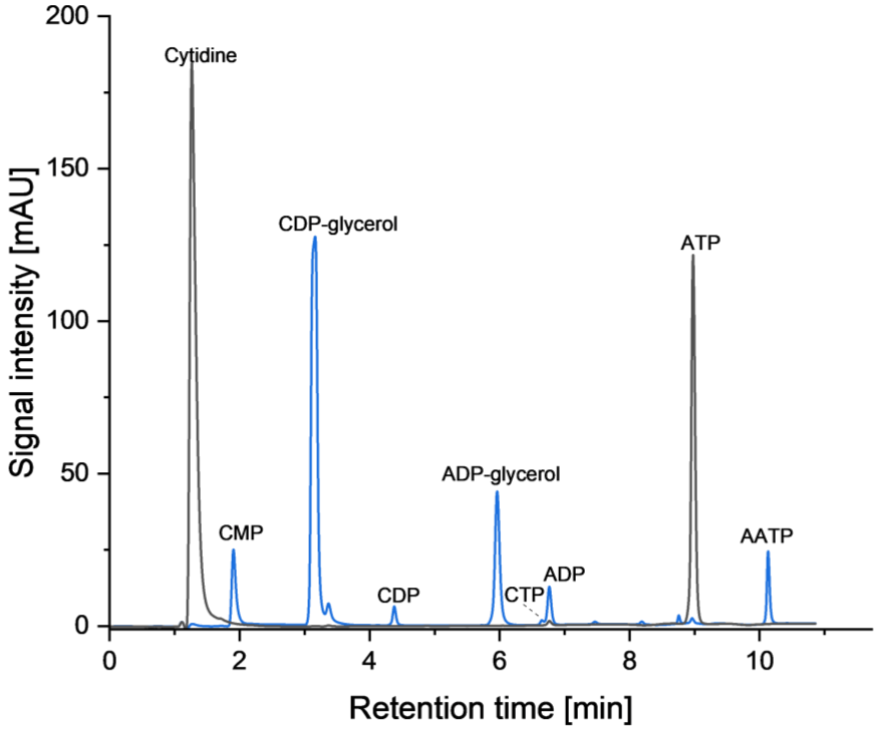


Figure 17 - Intermediates in the synthesis of CDP-glycerol. HPLC-UV chromatogram for the cascade reaction with optimised temperature, pH value and $MgCl_2$ concentration. In black: time zero, in blue: 18 h of incubation. Species were identified by their retention time compared to analytical standards, except for ADP-glycerol and AATP, which are the most likely species for the respective unidentified peaks.

4.1.3 Discussion

Individually purified enzymes

The purification of recombinant enzymes was the first step in the process, and it was conducted both with and without storage in 50% glycerol, a standard practice for cryoprotection. Glycerol serves as a substrate for the enzyme glpK, and including it in all enzyme stocks would complicate the control of its concentration in the reaction.

In the stock solution containing glycerol, tagD was purified to a final concentration of 28 g/L after buffer exchange and concentration. A total protein yield of 28 mg was obtained from a 200 mL bacterial culture under these conditions. The total enzyme yield was measured at 28.6 mg without glycerol. In comparison, Park et al.[194] reported a total enzyme yield of 46.3 mg from a 1 L bacterial culture. Notably, the pET expression system was employed in both studies; however, while affinity chromatography was utilised for purification in this study, Park et al. utilised anion exchange chromatography[194].

The purified GlpK was concentrated to 1.6 g/L after the addition of glycerol. From a 200 mL bacterial culture, a total protein yield of 6.4 mg was obtained. Without glycerol, the total enzyme yield was measured at 12.1 mg. Previous studies by Katsumi et al.[211] and Koga et al.[198] reported total recovered protein amounts of 5 mg and 9.2 mg from a 1 L culture after purification of this enzyme[211]. Therefore, the protein yield obtained in this study falls within the previously reported range for the purification of this enzyme.

The purified PPK3 was concentrated to 0.24 g/L after the addition of glycerol, yielding 0.96 mg of purified protein from a 200 mL bacterial culture. Without glycerol, the total enzyme yield was 14 mg from a 200 mL culture. Notably, the enzyme precipitated during the chromatographic step, appearing as white matter in the elution fractions. This recurring characteristic of PPK3 precipitation has been documented in previous literature[73, 115]. It is presumed that this phenomenon is attributed to PPK3's membrane-associated nature[214].

The purified URA6 was concentrated to 1.32 g/L after the addition of glycerol, yielding a total protein concentration of 2.64 mg from a 200 mL bacterial culture. The total enzyme yield was 31.5 mg from the same culture volume without the addition of glycerol. Mahour et al. reported a total protein yield of 42.84 mg from a 500 mL bacterial culture[71].

The purified UDK was concentrated to 0.91 g/L after the addition of glycerol, yielding a total protein amount of 3.64 mg from a 200 mL culture. The total enzyme yield was 11.2 mg from the same culture volume without the addition of glycerol. Notably, no studies have reported the recombinant overexpression and purification of this enzyme. Valentin-Hansen reported a recovery of 0.6 mg from a 10 L bacterial culture, although no overexpression was performed[213].

The purified PPA was concentrated to 12.34 g/L after the addition of glycerol, resulting in a total protein yield of 61.7 mg from a 200 mL cell culture. The total enzyme yield was 46 mg from the same culture volume without the addition of glycerol. Previous studies reported a total recovered enzyme amount of 2.05 mg from a 500 mL cell culture, whereas another reported 53 mg from a similar culture volume[71, 73]. The enzyme recovery in this study aligns with the range of previously reported amounts for PPA.

Proof of concept and CDP-glycerol identification

A proof-of-concept experiment using only the tagD and glpK enzymes to synthesise CDP-glycerol directly from CTP and glycerol was conducted. An unidentified compound was detected during HPLC-UV analysis on the sample taken after 18 h of reaction; this compound was subsequently identified as CDP-glycerol using MALDI-TOF MS/MS. Interestingly, another compound with a m/z of 500.07 was found, which is hypothesised to be ADP-glycerol. Although ATP is not expected to be a substrate for the enzyme tagD[93], the mass balance suggests that it is an adenosine-containing compound. It has been observed in the literature that ADP-fucose can be a co-product of GDP-fucose[33] synthesis. Therefore, the author believes that a similar side reaction is likely occurring in this enzymatic cascade.

Reaction engineering

When the whole cascade, including ATP regeneration, was tested, no CDP-glycerol formation could be initially observed. According to the compounds measured when running the cascade reaction using enzyme stocks stored in glycerol compared to stocks without it, CDP-glycerol could not be synthesised when an excess of glycerol is present. On the other hand, CDP-glycerol could be successfully synthesised by limiting the amount of glycerol in the reaction. The author hypothesises that the consumption of ATP by gplK is faster than its regeneration by the PPK3-PolyP system, quickly leading to an ATP deficiency. This is supported by literature where it has been previously reported that the k_m for ATP of glpK is as low as 0.0078 mM for the wild-type enzyme[215] while the lowest k_m reported for UDK is 0.062 mM[216], effectively favouring the consumption of ATP by glpK to form Gly-3P rather than CMP by UDK. All subsequent experiments on this cascade reaction were performed using recently purified enzymes stored without glycerol at 4 °C. Alternatively, controlling the initial concentration of glycerol can be a straightforward task, as it can be replaced by other cryoprotectant agents such as dimethyl sulfoxide or mannitol[217]. The author recommends this method for long-term storage of the enzymes.

The addition of the enzyme PPA, as initially conceived for the cascade design, was tested to determine whether it affects the yield of CDP-glycerol. As shown in section 4.1.2.4, no improvement was observed when adding PPA compared to setting up the cascade reaction without it. This observation appears to contradict previous studies. Park and collaborators described competitive inhibition of the enzyme tagD by PPi with an inhibition constant of 0.51 mM[194]. Their data were obtained in an assay using Gly-3P and CTP in a range of 0-8 mM and PPi of 0-10 mM, whereas the assay shown here in Figure 13 had a starting cytidine and glycerol concentration of 44 mM and 8 mM PolyP. The author believes that the inhibition effects observed by Park and colleagues at low concentrations of CTP and Gly-3P are not affecting the performance of this enzymatic cascade because the rate of CTP and Gly-3P to pyrophosphate is considerably higher than the one tested by them. Additionally, CMP accumulated in the reaction when PPA was added; it is hypothesised that PPA might not only degrade PPi but also CDP-glycerol into CMP and Gly-3P. Since the experiments showed no benefit from adding the enzyme PPA to the reaction, it was no longer used in the subsequent optimisation steps and was continued with only the other five enzymes.

Design of experiments

To optimise the operation conditions for the cascade reaction, a DoE screening was performed, considering temperature, pH value and $MgCl_2$ concentration as factors to investigate. These factors were chosen because they are known to influence enzymatic reactions. However, predicting their effect using kinetic modelling tools is very challenging. When working with enzymatic cascades, each enzyme has different characteristics and, therefore, can react differently to a change in temperature or pH, for example. A DoE screening approach can be effective in detecting effects, interactions, and possible nonlinear effects (MODDE® 13 does not support nonlinear models, such as exponential and logarithmic ones), even when the mechanisms behind these remain unknown. The design is defined within an experimental space, and the responses are modelled with a linear polynomial model. As shown in Figure 14, the values for R^2, Q^2, validity and reproducibility satisfy the model acceptance criteria. A possible nonlinear effect was identified for the temperature during the analysis, which reduces the predictive usefulness of the model and may explain its relatively low validity. Furthermore, the presence of outliers (Table 14) impacts the reproducibility of the model. The model was considered suiatble for running the built-in optimiser and approximating a set of parameters within the design space to yield a high response; this set of parameters was a pH of 8.76, a temperature of 29.2 °C and a $MgCl_2$ concentration of 58.52 mM. The model predicted a peak area value of 28 mAU/min under these conditions. The conditions suggested by the optimiser were tested experimentally, as described in section 4.1.2.5, and the resulting components measured over time are shown in Figure 16. The validation experiment reached a peak area of 22.5 mAU/min, equivalent to a final titer of 31.2 mM CDP-glycerol. Although the peak area of 28 mAU/min was not achieved, indicating the low predictive power of the model, the cytidine conversion yield was calculated to be 89 %. This conversion yield represents a significant improvement over the experiments performed before the optimisation with DoE, where a 50% conversion yield was achieved. This highlights the utility of the DoE approach chosen for improving enzymatic cascades. Interestingly, the other compounds in the reaction remain below 2.5 mM, with ATP being depleted after 18 h. This suggests that the ATP regeneration system may limit the progression of the cascade. The synthesis of adenosine-based co-products was observed. As confirmed by mass spectroscopy in section 4.1.2.3, ADP-glycerol is being produced, representing a dead end for the ATP regeneration system since PPK3 is not known to utillise ADP-glycerol as a substrate. It has been previously observed in other studies[218, 219] that PolyP kinases can synthesise adenosine tetraphosphate (AATP); the author believes that AATP may be produced in the cascade, because an unidentified component was observed (Figure 17) eluting later than ATP and CTP.

In summary, a novel enzymatic cascade for the synthesis of CDP-glycerol was established. The reaction was improved through a DoE approach. The experimental design used to model the response of the reaction was useful for identifying pH as the most significant factor, however, it had low predictive power and failed to identify interactions between the factors. Nevertheless, it was helpful to improve the substrate conversion yield to a final value of 89%. Further use of DoE methods to improve enzymatic cascades should not overlook factors that may

be important to represent the system, such as enzyme and substrate concentrations.

4.2 CMP-Neu5Ac

The development of an enzymatic cascade for the production of CMP-Neu5Ac is described in this chapter. The cascade synthesises CMP-Neu5Ac from CMP, GlcNac, pyruvate, ATP and PolyP as substrates. A schematic of the cascade reaction is shown in section 4.2.1 of this chapter. Section 4.2.2 shows the results of recombinant enzyme production using *E. coli* and co-expression of multiple enzymes in this strain, followed by all reaction engineering steps towards optimising the cascade reaction for higher yield and volume. This chapter contains work that was published in the B.Sc. thesis of Jasmin Vogel (Hochschule Flensburg, Germany) and in the M.Sc. thesis of Jesus Pardo (Aalborg University, Denmark).

The identity of bacterial plasmids, detailed methods and reaction conditions have been removed and are available upon request. This applies to sections 4.2.2.3, 4.2.2.5, 4.2.2.6, 4.2.3 of this chapter and Table 13 in the appendix.

4.2.1 A Cell-free Enzymatic Cascade for the Synthesis of CMP-Neu5Ac

A cascade reaction was designed to synthesise CMP-Neu5Ac cost-effectively. One of the most critical decisions in the design process was to define the starting substrates for the reaction. The shortest way to CMP-Neu5Ac is to use the enzyme CSS to synthesise it directly from CTP and Neu5Ac. However, the prices for these substrates were, at the time of writing, approximately 5,420 € per kilogram of Neu5Ac and 14,590 € per kilogram of CTP from Biosynth (Price calculated from the highest amount that the provider offers, e.g., 500 g of CMP and 100 g of CTP in May 2024). GlcNac and cytidine were considered candidates to be used as precursors for the enzymatic cascade. These substrates would cost 221 € per kilogram of GlcNac and 850 € per kilogram of cytidine. Interestingly, the price of CMP from the same provider was 780 € per kilogram. The costs were similar, and it would save one enzyme, so CMP was chosen as the starting substrate. This choice of substrates also implied adding pyruvate, which costs about 320 € per kilogram. The cascade reaction comprises six enzymes, including modules for ATP regeneration and pyrophosphate degradation. All the catalytic steps of the cascade are illustrated in Figure 18. First, the initial substrate CMP is phosphorylated to CDP by URA6; then CDP is phosphorylated to CTP by PPK3, which is also responsible for the regeneration of ATP from PolyP; in parallel, the initial substrate GlcNac is transformed to ManNac by the *N*-acylglucosamine 2-epimerase (AGE); from there, Neu5Ac is formed from ManNac and pyruvate by the *N*-acetylneuraminate lyase (NANA); at this point, the two parts of the cascade reaction converge when Neu5Ac and CTP are taken up by the *N*-acylneuraminate cytidylyltransferase (CSS) to form CMP-Neu5Ac and pyrophosphate; finally, the pyrophosphate is degraded into monophosphates by PPA to avoid product inhibition and drive the reaction equilibrium towards CMP-Neu5Ac formation.

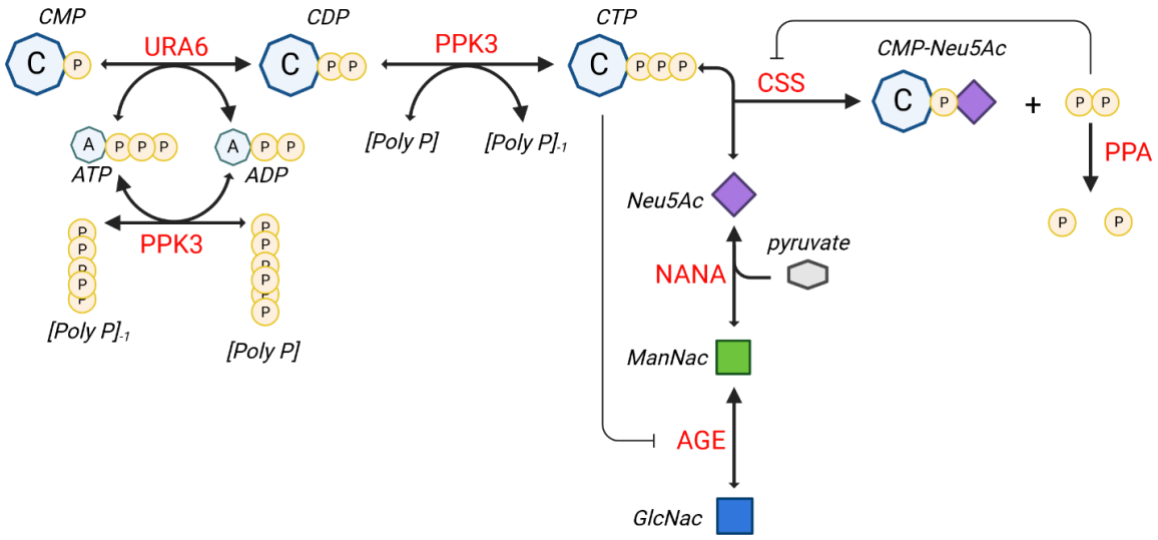


Figure 18 - Enzymatic cascade for the synthesis of CMP-Neu5Ac. Enzymes are depicted in red and components in black; enzymatic reactions are shown in black arrows weighted by size for preferred reaction direction when reversible.

As described in section 4.1.1, the design of enzymatic cascades for synthesising nucleotide sugars shares many features, such as ATP regeneration from PolyP, and pyrophosphate degradation with inorganic pyrophosphatase. Although similar strategies were used to design the cascades, each represents a challenge of its own. It has been described that CTP can inhibit the enzyme AGE. Klermund and colleagues calculated the K_i of CTP for AGE to be 1 mM[220] and, for this reason, developed a compartmentalisation strategy to prevent CTP from being in contact with AGE when running a three-enzyme cascade reaction to synthesise CMP-Neu5Aac. As the enzymatic cascade presented here does not include CTP as the starting substrate but rather as an intermediate, the challenge lies in whether it will be possible to prevent CTP from accumulating or whether it will hinder the cascade's performance by inhibiting the synthesis of ManNac from GlcNac.

4.2.2 Results

The development of a cell-free cascade capable of synthesising CMP-Neu5Ac started with the expression and purification of each enzyme to perform proof-of-concept experiments. A strategy to co-express all the required enzymes in a single *E. coli* strain was subsecuently developed to reduce the number of fermentations and enzyme purifications steps from six to only one. The process was improved through exploratory experimentation and DoE screening. Finally, a scale-up was performed to obtain gram amounts of product.

4.2.2.1 Individually Purified Enzymes

The enzymatic cascade for the synthesis of CMP-Neu5Ac comprises six different enzymes, each of which is recombinantly expressed and purified. Conveniently, the three enzymes PPK3, URA6 and PPA are shared with the cascade for CDP-glycerol and have already been discussed in section 4.1.3. For this reason, these are not thoroughly discussed in this section. This section focuses on the results of the expression and purification of the AGE, NANA and CSS enzymes.

The purified enzyme stocks were analysed using SDS-PAGE and BCA assay. The SDS-PAGE gel is shown in Figure 19. All enzymes produced for this cascade

could be successfully identified at their expected molecular weight and purity was considered sufficient for their use in the cascade reactions. The molecular weights for each enzyme in the cascade are the following: URA6—22.48 kDa, PPK3—34.74 kDa, AGE—45.61 kDa, NANA—32.60 kDa, CSS—24.89 kDa, PPA—19.31 kDa; plus 0.9 kDa belonging to the 6x His-Tag added to each enzyme. The SDS-PAGE reveals that the stocks for the enzymes NANA and CSS were highly concentrated and of high purity. Conversely, the stock for AGE contains a less pure enzyme, indicating poor overexpression. Because a larger volume could be added to the reaction mix to achieve the desired concentration, this was not deemed a critical issue during the proof-of-concept assays.

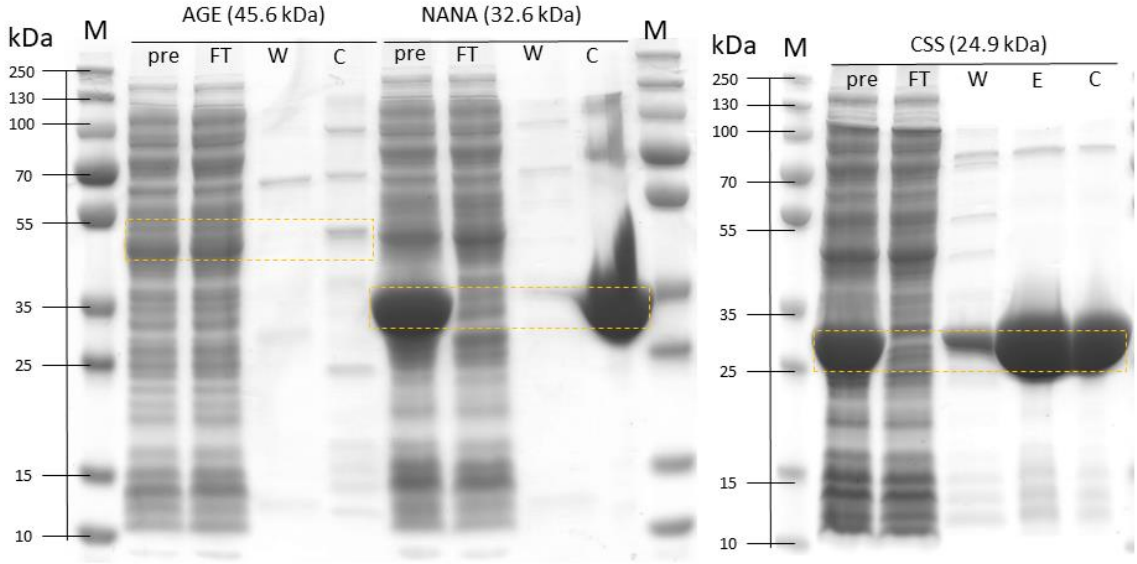


Figure 19 - SDS-PAGE showing enzymes for the CMP-Neu5Ac cascade. For each enzyme, metal affinity chromatography fractions are shown. kDa – kilodalton; M - Molecular weight marker; AGE, NANA and CSS – enzymes; pre – cell lysate fraction; FT – flowthrough fraction; W – wash fraction; E – elution fraction; C – concentrate fraction (10 kDa MWCO Amicon centrifugal filtered). Between 1-2 µg of total protein was loaded into each 10% polyacrylamide gel well. The gel was stained with Coomassie blue and scanned using a tabletop Epson scanner.

Table 7 - Summary of purified enzyme stocks of the CMP-Neu5Ac cascade. The highest amount of total purified target protein obtained in this work is compared with other works described in the literature.

Enzyme	Total target protein in this work [mg]	Bacterial culture volume [mL]	Total target protein previously reported [mg]	Bacterial culture volume [mL]
AGE	1.48	200	40 [199]	1000
NANA	258.8	200	250 [221]	1000

Enzyme	Total target protein in this work [mg]	Bacterial culture volume [mL]	Total target protein previously reported [mg]	Bacterial culture volume [mL]
PPK3	14	200	57.85 [71]	500
URA6	31.5	200	42.84 [71]	500
CSS	45.78	200	22, 90, 175 [222-224]	1000,6000
PPA	61.7	200	53 [73]	500

*AGE – *N-acylglucosamine 2-epimerase; NANA* – *N*-acylneuraminate lyase; PPK3 – polyphosphate kinase 3; URA6 – UMP-CMP kinase 3; CSS – *N*-acylneuraminate cytidylyltransferase; PPA – inorganic pyrophosphatase.

All purified enzyme stocks for this work were stored in 50% glycerol at -20 °C. Total protein concentrations were measured using a BCA assay, and the results for total protein produced per volume of bacterial culture can be found in Table 7. Here, a comparison between the protein yields obtained in this work and previous reports in the literature can also be found for each enzyme.

4.2.2.2 Proof of Concept

An assay was set up using all the purified enzymes to confirm if the synthesis of CMP-Neu5Ac could be readily achieved. The conditions of the assay were 100 mM TRIS/HCl (pH 8), 45 mM $MgCl_2$, 0.1 g/L URA6, 0.04 g/L PPK3, 1 g/L CSS, 0.1 g/L AGE, 1 g/L NANA, 0.1 g/L PPA, 11 mM CMP, 11 mM GlcNac, 5 mM ATP, 6 mM PolyP and 11 mM pyruvate at 37 °C with sampling over time for 21 h. The time course of the enzymatic cascade is illustrated in Figure 20. A titer of 5 mM CMP-Neu5Ac was observed, confirming that the cascade reaction could synthesize the nucleotide sugar; this established a baseline for the process intensification steps in the following sections. CDP, CTP, CMP, and AMP concentrations remain below 1 mM. The concentration of ATP remains stable throughout the reaction at around 3 mM.

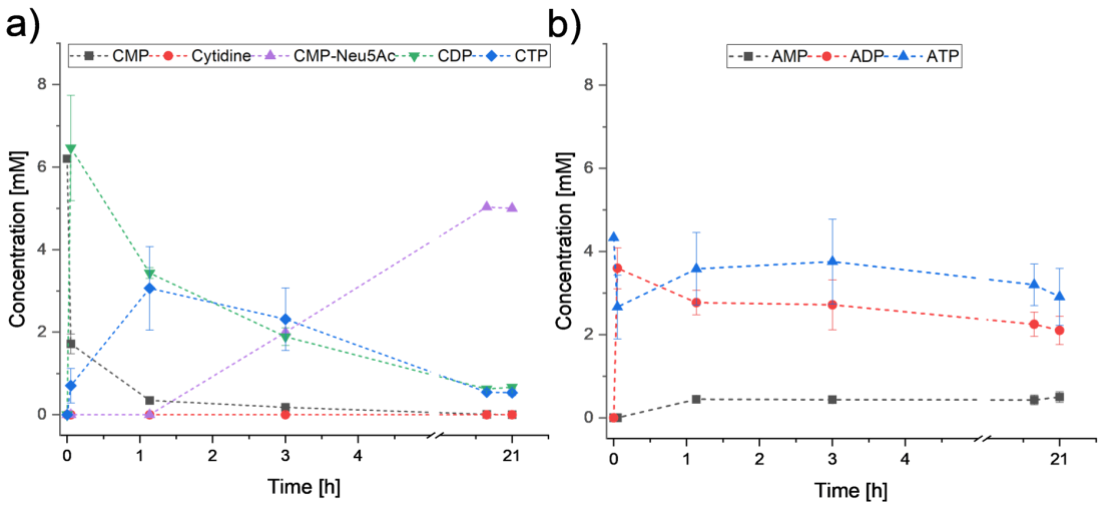


Figure 20 - Reaction with isolated enzymes CMP-Neu5Ac. The time course of the reaction was measured using HPLC-UV at separate time points. a) the cytidine-containing compounds, and b) the adenosine-containing compounds. A break in the timescale is introduced between 5 and 19 h (no sampling). Error bars represent the standard deviation of biological triplicates.

4.2.2.3 Co-expression with pDuet System

It was decided to explore the possibility of co-expressing all the cascade enzymes in a single *E. coli* strain; achieving this would substantially reduce the effort required to obtain the biocatalyst. The pDuet™ vectors developed by Novagen® were selected as a platform for the co-expression of multiple recombinant enzymes. Four different vector combinations were designed to account for variations in expression due to differences in the copy number of each plasmid and the cloning site of each enzyme (Table 8). These combinations were designated as A, B, C, and D. For each combination, various strains were produced, depending on the transformation scheme used.

Table 8 - pDuet combinations established for the CMP-Neu5Ac enzymatic cascade. In total, four different combinations were designed: A, B, C, and D. Each combination consists of three plasmids, and each plasmid has one gene per multicloning site.

Combination	MCS 1	MCS 2
B	UDK	AGE
A	UDK	AGE
A/B	NANA	CSS
A/B	URA6	PPK3
C	URA6	PPA
C	CSS	NANA
C	AGE	PPK3

Combination	MCS 1	MCS 2
D	PPK3	CSS
D	NANA	AGE
D	URA6	PPA

*MCS – multi-cloning site; UDK – uridine kinase; AGE – *N*-acylglucosamine 2-epimerase; NANA – *N*-acetylneuraminate lyase; CSS – *N*-acylneuraminate cytidylyltransferase; URA6 – UMP-CMP kinase 3; PPK3 – polyphosphate kinase 3; PPA – inorganic pyrophosphatase. Vector identities have been removed at request of eversyn GmbH.

The strategy to introduce each set of three vectors into the *E. coli* host was to co-transform each plasmid simultaneously, following the standard bacterial transformation protocol (section 3.1.3). However, the transformation efficiency was low in the case of vector combination C, for which a stepwise transformation method was used (data not shown). This method involved introducing two vectors in the first step, selecting and isolating transformants, treating the cells for chemical competency and introducing the third vector in the second step. Two strains were created with this strategy: C1 and C2 (order of vector transformation has been removed at request of eversyn GmbH). Through visual inspection of the expression patterns, it became evident that the order of transformation into the host affected the enzyme expression profile The SDS-PAGE gels for the His-tag purified proteins from the pDuet combination C transformed sequentially and all simultaneously are shown in Figure 21, as an example. The expression profile of strain “D”, which was transformed with all three vectors simultaneously, is shown alongside these to further emphasize the differences between different strains expressing the same recombinant enzymes.

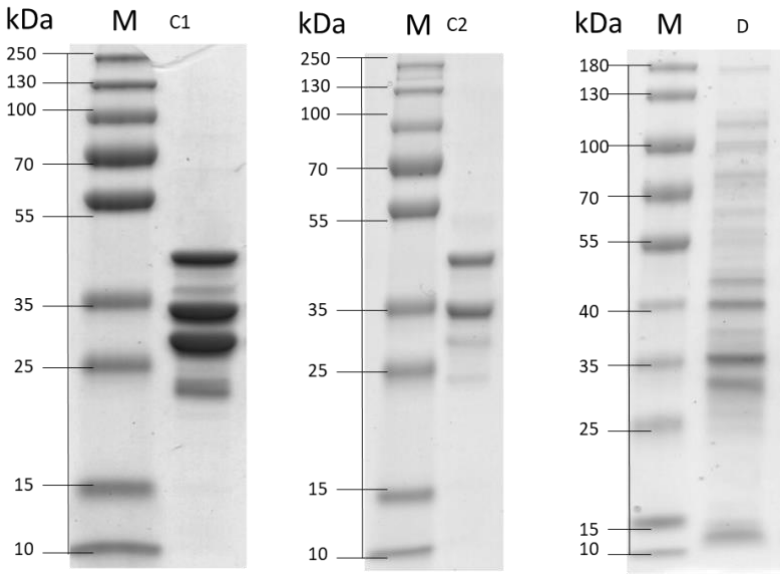


Figure 21 - SDS-PAGE of strain C and D expression profiles. The expected MW of each enzyme is approximately AGE: 45.61 kDa, PPK3: 34.74, NANA: 32.59 kDa, CSS: 24.89 kDa, URA6: 22.48, PPA: 19.31, plus 840 kDa to account for the 6x His-Tag.

The synthesis of CMP-Neu5Ac was used as the criterion to select the best-performing *E. coli* strain for further process intensification. The assay was performed combining 11 mM GlcNac, 11 mM ManNac, 11 mM CMP, 5 mM ATP, 11 mM pyruvate, 6 mM PolyP, 45 mM $MgCl_2$, 100 mM Tris/HCl (pH 8) and 2.5 g/L total protein of the co-expressed and purified enzyme stock. The reactions were carried out in biological triplicate at 37 °C in a 200 µL working volume. A summary of the results of the strains tested is shown in Figure 22. No synthesis of CMP-Neu5Ac could be observed in the strain transformed with vector combination A. For this reason, this combination and combination B (which was very similar and only varied in the selection of one vector) were discarded as potential candidates at this point. The best strain was obtained using the vector combination D transformed in a single step. Therefore, this strain was selected to continue the development of the process.

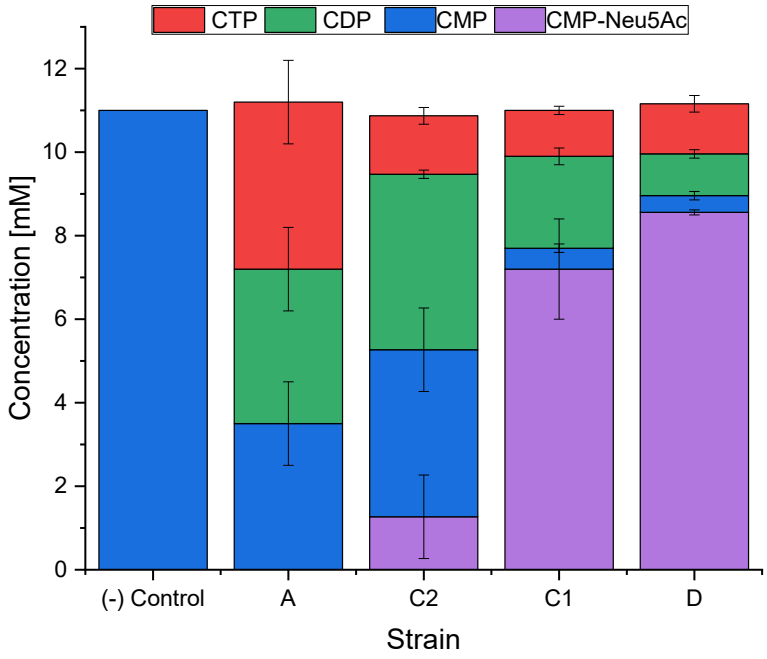


Figure 22 - CMP-Neu5Ac synthesis with different strains. Each stacked column represents the concentrations of cytidine-containing components measured by HPLC-UV. The samples were taken after 20 h of incubation. Error bars represent standard deviation from biological triplicates.

4.2.2.4 Design of Experiments

The DoE approach was a helpful starting point for optimising the enzymatic cascade for the production of CDP-glycerol. For this reason, a DoE approach was chosen to screen and analyse the cascade reaction for synthesising CMP-Neu5Ac. This setup included the factors temperature, pH value and $MgCl_2$ concentration, in the ranges of 7.5-9 for pH, 25-45 °C for temperature and 10-100 mM for $MgCl_2$. The remaining variables were kept constant at 1 g/L enzyme, 10 mM CMP, 10 mM GlcNac, 10 mM pyruvate, 5 mM ATP and 6 mM PolyP. In this case, a D-optimal quadratic design was chosen with 26 conditions to be performed in biological triplicates, resulting in 78 independent experiments. Although one of the goals of using DoE methods is to reduce the number of experiments, a higher number of experiments was used in this design to obtain a model that can better identify the effects and interactions of the factors involved. After the experimental runs were completed, no outliers were found with an outlier threshold of 3σ; therefore, the model was fitted using 78 experiments (DF = 68). The experimental conditions for each experiment, along with their measured responses (in mM of CMP-Neu5Ac, as determined by HPLC-UV), are presented in Table 15 in the Annex.

MODDE® built-in statistical tools were used to refine the model. The coefficient plot was used to identify and exclude non-significant factors, and no transformation of the responses distribution was done for this design. The model evaluation criteria computed by the software are presented in Figure 23 with an R^2 value of 0.89, a Q^2 value of 0.84, a model validity value of -0.2 and a reproducibility value of 0.97. The values obtained for R^2, Q^2 and reproducibility point to an excellent set

of data that could be useful for prediction and analysis. However, a model validity of -0.2 indicates a potential lack of fit. This value may be due to the presence of non-linear effects and interactions that are not represented by the model. Furthermore, a low value for validity can also occur in models with high values for Q2 (Q2 > 0.9) or those with extremely good replicates. The sensitivity analysis revealed that the factor with the most significant impact was the concentration of $MgCl_2$, followed by the temperature. The analysis also revealed that the concentration of $MgCl_2$, temperature, and pH value have potential non-linear effects on the response. Possible interactions were found between the temperature, concentration of $MgCl_2$ and the pH value of the reaction; this is interesting and could be explained by the temperature sensitivity of TRIS/HCl, the buffer used for the cascade reactions.

An optimisation was performed using the MODDE® built-in optimiser tool to maximise the response value. The values suggested by the optimisation were a temperature of 28 °C, a pH value of 8, and a $MgCl_2$ concentration of 83 mM; this setpoint predicted a CMP-Neu5Ac response of 10 mM, corresponding to a 100 % substrate conversion yield.

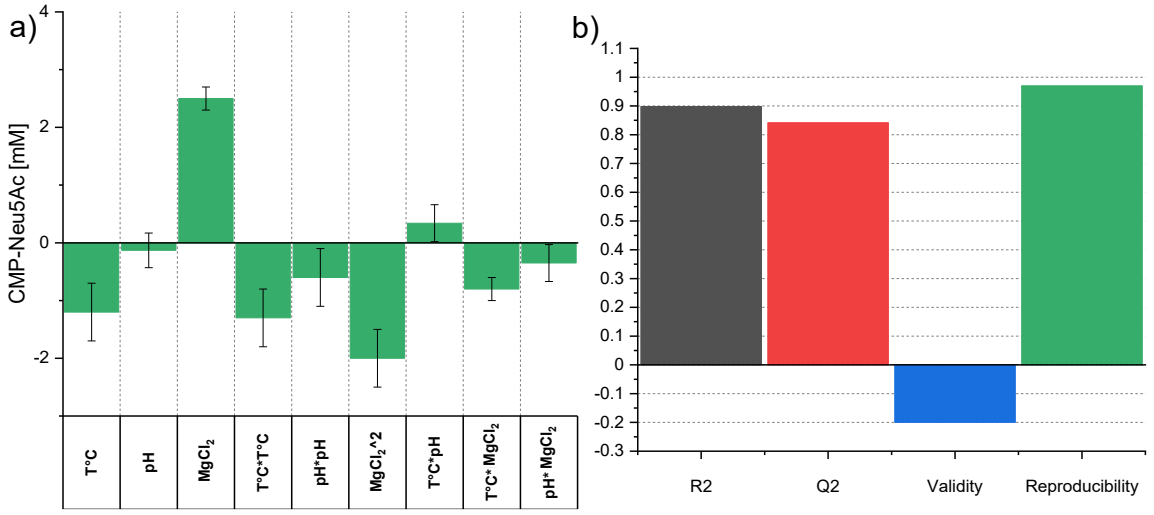


Figure 23 - DoE model evaluation for the CMP-Neu5Ac enzymatic cascade (3σ). Sensitivity analysis and model evaluation. a) Effect graph for each factor. Green bars represent the magnitude of a positive or negative influence on the response, while the error bars represent the significance of the response. Single-factor terms, interaction, and quadratic terms are included for each factor. b) Summary of fit. R^2: goodness of fit; Q^2: prediction precision; model validity: a test for diverse model problems; reproducibility: variation within replicates. N=78, DF=68. Model fit and analysis were performed by the built-in tools in the software package MODDE® 13.

As an exploratory exercise, the model was refined in an alternative manner to obtain a validity score greater than 0.25; in this case, the outlier threshold was reduced from 3σ to 2σ. The result was the exclusion of 17 data points (Table 15) and a model fit with alternative results for the coefficient plot and the validation metrics, as shown in Figure 24. The results for the recomputed metrics are an R^2 of 0.98, a Q^2 of 0.93, a validity of 0.27 and a reproducibility of 0.99. Considering these values, this model can be considered acceptable for analysis and prediction purposes. The sensitivity analysis reveals that the concentration of $MgCl_2$ is the factor with the highest effect on the response, which is in accordance with the first refining of the model. The pH value, Temperature x Temperature, $MgCl^2$ x $MgCl^2$

and Temperature x pH value were considered significant in this model fit and the temperature effect was not found to be significant. The identification of the quadratic effects of temperature and $MgCl_2$ is maintained, but not that of the pH value. Lastly, only the interaction between the temperature and the pH was identified. It is the author's belief that the removal of data points had as a consequence the loss of analytical power of the model. The same optimisation process was performed as before, and the values obtained were a temperature of 20 °C, a pH value of 8 and a $MgCl_2$ concentration of 83 mM; the predicted response was a CMP-Neu5Ac concentration of 9.4 mM, corresponding to a 94 % substrate conversion yield.

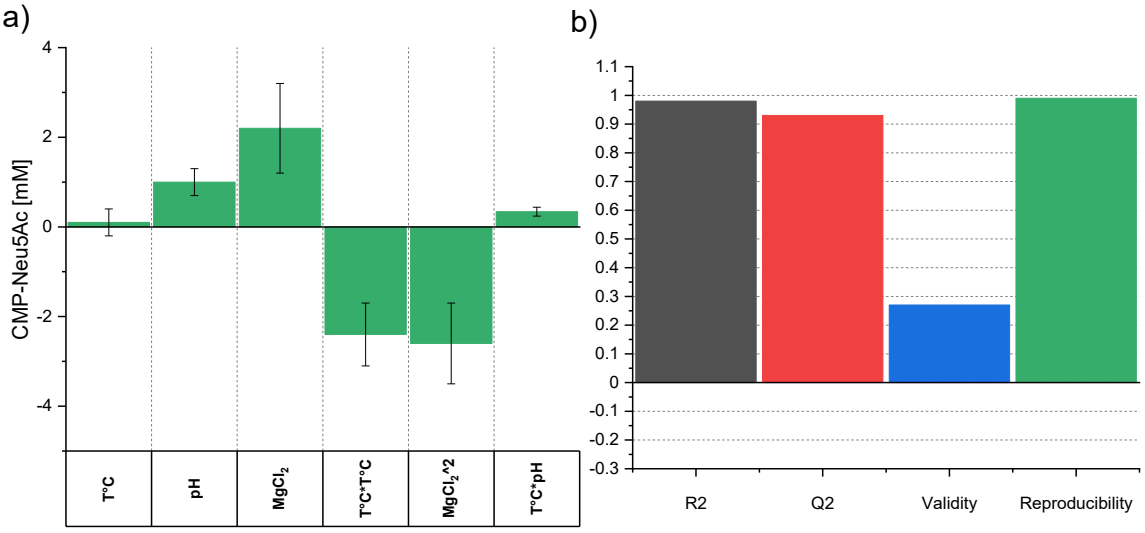


Figure 24 - DoE model evaluation for the CMP-Neu5Ac enzymatic cascade (2σ). Sensitivity analysis and model evaluation. a) Effect graph for each factor. Green bars represent the magnitude of a positive or negative influence on the response, while the error bars represent the significance of the response. Single-factor terms, interaction, and quadratic terms are included for each factor. b) Summary of fit. R^2: goodness of fit; Q^2: prediction precision; model validity: a test for diverse model problems; reproducibility: variation within replicates. N=61, DF=54. Model fit and analysis were performed by the built-in tools in the software package MODDE® 13.

The optimisation results for both models have similar values and share a counterintuitive characteristic, a negative effect of the temperature on the synthesis of the product. It is expected to have a temperature optimum for enzymatic reactions, as higher temperatures tend to increase enzyme reaction rate up to an optimum, and then decrease as the enzyme denatures (at least in mesophilic enzymes)[225]. According to the predictions made with these models, lower temperatures are proposed as optimal (25-28°C). A validation experiment for this setup was designed, encompassing a range of temperature values from 20 °C to 35 °C to observe the non-linear behaviour and the predicted local maxima described earlier. The assay was set up under the same conditions as the DoE experiments and only varied in temperature. Each experiment was performed in biological triplicate, and the results are shown in Figure 25. It was observed that, unlike in the model prediction, the final product titer is higher at higher temperatures. The optimum temperature as predicted by the model was not observed. The model's predictive power appears to be limited in this particular case, as exemplified by the temperature's effect not being well represented. It is worth noting that other factors, such as ATP and PolyP concentrations, were kept

constant instead of being included in the screening. These factors have the potential to interact significantly with each other, and excluding them from the model could have rendered the model's usefulness relatively low. Nevertheless, the approach proved useful to some extent, as plenty of insight can be drawn from the analysis. Notably, the Yp/s of CMP-Neu5Ac was effectively improved from 71 % to roughly 80 %.

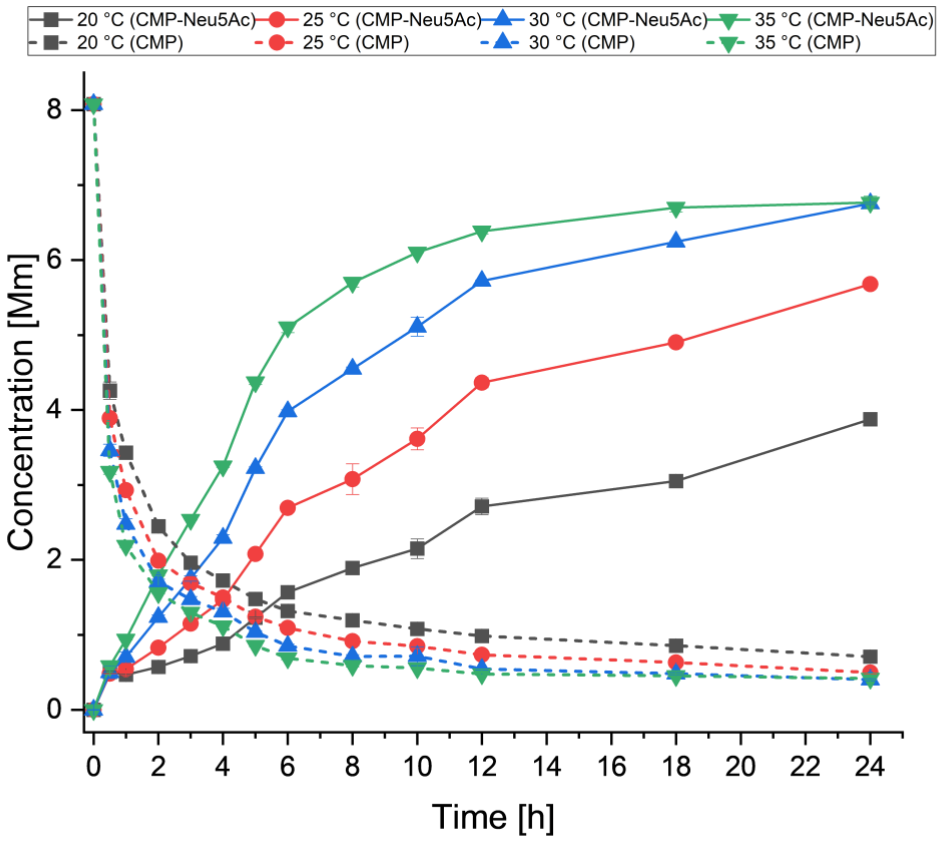


Figure 25 - CMP-Neu5Ac DoE validation experiment. Time course measurements of component concentrations measured by HPLC-UV from cascade reactions set up at 20, 25, 30 and 35 °C. Only the initial substrate CMP and the final product CMP-Neu5Ac are shown for simplicity. Experiments were performed in biological triplicates; error bars represent the standard deviation.

4.2.2.5 Reaction engineering

After the DoE screening approach, a final substrate conversion yield of 80 % was deemed good enough to move on to the following reaction engineering goal: increasing the substrate concentration load. For this purpose, the setup obtained after DoE screening was tested against a goal substrate concentration of 100 mM CMP. The setup was complemented by exploring the effect of the factors not included before in the DoE screening, particularly the concentrations of PolyPs, CMP, GlcNac and pyruvate. Another process engineering goal was to switch the biocatalyst from purified enzymes to crude cell lysates, significantly reducing overall costs. Lastly, the reaction was scaled up to produce gram amounts of CMP-Neu5Ac.

4.2.2.5.1 Increasing of Substrate Load

To explore the reaction space when the substrate concentrations are increased to 100 mM, experiments were set up with the following conditions: 100 mM CMP, 100 mM pyruvate, 100 mM GlcNac, 5 mM ATP, 150 mM TRIS/HCl, 60-200 mM $MgCl_2$ and 5–65 mM PolyP. The concentrations of $MgCl_2$ and PolyP were varied relative to each other to determine an optimal combination that would yield the highest substrate conversion in the cascade reaction. Observations from time point sampling indicated that the reaction reached equilibrium after approximately 8 h (data not shown), and therefore, the sampling time for these experiments was defined as 8 h. The results from these experiments are illustrated graphically in Figure 26.

A pattern emerged in which, as the concentration of PolyP approached 25 mM and the concentration of $MgCl_2$ approached 90 mM, the substrate conversion yield of CMP-Neu5Ac increased to a local maximum of 84 % Yp/s. Therefore, these concentrations of PolyP and $MgCl_2$ were selected for carrying out the cascade reaction with a substrate load of 100 mM.

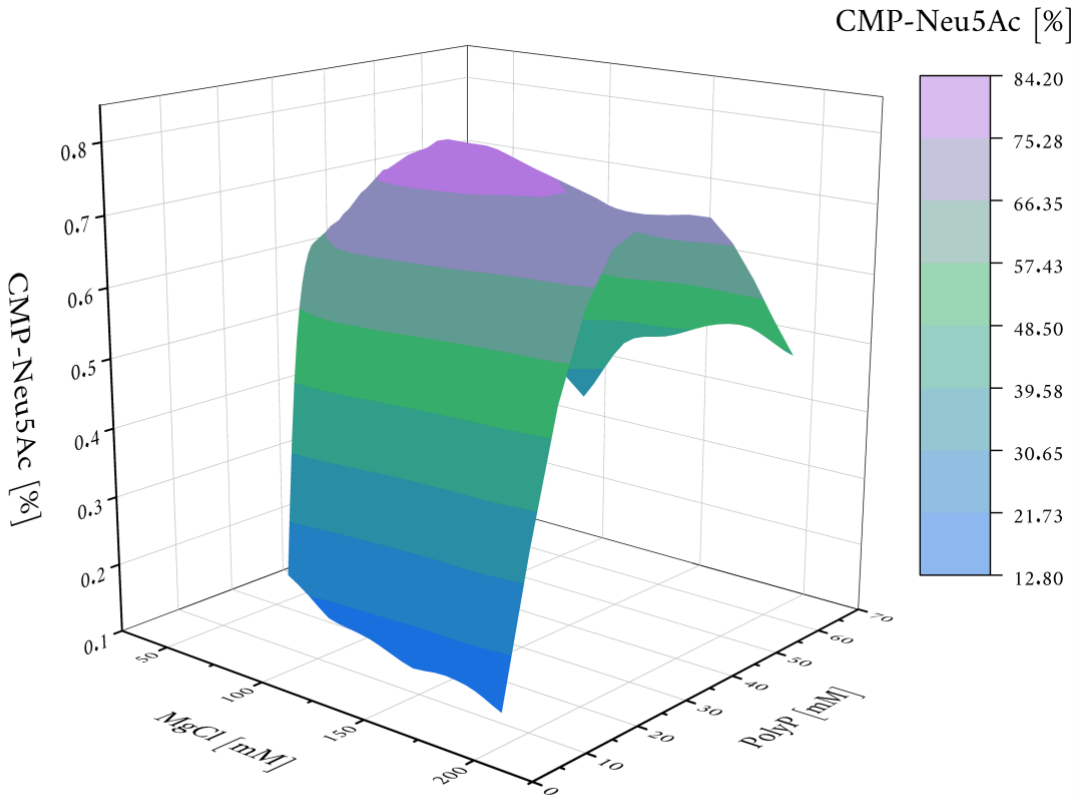


Figure 26 - Substrate increase and PolyP/$MgCl_2$ screen for CMP-Neu5Ac. A surface plot from 60 independent data points obtained from cascade reactions varying PolyP and $MgCl_2$, sampled after 8 h of reaction and measures using HPLC-UV. The synthesis of CMP-Neu5Ac is reported here in the substrate conversion yield: the percentage of CMP-Neu5Ac synthesised from the total initial amount of CMP (100 mM).

As a final note, it was theorised that changing the substrate load ratios can be done to direct the reaction equilibrium towards the product. In the case of this enzymatic cascade, the substrate that constitutes the highest cost is CMP (CMP:

780€/Kg, GlcNac: 58€/Kg, pyruvate: 260 €/Kg, in June 2024). An experiment was set up to evaluate if adding GlcNac and pyruvate in excess could result in higher CMP-Neu5Ac yields. Three different ratios of CMP:GlcNac:pyruvate were tested: 1:1.25:1.25, 1:1.5:1.5 and 1:2:2. The experiment conditions were as follows: 100 mM, CMP, 100 – 200 mM pyruvate/GlcNac, 5 mM ATP, 90 mM $MgCl_2$, 25 mM PolyP, 150 mM TRIS/HCl_2 sampled one time after eight hours of reaction. A small improvement was observed when GlcNac and pyruvate were added in an excess of 1.25 to 1 with respect to CMP (data not shown graphically). The synthesis yield with equimolar (1:1) ratio was 80.2 %, 86.6 % when using a ratio of 1:1.25:1.25, and 84.8 % with 1:2:2. Since a small improvement could be observed, it was considered a viable strategy for slightly improving the yield, although not a necessary one.

4.2.2.5.2 Cell Lysate as Biocatalyst

The use of crude cell lysate as a biocatalyst instead of purified enzymes is strategy to simplify the overall process by dramatically shortening the downstream processing of the biocatalyst, and it has been applied successfully in several examples of enzymatic cascades[13]. It was tested by setting up a reaction using cell lysate from the fermentation of *E. coli* “D” strain, co-expressing the cascade enzymes under the same conditions as the cascade reaction using purified enzymes to date. Since the purified enzymes used were at -20 °C in 50% glycerol for long-term storage, a batch of cell lysate stored under the same conditions was also used for comparison. It is worth noting that glycerol was used as a cryoprotectant agent when storing the cell lysate at -20°C; however, it was not added to the fresh lysates, thereby introducing an additional variable to the experiment. To accurately assess the difference in cell lysate storage, an experiment was conducted using cell lysates stored with and without glycerol. The experiment was set up with the following conditions: 100 mM CMP, 100 mM pyruvate, 100 mM GlcNac, 5 mM ATP, 90 mM $MgCl_2$, 25 mM PolyP and 150 mM TRIS/HCl_2. Biological triplicates were sampled over a period of 24 h. The samples taken were analysed by HPLC-UV for quantification of UV-active components. The results of this experiment are shown in Figure 27. Although not initially expected, glycerol in the cell lysates dramatically decreased the synthesis yield. This had to be considered for the process, as the cell lysate used as a biocatalyst had to be either fresh or frozen at -20 °C without further additives.

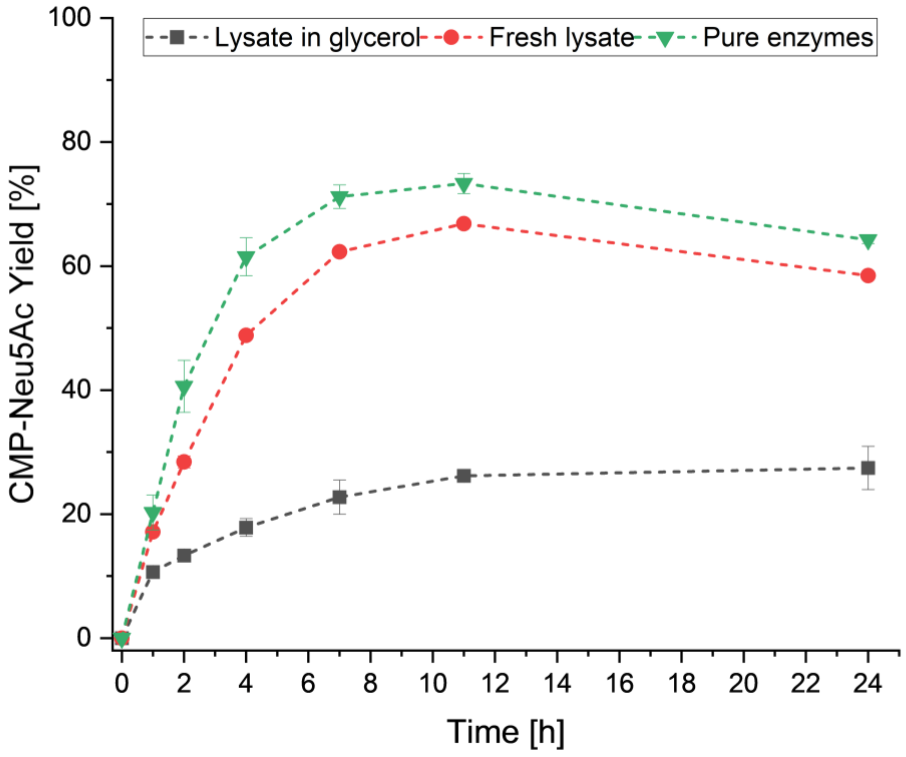


Figure 27 - Synthesis of CMP-Neu5Ac using a cell lysate. Time course measurements of cascade reactions using purified enzymes, fresh cell lysate and cell lysate stored at -20 °C in 50% glycerol. Measurements were done by HPLC-UV. Values displayed in product yield [%]. Each experiment was performed in biological triplicates, and the error bars represent standard deviation.

4.2.2.6 Gram-scale Synthesis

The final reaction conditions tested at µgram scale were directly transferred to a gram scale reaction in a shake flask set up using an incubator for temperature and shaking control. The experiment was done with a working volume of 100 mL and the following conditions: 100 mM $MgCl_2$, 150 mM TRIS/HCl, 100 mM CMP, 125 mM GlcNac, 125 mM pyruvate, 5 mM ATP. It is of note that the cell lysate used in this reaction was obtained from a different batch of *E. Coli* strain D cryo-cultures, one that was created directly from the master bank instead than from a working cell bank like the ones used for the experiments in section 4.2.2.5.2. The results of this experiment are presented in Figure 27, which shows that a substrate conversion of approximately 98 % was achieved after 24 h; this corresponds to a product titer of 98 mM (62 g/L). Notably, a substrate conversion of 93.9 % was achieved after only 8 h, opening the possibility of substantially decreasing the reaction time from 24 to 8 h with a reduction of only 4 % conversion yield as a trade-off.

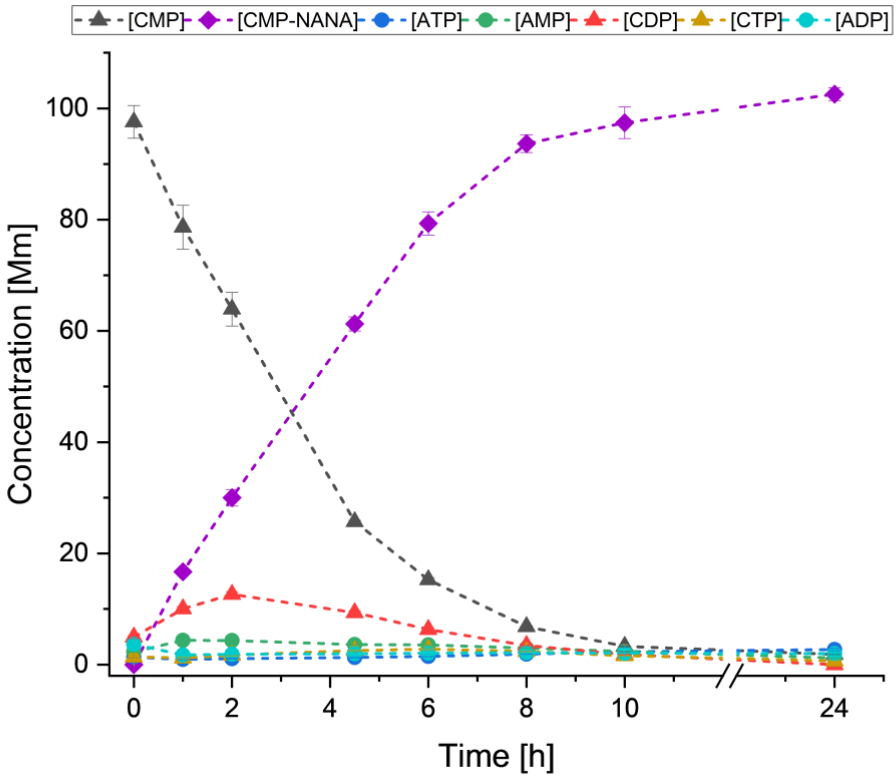


Figure 28 - Synthesis of CMP-Neu5Ac at 100 mL scale with cell lysate as biocatalyst. Cell lysate from the *E. coli* strain D was used directly after expression and cell lysis. Measurement was done by HPLC-UV. Error bars represent standard deviation from biological triplicates.

4.2.3 Discussion

Individually purified enzymes

After designing the cascade, the first step in developing a cascade reaction is to express the recombinant enzymes selected for the specific cascade. The cascade reaction for the synthesis of CMP-Neu5Ac presented in this chapter consists of six enzymes; conveniently, three of those –PPK3, URA6 and PPA— coincide with the CDP-glycerol synthesis cascade discussed in Chapter 3. For this reason, only the recombinant expression of the enzymes AGE, CSS and NANA are discussed in this section.

After purification and concentration, 1.48 mg of total protein was recovered from 200 mL of bacterial culture (Table 7). This low yield could be attributed to the experimental method used in this work because it has been reported in other research studies that the same enzyme, although from different species of cyanobacteria, could be obtained in quantities of 40 to 80 mg of total enzyme per litre of bacterial culture[199]. In work by Klermund and colleagues, it was observed that enzymes from the AGE family tend to form inclusion bodies and become insoluble[226]. It can be the case that the AGE enzyme used in this work (*Anabaena variabilis*) could not be obtained in a soluble form using our purification methods. The amount obtained was considered enough for the proof-of-concept experiments performed in section 4.2.2.2.

In the case of the enzyme CSS, 45.78 mg of total protein could be obtained from a 200 mL bacterial culture. Earlier studies, such as the one by Ganguli and colleagues[222] reported 22 mg of total protein obtained from a 6 L bacterial culture. Later research in 2012 reported that as much as 90 mg of total protein was obtained from a 1 L fermentation[223]. Lastly, in a more recent publication, Mathews and colleagues report being able to routinely express and purify CSS with yields of 175 mg per litre[224]. Altogether, the amount of protein obtained in this work agrees with what other workgroups have obtained using the same expression platform and similar protocols. Overall, high expression of CSSs from different microorganisms has been observed[65], and we can conclude that expression or solubility are not an issue when producing this enzyme.

Lastly, 258.8 mg of total protein was obtained from a 200 mL bacterial culture for the enzyme NANA. This makes it the enzyme with the highest yield among all the enzymes produced in this work. As a comparison, Li and colleagues reported a yield of 250 mg total protein from a 1 L fermentation[221]. Other studies from the recombinant expression of similar enzymes include the *N*-acetyl neuraminate lyase from *Lactobacillus plantarum* with a yield of 160 mg/L of culture[227] and the *N*-Acetylneuraminate Lyase from *Bathynomus jamesi* which has reportedly been expressed with a yield of 497 mg/L of fermentation[228]. This enzyme class seems to be efficiently expressed in *E. coli* and in an active, soluble form.

Proof of concept

In the proof-of-concept experiment, the individually expressed and purified enzymes were tested to observe whether or not the reaction would result in the final product or it would stop and accumulate one of the intermediates. The result (shown in section 4.2.2.2) was that 5 mM CMP-Neu5Ac could be synthesised in a 200 µl reaction, with a 45 % Yp/s. Based on research done by another

workgroup[226], it was reasonably expected that having the enzyme PPK3 and AGE in the same pot would inhibit AGE by CTP produced by PPK3. This inhibitory effect, however, was not significant, likely because the concentration of CTP reached an equilibrium below 1 mM during the reaction time. The concentration of ATP remained fairly stable during the reaction time, at around 3 mM. This, along with the synthesis of CDP and CTP, is a result of the ATP regeneration system involving PPK3 and PolyP.

Co-expression with the pDuet system

When working with multi-enzymatic cascades, one of the challenges is that independent fermentations are required per enzyme included in the process. The more enzymes used, the more expensive and time-consuming it is to obtain the biocatalyst. This can be challenging to overcome as economic metrics become increasingly important. Some studies estimate the cost of producing a single technical enzyme to be as much as 316 US$ per kilogram of enzyme[229], which will be multiplied by each new enzyme added to the cascade. Applying a strategy of co-expressing the cascade enzymes was chosen as the next process intensification step (Section 4.2.2.3). No other steps further than proof-of-concept were taken with the single purified enzymes because the optimisation steps would have to be performed again after the catalyst change, from single-expressed to co-expressed enzymes, was done. The platform chosen for co-expression was the pDuet™ vectors developed by Novagen®, which can clone up to eight genes in four compatible plasmids[158]. Since the copy number of each vector was different, which could affect the expression of the proteins, four different vector combinations were designed (Table 8). Even though the platform theoretically allows for the expression of up to eight enzymes, it was decided to keep the number at six because that would reduce the number of antibiotics used, make it easier to transform into the host and mean less metabolic strain during expression.

Combination A and combination B leave out the enzyme PPA because it was hypothesised that this enzyme might not be essential for the performance of the reaction (see Figure 13). It was hypothesized that cloning a gene into a higher copy number vector would result in higher expression of enzyme, but this was quickly disproved. The genes cloned in the plasmid with the highest copy number did not necessarily result in the highest expressed enzymes. Furthermore, simultaneous transformation of three plasmids required many tries before viable colonies were obtained (data not shown). When switching to a step-wise transformation method, in which two plasmids were transformed simultaneously, and the third one would be transformed in a second step, it was clear that the method of transformation had a direct effect on the enzyme expression patterns. This is observed in Figure 21. In the SDS-PAGE profiles, it can be observed that three different enzymes showed strong overexpression when the transformation was carried out in two steps; these enzymes seem to be AGE, PPK3, and NANA. Only two strongly overexpressed proteins can be observed in the transformation done with the three plasmids simultaneously in one step.

Notably, combination C was designed with the objective of increasing the expression of CSS; however, this enzyme exhibited very low expression. This leads to the argument that controlling enzyme expression in an expression system composed of pDuet vectors by selecting the highest copy number of plasmids for the desired higher-expressing enzymes is an oversimplification. Many factors

contribute to recombinant enzyme expression[230], and to more accurately predict enzyme expression in such a system, one would need to account for factors at the mRNA and protein levels, such as mRNA stability, target protein size and solubility. Characterising each possible combination of protein-vector-transformation scheme in detail would result in a time and resource-intensive investigation that was not considered for the scope of the present work. Instead, a few positive transformation hits were tested for CMP-Neu5Ac synthesis activity to select the best-performing strain for further steps. Figure 22 shows the synthesis of CMP-Neu5Ac using different purified His-tagged enzymes from each strain. Combination A had no product synthesis, which was attributed to the fact that the enzyme PPA was not included in this combination. For this reason, combinations A and B were no longer considered potential candidates. Two different *E. coli* strains for combination C were tested, both transformed in two steps but with vectors transformed in different order. This resulted in different transformation patterns. The impact of different pDuet vectors and the transformation order is a critical factor affecting protein expression profiles. To finalize the selection of a co-expressing *E. coli* strain, the strain created by transforming the vector combination D in a single step was chosen as the best option to proceed with the subsequent reaction engineering steps.

Design of experiments

Once a strain was selected for the co-expression of the cascade enzymes, the next step was to optimize the reaction conditions for optimal performance of the cascade. A DoE approach was chosen, and the resulting model evaluation is shown in Figure 23 andFigure 24. The R^2, Q^2, and reproducibility values indicate a well-fitting model. However, a value of -0.2 for model validity indicates a lack of fit, suggesting potential underlying issues with the model's fitting. An alternative model refinement was done as an exploratory exercise to observe how different the model would look if the outlier threshold was reduced from 3σ to 2σ. This resulted in the exclusion of 17 data points and values for R^2, Q^2, validity, and reproducibility that fell within the acceptable range. The sensitivity analysis shows that the most significant factor is the concentration of $MgCl_2$ followed by the temperature. Quadratic effects were detected for the $MgCl_2$ concentration, temperature and pH. This is in accordance with previous knowledge on enzyme kinetics, where these three factors have an optimum, after which the enzyme becomes inhibited or denatured[231]. Factor interactions are also identified, notably the interaction between temperature and pH. This interaction can potentially be explained a priori, as it is known that the pH value of the TRIS buffer used for the reaction is temperature dependent[232]. The model, refined by loosening the outlier threshold, failed to identify most of these effects and interactions, although it had an acceptable validity value. The exclusion of data points resulted in a loss of information, which prevented the model from being as useful as in the first iteration. This brings us to the idea that other complex interactions may remain unidentified, and these could have an effect on the reaction, especially from factors that were not considered for the DoE scheme, like PolyP concentration, which has the potential to interact with ions like Mg^{+2} [233]. These complex interactions can limit the model's predictive power when used for optimisation. In both models, the optimisation resulted in very similar reaction conditions. An assay was set up to experimentaly validate these conditions, and the results (Figure 25) show that,

contrary to the model prediction, the product output is lower at 25 °C than at 35 °C. This experiment highlights the limitations in the model's predictive power and explains the value obtained for the model validity test performed by the software. Nevertheless, the model proved helpful for understanding the cascade reaction, and the optimisation yielded a set of conditions that were better than the ones used before, with a CMP-Neu5Ac conversion yield of 81 %.

Reaction engineering

A product titer close to 100 mM was desired for the reaction to be industrially interesting; this concentration is equivalent to 61.45 g/L and would be roughly triple the highest titer obtained in the other studies[109-111, 113-115, 234]; this titer is also necessary to achieve a biocatalyst load of over 50 g product per 1 g of enzyme, assuming that an enzyme concentration of 1 g/L is used in the process. A few considerations were made to increase the concentration of substrates CMP, GlcNAc and pyruvate. For instance, only two phosphorylation reactions are required to produce one molecule of CMP-Neu5Ac within our enzymatic cascade. These phosphates are transferred from ATP to CMP and CDP to form CTP. The leftover ADP is regenerated using PolyP. Assuming PolyP with an average chain length of 25 is used, one molecule of PolyP is stoichiometrically enough for the synthesis of roughly ten molecules of CMP-NeuAc. In reality, not only is the mechanism in which PPK3 phosphorylates ADP reversible[197], but also the negative charge of the molecule is prone to interact with divalent ions like Mg^{+2}, which happens to be an essential cofactor for most of the enzymes present in the cascade. The amount of PolyP and $MgCl_2$ that has to be added in the reaction to be available for the enzymes is effectively higher than what can be calculated by looking at the reaction stoichiometry. The complex relationship between these components makes it challenging to rationally calculate the optimal ratio at which PolyP, $MgCl_2$, PPK3, ATP and CMP should be present in the reaction.

Using the cell lysate directly was pursued to bypass the enzyme purification steps. It is unknown why the cell lysate stored in 50% glycerol had reduced enzymatic activity compared to cell lysate used fresh or frozen without additives. One can theorise that glycerol might be an inhibitor for one of the enzymes of the cascade; this thought is quickly discarded as the cascade reaction worked with a product yield of over 80% when using purified enzymes, which were also stored in 50% glycerol. Another possible theory is that, since glycerol can be metabolised by *E. coli*[235], it is being transformed by the endogenous enzymes in the lysate to another molecule, which can inhibit the enzymes that are part of the cascade reaction.

Gram-scale synthesis

The synthesis of gram amounts of the product was performed as a final goal for the process intensification of CMP-Neu5Ac. The reaction conditions in small-scale experiments were directly transferred to a shake flask setup with 100 mL working volume. It was thought that the second and third-generation cryo-cultures would share the same characteristics as the master cell bank created from the original bacterial transformation. These third-generation stocks, also known as working cell banks, showed experimental yields of approximately 80% Yp/s. When a second-

generation stock, or master strain, was tested under the same reaction conditions, the Yp/s was around 98%. This result underscores the importance of strain management in ensuring the robustness of a process. Expression systems based on plasmids are susceptible to plasmid loss due to plasmid segregation phenomena[236]. Since daughter cells formed without plasmid tend to have a higher growth rate, a whole culture can significantly lose expression efficiency if such plasmidless cells happen early in the fermentation. The author hypothesises that such an event occurred during the fermentation step when producing a working cell bank of our *E. coli* strain D. Fortunately, a new cell bank could be created from the master cell bank used in the cascade reaction shown in section 4.2.2.6.

On a final note, even though an inhibitory effect of CTP towards AGE was observed before by Klermund and colleagues[220], it was not observed in this work. Since the K_i was calculated to be 1 mM, the author believes that the fluxes in the cascade reaction did not allow CTP to accumulate to concentrations at which inhibition could be an issue. Nevertheless, the author suggests that this inhibitory effect should be taken into consideration in future work, since changing enzyme concentrations or reaction conditions can lead to CTP accumulation and, therefore, AGE inhibition.

In summary, the cell-free enzymatic cascade was brought from early-stage µg-scale proof-of-concept experiments using purified single enzymes to multi-gram-scale reactions using crude cell lysates as biocatalysts, achieving a Yp/s up to 98%. With an estimated space-time yield of 7.75 g/L*h and a biocatalyst-to-product ratio of 60:1, this enzymatic cascade can now be developed into an industrial process.

4.3 GDP-fucose

The development of an enzymatic cascade for the synthesis of GDP-fucose from GMP, fucose, ATP and PolyP is described in this chapter. This work builds upon early proof of concept research done on the enzymatic cascade[32], the recombinant expression of enzymes in distinct hosts and through co-expression in a single one is shown, followed by the various optimisation approaches used to enhance the product yield and scale of the reaction. Some of the content in this chapter has been published in collaboration with Nicolas Huber from the Analysis and Redesign of Metabolic Networks group at the MPI for DCTS, Magdeburg[207], in particular, Figures 36-38 are reproduced from the original publication under the open-access license CC BY 4.0.

4.3.1 A Cell-free Enzymatic Cascade for the Synthesis of GDP-fucose

The design of an enzymatic cell-free cascade for synthesising GDP-fucose was made to start with GMP, PolyP, fucose and ATP as substrates and produce GDP-fucose. From an economic point of view, using GMP as a starting substrate represents an opportunity because it is already used in the food industry as an additive for flavour enhancement[237, 238]. On the other hand, fucose is not yet widely available in bulk amounts (100 g of L-fucose costs €1,101 from the supplier Biosynth®, summer 2024), despite its recent potential for therapeutic applications[239]. However, recent advancements in metabolic engineering and fermentation technology may increase accessibility of L-fucose in the near future[240-242]. For this reason, it was decided to develop a cascade reaction based on the salvage pathway rather than on the de novo pathway, which requires a larger number of enzymes plus NADH as cofactor [243]. The enzymatic cascade consists of four enzymes that catalyse six enzymatic reactions. The guanylate kinase (GMK) first phosphorylates GMP by transferring a phosphate group from ATP, the donor, to form GDP. The enzyme PPK3 phosphorylates GDP by transferring one phosphate from PolyP to form GTP. This enzyme's substrate promiscuity can also phosphorylate ADP to ATP, effectively recycling it within the system. The bifunctional L-fucokinase/L-fucose-1-phosphate guanyliltransferase (FKP) performs the ATP-dependent phosphorylation of fucose to form fucose-1-phosphate and then, converging with the other side of the cascade, the GDP-fucose synthesis reaction from GTP and fucose-1-phosphate. Pyrophosphate is formed as a side product of the GDP-fucose synthesis reaction, which is then degraded by PPA.

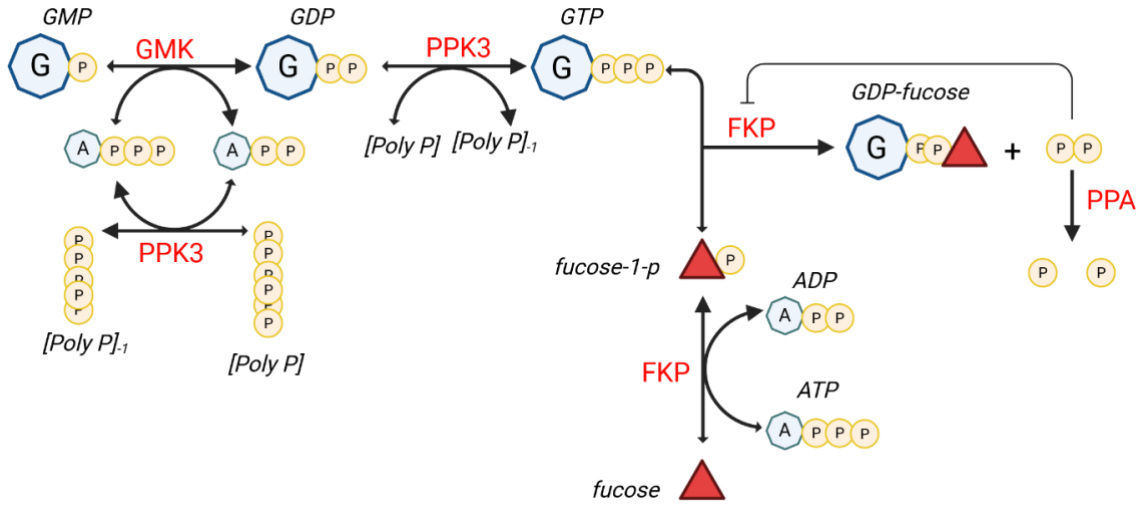


Figure 29 - Enzymatic cascade for the synthesis of GDP-fucose. Enzymes are depicted in red and components in black; enzymatic reactions are shown in black arrows weighted by size for preferred reaction direction when reversible.

A particular characteristic of this enzymatic cascade design is that utilizing the bifunctional enzyme FKP can allow fewer enzymes to complete the reaction. This enzyme was discovered in *Arabidopsis*[244] and, more importantly, in the microorganism *Bacteroides fragilis*[245]. Several research groups have used it to synthesise GDP-fucose [10, 23, 32, 33, 202].

In contrast to other nucleotides, guanosine has the quality of self-assembling under aqueous conditions[246], which can result in hydrogel formation in the presence of divalent ions[247]. This complex phenomenon will inevitably be a reaction engineering challenge specific to this enzymatic cascade, especially when working with a high substrate concentration load is desired.

4.3.2 Results

This section first describes the recombinant expression and purification of each enzyme required for the enzymatic cascade synthesising GDP-fucose. A mathematical model of enzyme kinetics was developed to predict the optimal amount of biocatalyst for use in a synthesis reaction. After experimental validation of the model, co-expression vectors were designed to allow for the simultaneous production of all necessary biocatalysts in a single fermentation step. Later, several reaction engineering and intensification steps were done, including biocatalyst blending, DoE screening, fed-batch reaction setup and enzyme mining.

4.3.2.1 Individually Purified Enzymes

Since PPK3 and PPA are discussed in Section 4.1.2.1., this section will focus on the enzymes GMK and FKP. The enzymes were recombinantly expressed as described in the Materials and Methods. The SDS-PAGE showing different fractions of the IMAC purification step is presented Figure 30. All the enzymes could be visually identified at their expected molecular weight: PPK3—34.74 kDa, FKP—105.66 kDa, GMK—23.59, kDa and PPA—19.31 kDa. The molecular weight of each enzyme is increased by approximately 0.9 kDa because of the inclusion of 6x His-tags. It can be observed that all enzymes were obtained in a sufficient purity to be used as catalysts for the enzymatic cascades.

After purification and concentration of the enzymes. The stocks were analysed by BCA assay, and total protein yield was calculated for each fermentation. Table 9 presents the results for the highest protein concentration achieved for each enzyme in this study, along with a comparison to values previously reported in the literature.

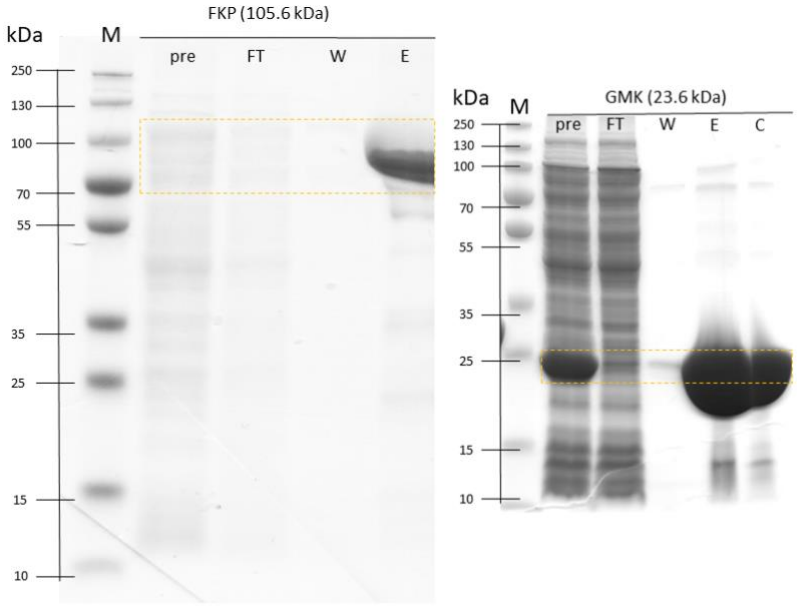


Figure 30 - SDS-PAGE showing FKP and GMK enzymes for the GDP-fucose cascade. For each enzyme, metal affinity chromatography fractions are shown. kDa – kilodalton; M - Molecular weight marker; FKP and GMK – enzymes; pre – cell lysate fraction; FT – flowthrough fraction; W – wash fraction; E – elution fraction; C – concentrate fraction (10 kDa MWCO Amicon centrifugal filtered). Between 1-2 µg of total protein was loaded into each 10% polyacrylamide gel well. The gel was stained with Coomassie blue and scanned using a tabletop Epson scanner.

Table 9 - Summary of purified enzyme stocks for GDP-fucose cascade. The highest amount of total purified target protein obtained in this work is compared with other works reported in the literature.

Enzyme	Total target protein in this work [mg]	Bacterial culture volume [mL]	Total target protein previously reported [mg]	Bacterial culture volume [mL]
GMK	68.3	200	14 [203]	1000
FKP	32.1	200	10 [248]	1000
PPK3	14	200	57.85 [71]	500
PPA	61.7	200	53 [73]	500

*GMK – guanlylate kinase; FKP – bifunctional L-fucokinase/L-fucose-1-P guanylyltransferase; PPK3 – polyphosphate kinase 3; PPA – inorganic pyrophosphatase.

4.3.2.2 Kinetic Modelling

The relative amount of each enzyme added to an enzymatic cascade is critical in determining its performance. There are well-established equations that can describe enzymatic reactions in terms of concentration of substrates and enzymes in a mathematical way, most notably the Michaelis-Menten equation and its variations[249]. The more enzymes are in the cascade, the more challenging it becomes to build a useful model of the enzymatic cascade. The enzymatic cascade for the production of GDP-fucose has four enzymes, while the enzymatic cascades for the production of CDP-glycerol and CMP-Neu5Ac have six. Therefore, this enzymatic cascade was selected as the better candidate to test this approach. A kinetic modelling approach was chosen to estimate the optimal amounts of each enzyme that must be added to the reaction. A methodology was developed to work under the uncertainty of kinetic parameters (K_m, K_{cat}, K_{eq}), while still obtaining a useful prediction from the model. A simplified description of the method is shown in Figure 31. The focus of this thesis was on product titer maximisation, but the methodology is helpful for any optimisation problem, like enzyme load minimization and total cost minimisation. For an in-depth look at these examples, see the original publication[207]. First, exploratory experiments used a theory-informed choice of enzyme concentrations, and the resulting product yields were measured over time. The experimental data was then used to fit the kinetic parameters of the model several times; n=100 was used to limit the computing power required. Each parameter set was different because of the global solution algorithm's stochastic nature. An optimisation problem was solved using each parameter set; in this case, the optimisation problem was defined as product titer maximisation but kept the total enzyme concentration constant as a constraint. This process resulted in 100 solutions (enzyme concentrations) for the optimisation problem and their respective predicted product titer output. These solutions were cross-validated against the other 99 parameters, set to fare how well they perform under kinetic parameter uncertainty.

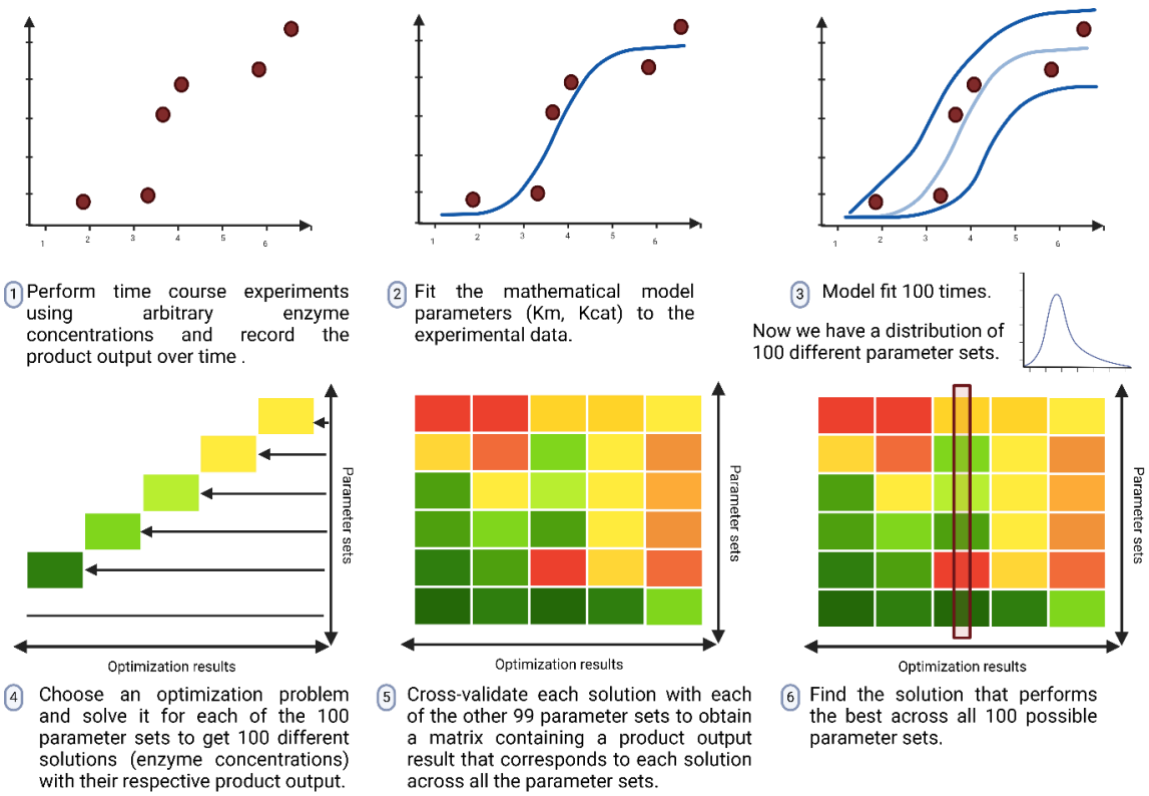
Lastly, the best-performing solution across all 100 parameter sets was selected by scoring criteria defined in the following way:

$$Score(i) = \frac{M_i}{M_{max}} + \frac{T_i}{T_{max}} + \frac{S_i}{S_{max}}$$

Equation 5 - Scoring criteria

Where:

- M_i: the median titer of column i.
- M_{max}: the overall highest median titer.
- T_i: the minimum titer of column i.
- T_{max}: the overall highest minimum titer.
- S_i: the sum of all titers of column i.
- S_{max}: the overall highest titer sum.



To generate the original time course data, individually purified enzymes were

Figure 31 - Summary of kinetic modelling methodology. (1) Time course data generation, (2) model fit, (3) n=100 repetition of model fit, (4) optimization problem solving for each parameter set, (5) cross validation, (6) scoring and selection of best solution.

used for a baseline experiment under the following conditions: 100 mM TRIS/HCl (pH 8), 45 mM $MgCl_2$, 0.78 g/L GMK, 0.1 g/L PPK3, 0.7 g/L FKP, 0.6 g/L PPA, 25 mM GMP, 25 mM fucose, 5 mM ATP, 7.5 PolyP and a temperature of 37 °C in a working volume of 200 µL. The result from the baseline experiment is shown in Figure 32. As all reaction parameters were selected based on previous knowledge but not optimised, the final Yp/s after 24 h was around 52.6 %. Time course measurements were used to fit the kinetic model, along with five other experiments of varying initial substrate concentrations and enzymes (for the detailed experimental and model data, see the supplementary data in the original publication[207]). The reaction was run for 28 h, and it was observed that for the guanosine-containing species, the substrate GMP is quickly consumed to form the intermediates GDP and GTP, after which GDP-fucose is slowly formed throughout the reaction. As for the adenosine-containing species (ATP, ADP and AMP) there is an initial fast consumption of ATP, followed by its regeneration from ADP over the next 8 h; the concentrations of the three species remain constant after 8 h until the end of the experiment.

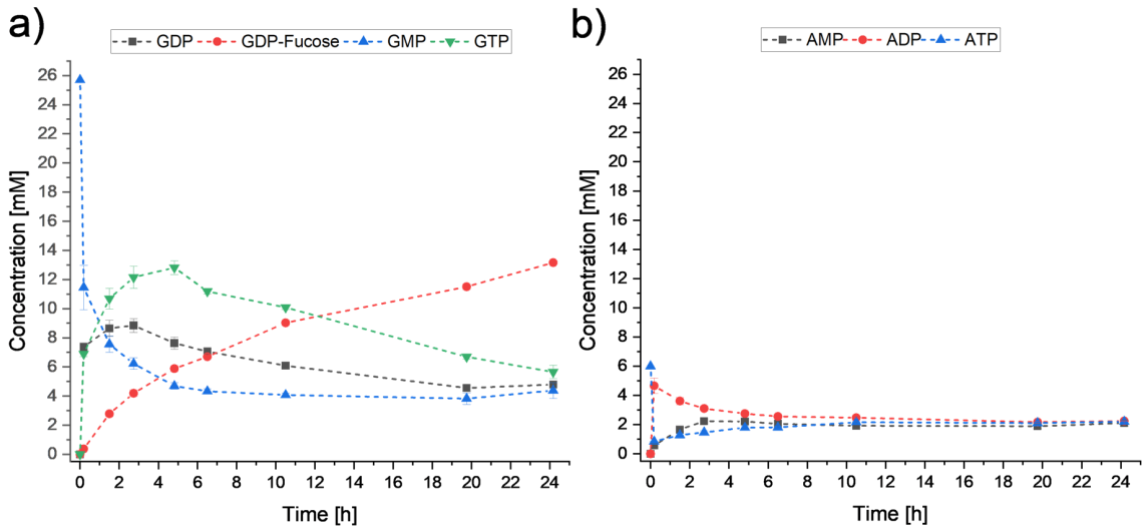


Figure 32 - Baseline experiment for the synthesis of GDP-fucose. One of the five independent experiments used for model fit. a) Guanosine-based, and b) adenosine-based components. Sampling from a 200 µL scale reaction and time course measurements done with HPLC-UV. Reactions made in biological triplicates: error bars represent standard deviation.

The kinetic model was constructed with 26 parameters corresponding to each enzymatic reaction. Some of these kinetic parameters have been estimated experimentally by other workgroups and published in the literature, but uncertainty remains because of their intrinsic variability when estimated under different conditions. To tackle this uncertainty, the available literature values were used only as starting points for the parameter estimation algorithm, which was run 100 times. The algorithm used was an evolutionary strategy (SRES) global solution algorithm implemented directly in the software COPASI. This method has an intrinsic stochastic characteristic, which is the reason for the variability of the parameterisation results even when the starting conditions are the same.

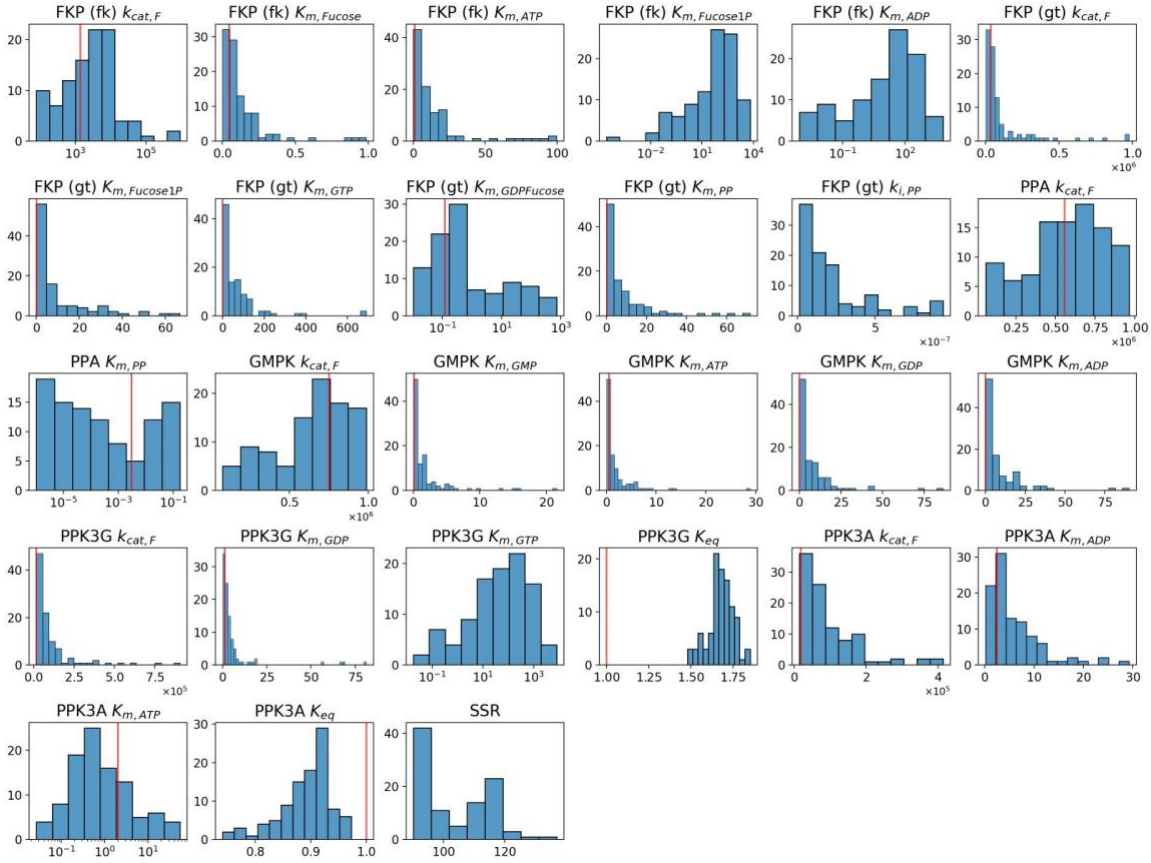


Figure 33 - Distribution of parameter sets estimated 100 times. The 26 unknown parameter values of the kinetic model are shown. Each was estimated using the evolutionary strategy (SRES) global solution algorithm in COPASI. Red lines show values obtained from literature when available. Each histogram shows parameter values in the x-axis (K_m, K_i [mM], K_{cat} [h^{-1}], K_{eq} [1]) and absolute frequency in the y-axis. Figure by Nicolas Huber from the original publication[207].

A summary of the parametrisation results is shown in Figure 33. It was found that for specific parameters, such as the Km of FKP for fucose and ATP, the distribution did not deviate significantly from the literature value.

An ensemble of 100 parameter sets was used to visualise the model and determine if it fits the experimental data. The enzyme concentrations were set to the ones used in the baseline experiment, and the trajectories for each parameter set were calculated. This visualisation is shown in Figure 34. It can be observed that the trajectories for the initial substrate GMP and the product GDP-fucose are the closest to the experimental data. The intermediate components GDP, GTP, ADP and ATP follow the curve trend but are not as close to the actual experimental values, and the species fucose and fucose-1-p show an extensive variation; this is because these species were not measured; for this reason, there was no data for the model to fit to.

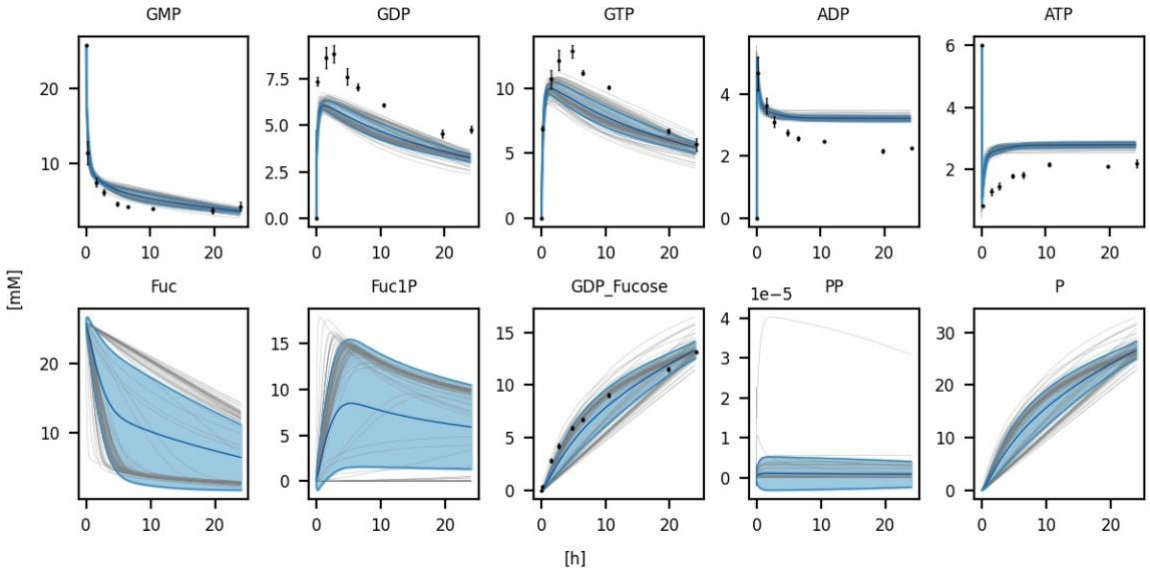


Figure 34 - Prediction trajectories of 100 model fits. An ensemble of 100 estimated kinetic parameter sets was used to simulate the time course curve of the enzymatic cascade. Each graph shows concentration [mM] against time [h] for a different species intermediate of the reaction. Black dots represent experimental data with error bars as standard deviation from biological triplicates. Grey lines show individual simulation trajectories, the central blue line shows the average of all simulation trajectories, and the outer blue lines show their +/- standard deviation. Figure by Nicolas Huber from the original publication[207].

The next step was to solve an optimisation problem with each parameter set to obtain a combination of enzyme concentrations. In this case, the optimisation problem was defined as maximising product titer while keeping the total amount of enzyme constant. Solving the optimisation problem for each of the 100 parameter sets resulted in 100 enzyme concentration combinations. These were cross-validated with the remaining parameter sets to create a matrix where parameter sets are in the y-axis, solutions (enzyme concentrations) are in the x-axis, and the performance (product titer) is represented with a heat map. The matrix is shown in Figure 35. The scoring criteria were then used to select a solution that performs the best across all parameter sets, as marked in a red window.

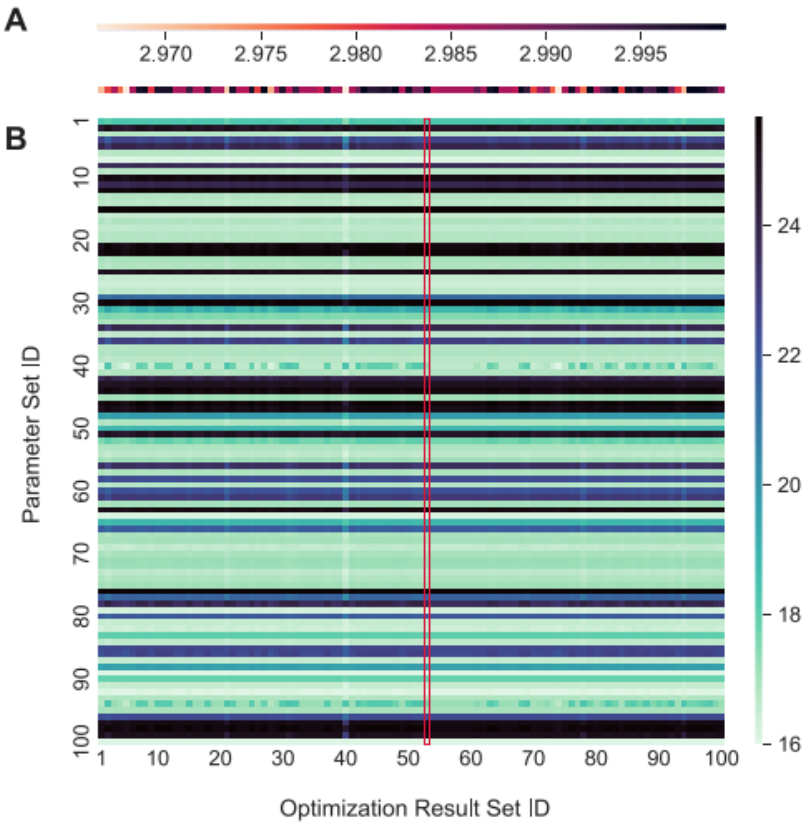


Figure 35 - Simulated GDP-fucose titers and scores of each optimisation solution.Scores of each proposed optimisation solution according to Equation 5. The solution with the highest score, solution ID 53, is marked in a red window. (B) Heat map of GDP-fucose titers simulated with each solution (enzyme concentrations) and each kinetic parameter set. Figure by Nicolas Huber from the original publication[207].

The selected solution was validated experimentally by setting up a reaction with the following conditions: 100 mM TRIS/HCl (pH 8), 45 mM $MgCl_2$, 0.057 g/L GMK, 0.32 g/L PPK3, 1.57 g/L FKP, 0.39 PPA, 22 mM GMP, 22 mM fucose, 5 mM ATP, 7.5 mM PolyP and a temperature of 37 °C. The reaction was carried out in a working volume of 200 µL, and samples were measured using HPLC-UV. The result of these time course measurements compared to the curves predicted by the kinetic model is shown in Figure 36. It is worth noting that for species that are not measurable by HPLC-UV, the predicted trajectories cannot be compared to experimental values. Although the trends appear to be correct, the experimental values do not closely match the predicted values. In the case of GMP, GDP, and GTP, the prediction underestimated the reaction speed, estimating a slower consumption of GMP in the first hours; it then estimated an equilibrium approximately 4 mM above the actual values after 20 h. The adenosine-based species equilibrium was also estimated slightly above the real values. The component AMP was observed in the experimental validation, but was not a part of the kinetic model because no reported enzymatic reaction could form AMP. It could be the product of a side reaction of spontaneous degradation of ADP, which is thus being formed, affecting the dynamics of the reaction and the precision of the model.

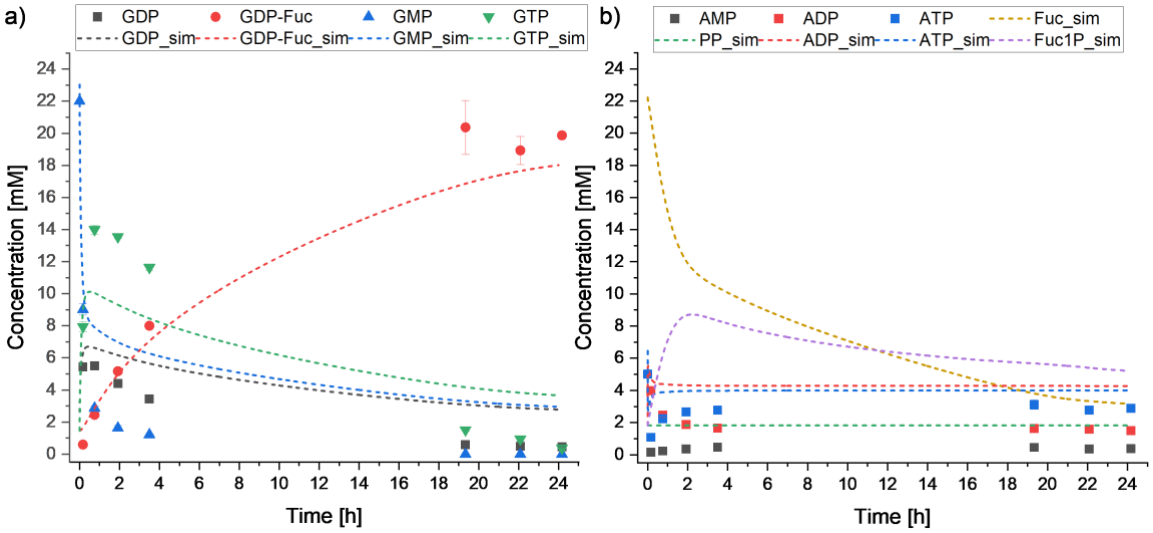


Figure 36 - Simulated vs experimental synthesis of GDP-fucose. Time course measurements from the experimental validation of the model simulation with optimised enzyme concentrations for maximum product titer. a) guanosine-containing compounds, and b) adenosine-containing compounds (also simulations for fucose and PP). Symbols represent measured data points, and dotted lines represent kinetic model simulations. Fucose, fucose-1-P and PP were not measured. The reactions were sampled over 24 h. Error bars represent the standard deviation of biological triplicates.

The measured final titer of GDP-fucose in the validation experiment was 19.86 mM, significantly higher than the value of 13.16 mM obtained in the baseline experiment. The kinetic modelling method to change the starting enzyme concentrations in the cascade reaction improved substrate conversion yield from 51.2% to 90.3%.

4.3.2.3 Co-expression with pDuet Platform

A co-expression strategy was developed to offset the apparent disadvantage of using cell-free enzymatic cascades, where each enzyme must be expressed independently. Co-expressing the cascade enzymes would reduce the number of independent fermentations required, significantly reducing costs and simplifying the overall process. A total of six different pDuet vector combinations were designed for the co-expression of enzymes necessary in the GDP-fucose synthetising cascade. One of the vectors could not be successfully cloned (data not shown). Therefore, this combination was discarded early on. Five vector combinations were tested, shown in Table 10. This section describes the selection process to elucidate which of these vector combinations performed the best in terms of GDP-fucose synthesis.

Table 10 - Vector combinations for co-expression of GDP-fucose cascade enzymes. The different combinations of the pDuet vectors and genes cloned in the multi-cloning sites 1 & 2 are shown.

Combination	pDuet	MCS 1	MCS 2
1	RSF	FKP	PPA

Combination	pDuet	MCS 1	MCS 2
1	CDF	PPK3	GMK
3	ACYC	FKP	PPA
3	CDF	PPK3	GMK
4	pET	FKP	PPA
4	CDF	PPK3	GMK
5	ACYC	FKP	PPK3
5	CDF	GMK	PPA
6	pET	FKP	PPK3
6	CDF	GMK	PPA

*/RSF/CDF/ACYC/pET – Duet plasmid´s origin of replication; MCS – multi-cloning site.

The vector combination one was tested first. The two plasmids were transformed into *E. coli*, and the standard enzyme expression and purification protocols were performed as described in the Materials and Methods. An SDS-PAGE showing the IMAC protein fractions is shown in Figure 37. Samples of the cell lysate, flowthrough, wash, elution and concentrated stock were loaded into the gel. The four proteins of the cascade could be visually identified at their expected molecular weight. According to the size of the observed bands, the enzymes GMK and PPA appear to be in a much higher concentration than the enzymes PPK3 and FKP.

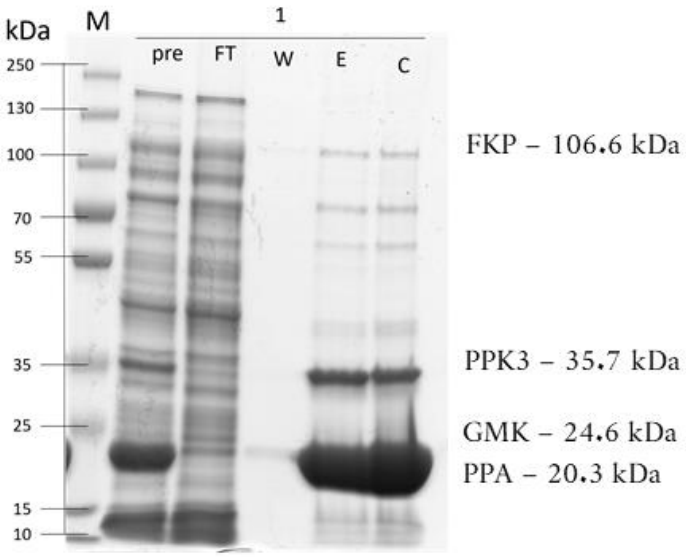


Figure 37 - GDP-fucose strain 1 protein purification. Bacterial cultured cells were lysed and centrifuged to remove cell debris, the resulting crude cell lysate was purified through IMAC and then the eluted his-tagged proteins were concentrated using Amicon® centrifuge filters. SDS-PAGE was done as described in Materials and Methods, and approximately 1-2 µg of

protein was added into each lane. Lane key: M – molecular weight marker, GDP-fucose duet 1 – strain ID, pre – crude cell lysate, FT – flowthrough, W – wash, E – elution, C – concentrate. Approximated molecular weight of enzymes, including 6x poly-Histidine tag, is: FKP – 106.6 kDa, PPK3 – 35.7 kDa, GMK – 24.6 kDa, PPA – 20.3 kDa.

Characterising the expression patterns of each combination of vectors individually can be time-consuming. To quickly and effectively screen for the best combination of expression vectors, the micro-fermentation platform BioLector® was utilized. This device enables 48 independent fermentations to be carried out simultaneously, opening the possibility of performing biological replicates for the fermentation/expression step and screening several strains simultaneously. The vector combinations 3, 4, 5 and 6 were transformed into *E. coli*. After that, one colony was picked from the selection plate of combination 3, and 2 colonies were picked from the selection plate of combinations 4, 5 and 6. The resulting *E. coli* strains were designated 3, 4a, 4b, 5a, 5c, 6a and 6b. Parallel micro-fermentation was performed to screen the enzyme expression patterns of each strain. The result of this screening is shown in Figure 38. Overall, two distinct expression patterns were identified: high expression of GMK and PPA with low expression of PPK3 and FKP was observed in strains 3, 4a, 4b, and 5c; and high expression of PPK3 with very low expression of FKP, GMK and PPA was observed in strains 6a and 6b.

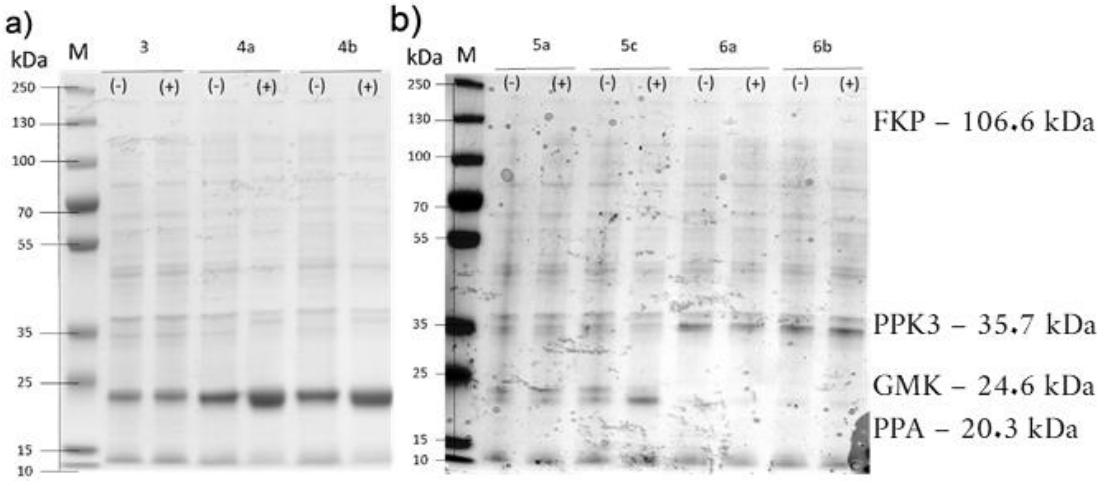


Figure 38 - Expression profiles screen of GDP-fucose strains. SDS-PAGE with lysate of bacterial culture harvested after 24 h, induced after 8 h with 0.4 mM IPTG (+) and uninduced culture (-). a) shows protein expression profiles for strains 3, 4a and 4b. b) shows protein expression profiles for 5a, 5c, 6a and 6b. Approximated molecular weight of enzymes, including 6x poly-Histidine tag, is: FKP – 106.6 kDa, PPK3 – 35.7 kDa, GMK – 24.6 kDa, PPA – 20.3 kDa.

After the screening, two representative strains were selected for further purification and testing of enzymatic activity. Strain 4a was chosen with the expression profile of high GMK, and strain 6a was selected with the expression profile of lower GMK but higher PPK3. The results of IMAC enzyme purification are shown in Figure 39. FKP was barely detectable in SDS-PAGE on all strains tested. Therefore, concentration by IMAC purification was performed to verify its expression, although at considerably lower levels than when the enzyme is expressed individually, as shown in section 4.3.2.1.

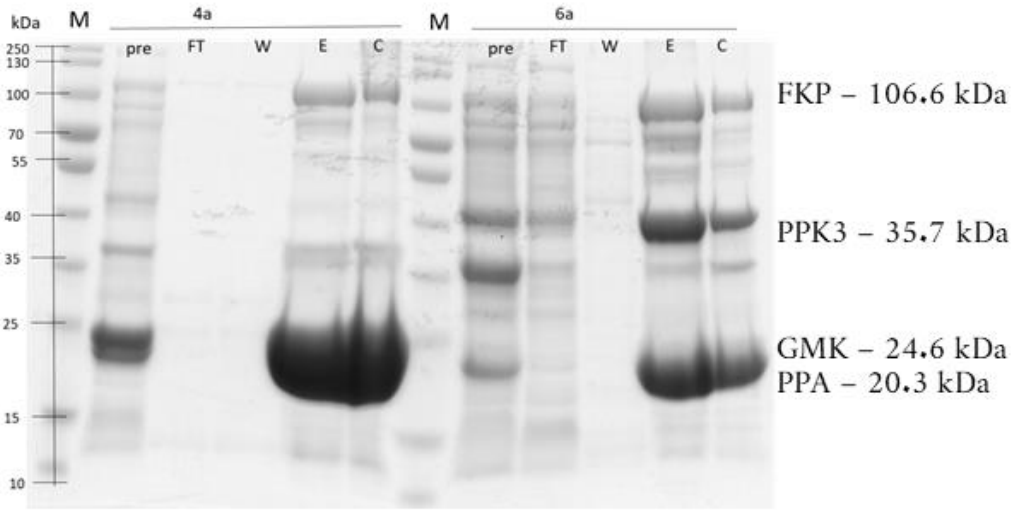


Figure 39 - GDP-fucose strains 4a and 6a protein purification. The cells were lysed and centrifuged to remove cell debris. The resulting crude cell lysate was then purified through IMAC, and the eluted His-tagged proteins were concentrated using Amicon centrifuge filters. SDS-PAGE was done as described in the Materials and Methods, and approximately 1-2 μg of protein was added to each lane. Lane key: M – molecular weight marker, 4a/6a – strain ID, pre – crude cell lysate, FT – flowthrough, W – wash, E – elution, C – concentrate. The approximate molecular weights of the enzymes, including the 6x poly-Histidine tag, are: FKP – 106.6 kDa, PPK3 – 35.7 kDa, GMK – 24.6 kDa, and PPA – 20.3 kDa.

To test the synthesis of GDP-fucose with the candidate strains 4a and 6a, an enzymatic reaction was set up with the following conditions: 100 mM TRIS/HCl (pH 8), 45 mM $MgCl_2$, 1 g/L purified enzyme, 15 mM GMP, 15 mM fucose, 5 mM ATP, 7.5 mM PolyP and a temperature of 37 °C. The reaction was sampled over 48 h, as the previous 24 h reaction period had been observed using purified enzymes, and it was expected that using co-expressed proteins would result in a slower reaction speed. The result of these reactions is shown in Figure 40. At first glance, both enzyme preparations successfully synthesised GDP-fucose from GMP and fucose as starting substrates. Strain 6a reached a product titer of 4 mM in 24 h, while strain 4a reached only 2 mM in the same time. After 48 h, both strains reach a concentration of 4 mM. The concentration of ATP was in equilibrium for most of the reaction, at approximately 2 mM. Based on these observations, strain 6a was selected as the best-performing strain in the screening and was chosen for further development.

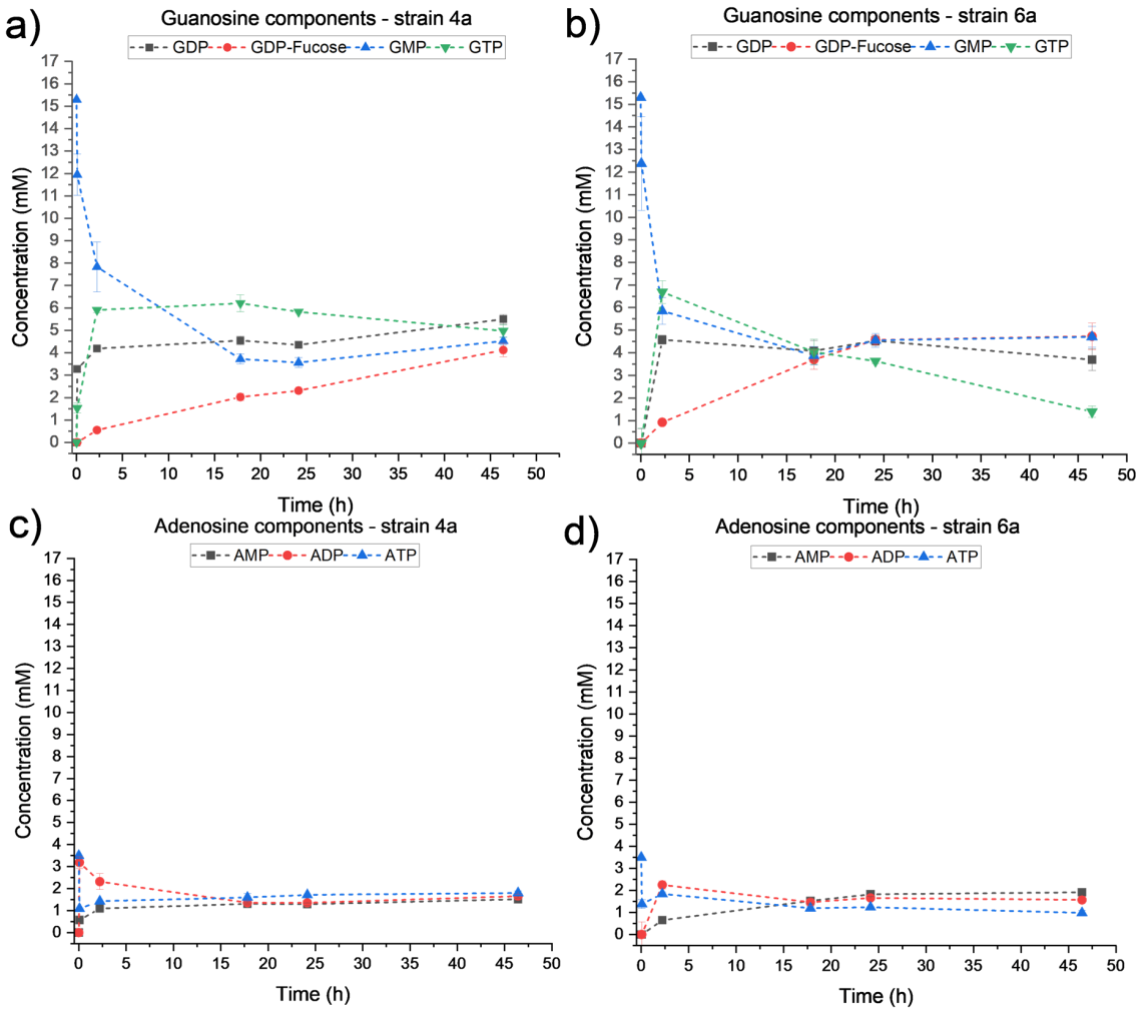


Figure 40 - GDP-fucose synthesis with *E. coli* strains 4a and 6a. Time course measurements were taken with HPLC-UV. a) time course measurements for reaction with purified enzymes from strain 4a , guanosine-containing intermediates. b) time course measurements for reaction with purified enzymes from strain 6a, guanosine-containing intermediates. c) time course measurements for reaction with purified enzymes from strain 4a, adenosine-containing intermediates. d) time course measurements for reaction with purified enzymes from strain 6a, adenosine-containing intermediates. The reaction was sampled for 48 h. Error bars represent the standard deviation of biological triplicates.

4.3.2.4 Reaction Engineering

After selecting a suitable co-expression strain, the process intensification steps focused on optimising the process variables and setup. A biocatalyst blending approach was considered to offset the low FKP expression observed when co-expressed in *E. coli* (Figure 39). This approach included mixing the co-expressed proteins with individually expressed and purified FKP. The use of cell lysate as a biocatalyst was also implemented, skipping the enzyme purification steps (data not shown). Lastly, efforts were made to define the substrate load limitations and, if possible, increase the substrate load in a manner similar to that of the enzymatic cascade CMP-Neu5Ac production in section 4.2.2.5.1.

Increasing Substrate Load

Following the cascade reaction for the synthesis of CMP-Neu5Ac in section 4.2.5.1, a substrate load of 100 mM was also desired for the synthesis of GDP-fucose. Using the reaction conditions obtained from the kinetic modelling optimisation as a basis, an experiment was set up to increase the substrate load while maintaining the substrate conversion yield. The reaction conditions for this experiment were as follows: 200 µL reaction volume in 2 mL centrifugal tubes at 37 °C, 100 mM Tris/HCl (pH 8), 100 mM $MgCl_2$, 0.057 g/L GMK, 0.32 g/L PPK3, 1.57 g7L FKP, 0.39 g/L PPA, 30-100 mM GMP and fucose, 5 mM ATP, 25 mM PolyP. When mixing higher concentrations of GMP into the reaction, it was observed that visible white precipitates were formed. This could be avoided to a certain degree by slowly adding the compound, accompanied by heating and sonication. The reactions were measured at various time points by HPC-UV; the results are shown in Figure 41. To compare substrate conversion yields when the initial substrate concentrations change, the results are normalised as a percentage of the total component concentrations of each reaction. As the concentrations of the substrates GMP and fucose rise, the conversion yield noticeably declines. At the lowest concentration tested in this set of experiments, 30 mM of GMP and fucose, the Yp/s was close to 50%, reaching only 10% when the substrate load was 100 mM.

As the substrate conversion yield was found to decrease significantly with increasing substrate load, the aim of further process intensification steps for this cascade reaction was to improve the conversion yield at moderate substrate loads (<30 mM).

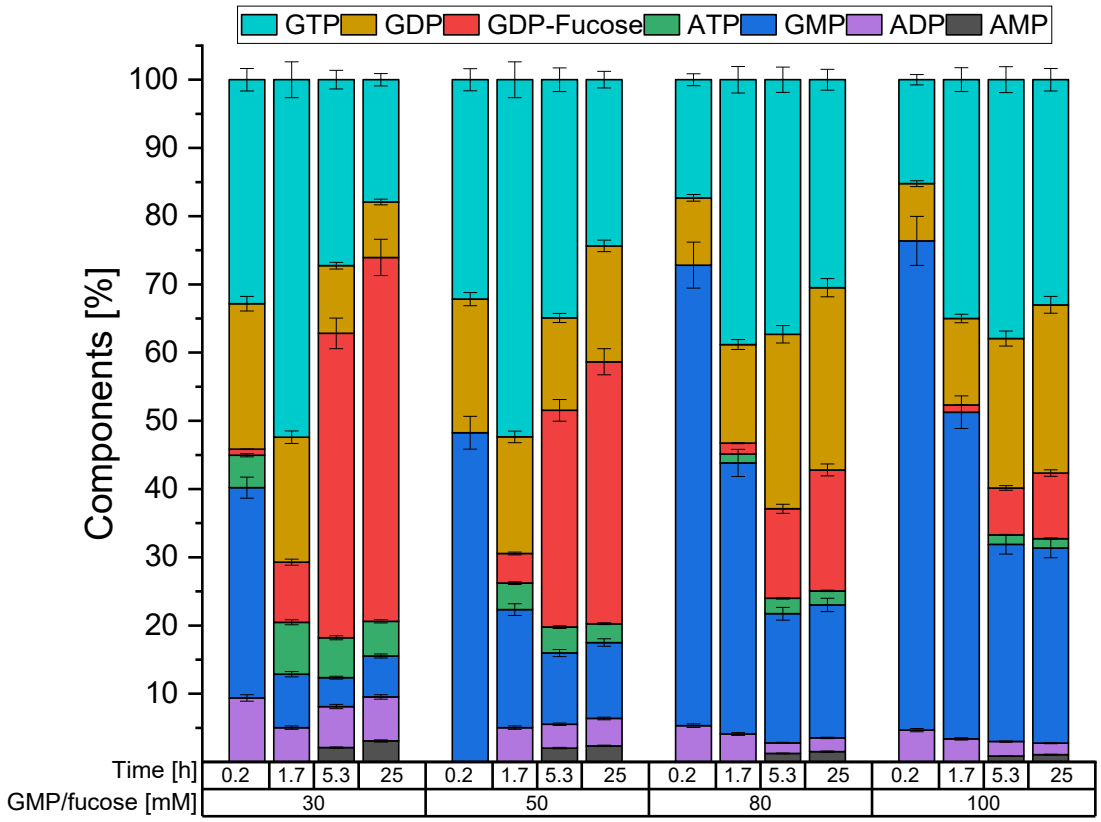


Figure 41 - Increasing substrate load for the GDP-fucose cascade. Measurements were done by HPLC-UV; error bars represent measurement error. Reaction components are represented as stacked columns normalised by the percentage of the total concentrations of UV-active components, which facilitates the comparison between conversion yields over time for different substrate loads.

4.3.2.5 Design of Experiments

The setup aimed to encompass a broader range of factors than those designed for the CDP-glycerol and CMP-Neu5Ac enzymatic cascades. A fractional-factorial level IV design was made to perform the minimum amount of experiments to obtain a useful model from nine independent input factors (pH value, polyP [mM], $MgCl_2$ [mM], GMP/fucose [mM], T °C, FKP [g/L], ATP [mM], co-expressed protein in cell lysate [g/L] and $MnCl_2$ [mM]) and measuring the output as substrate conversion yield [%] from GMP to GDP-fucose after 18 h of reaction time. A total of 54 independent experiments were performed with the centre point repeated three times. The detailed combination factor ranges and measured responses are presented in Table 11 in the Appendix. Based on the results observed from the kinetic modelling, which indicated that the concentration of this enzyme is a critical factor in determining the reaction performance, one of the factors included in the experimental design was the addition of purified FKP to the reaction. Other factors to consider were using the cell lysate of the co-expression strains as a biocatalyst and adding $MnCl_2$ to the reaction. This approach was inspired by the results obtained with the CMP-NANA synthesis cascade (Section 4.2.2.5.2).

Table 11 - Input factors for DoE setup with a fractional factorial IV design.

Name	Units	Range
pH	n.a.	7-9
polyP	mM	5-50
$MgCl_2$	mM	1-50
GMP & fucose	mM	5-30
Temperature	°C	23-50
FKP	g/L	0.5-2
ATP	mM	2-10
Enzymes	g/L	0.3-1.5
$MnCl_2$	mM	0-10

*PolyP – polyphosphate; GMP – guanosine monophosphate; ATP – adenosine triphosphate.

The model was fitted using PLS and then refined. An outlier threshold of 3σ was defined, and no outliers were identified using this criterion. The distribution of the response exhibited positive skewness, but the logarithmic transformation did not improve the model, so no transformation was applied. The coefficient plot was used to remove non-significant terms. The total number of experiments used for the model fit was N=54 with DF=40. The statistical analysis using MODDE® resulted in values for R^2 = 0.068, Q^2 = 0.40, validity = -0.2, and reproducibility = 0.96. The model validity was calculated to be <0.25, which suggests a lack of fit issue with the model. Furthermore, a value of Q^2 > 0.5 implies a model with low predictive power. A summary of the model analysis is presented in Figure 42.

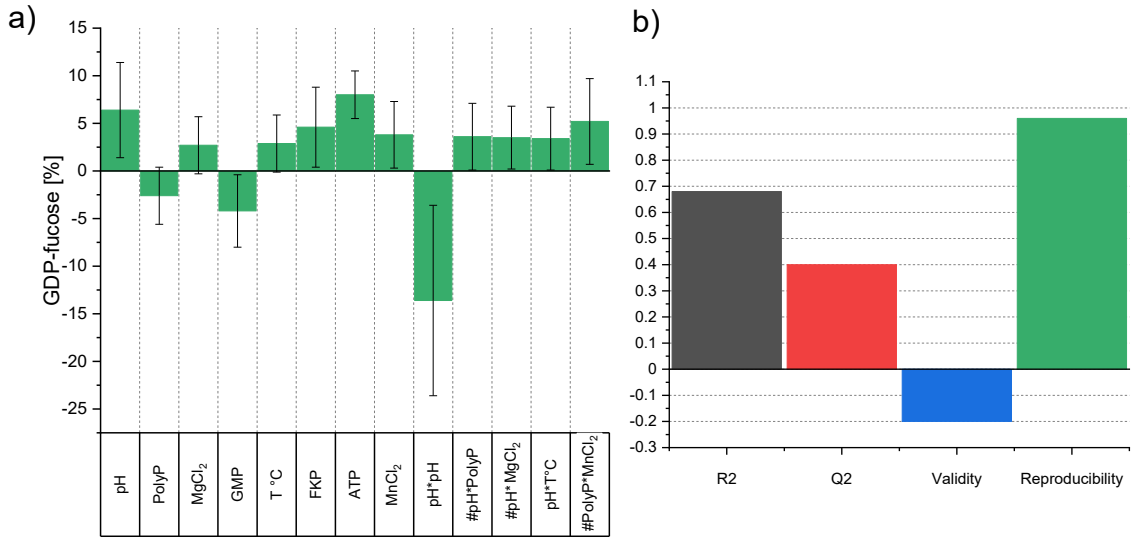


Figure 42 - DoE model evaluation for the GDP-fucose enzymatic cascade (3σ). Sensitivity analysis and model evaluation. a) Effect graph for each factor. Green bars represent the magnitude of a positive or negative influence on the response, while the error bars represent the significance of the response. Single-factor terms, interaction, and quadratic terms are

included for each factor. b) Summary of fit. R^2: goodness of fit; Q^2: prediction precision; model validity: a test for diverse model problems; reproducibility: variation within replicates. N=54, DF=40. (#) confounded interaction. Model fit and analysis were performed by the built-in tools in the software package MODDE® 13.

The sensitivity analysis performed using the MODDE® software is shown in Figure 42 (a). Out of the nine initial factors used in the design, only one (the concentration of cell lysate) was excluded from the model fit, as the analysis found no significant effect. The factor with the highest impact was the pH value, followed by the concentrations of purified ATP and FKP. A quadratic effect was identified for the pH value, and several factor-factor interactions were also observed. The interaction between pH and temperature was identified. However, the interactions between pH-polyP, pH-$MgCl_2$ and PolyP-$MnCl_2$ are confounded, and more experimental data would be necessary to resolve them.

As an exploratory exercise, the outlier threshold was reduced from 3σ to 2σ. When this was done, 10 potential outliers were identified and excluded. The evaluation of the model is presented in Figure 43. The values of R = 0.92, Q2 = 0.71, validity = 0.3, and reproducibility = 0.95 indicate that this model may be useful for prediction purposes. Although the model fit was improved, it can be observed in the coefficient plot that less significant factors and interactions were identified, this reflects the loss of information resulting from excluding data points.

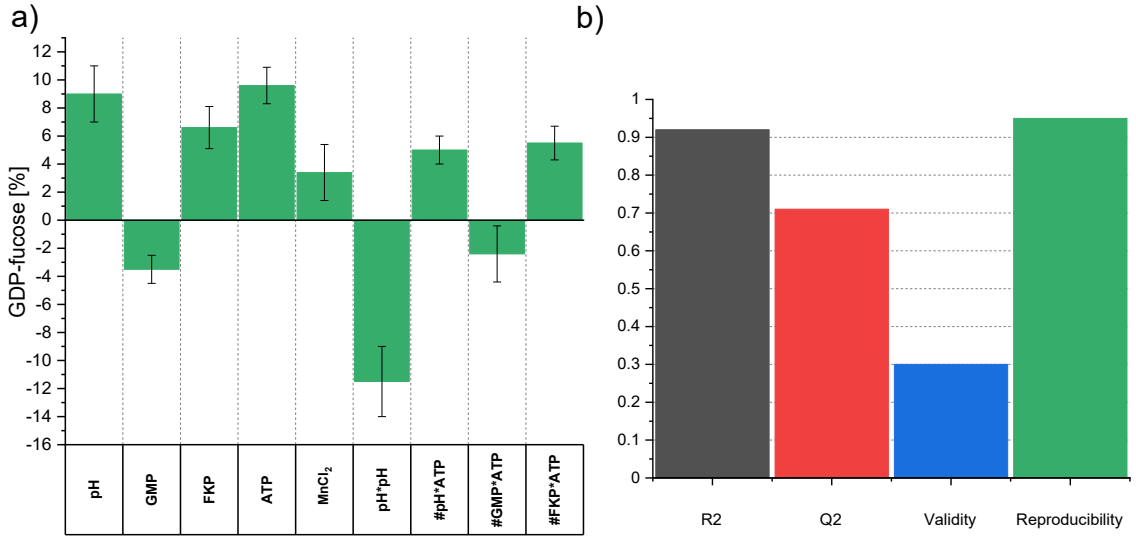


Figure 43 - DoE model evaluation for the GDP-fucose enzymatic cascade (2σ). Sensitivity analysis and model evaluation. a) Effect graph for each factor. Green bars represent the magnitude of a positive or negative influence on the response, while the error bars represent the significance of the response. Single-factor terms, interaction, and quadratic terms are included for each factor. b) Summary of fit. R^2: goodness of fit; Q^2: prediction precision; model validity: a test for diverse model problems; reproducibility: variation within replicates. N=54, DF=40. (#) confounded interaction. Model fit and analysis were performed by the built-in tools in the software package MODDE® 13.

The model was used to identify a combination of parameters that would yield the maximum conversion yield of the initial substrate GMP to GDP-fucose. The built-in optimiser from MODDE® was used for this purpose, which calculated various set points in which the yield could be close to 100% (see Figure 44). After selecting for high initial substrate load the following conditions were chosen: T=32 °C, Tris/HCl 100 mM, (pH 9), 45 mM $MgCl_2$, 0.55 g/L enzymes (strain 6a lysate),

1.7 g/L FKP, 30 mM GMP, 30 mM fucose, 10 mM ATP, 22 mM PolyP, 10 mM $MnCl_2$.

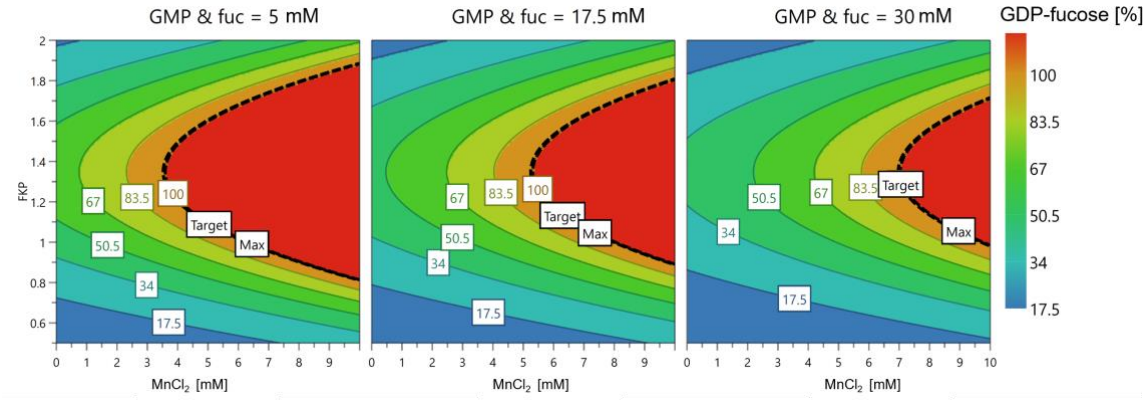


Figure 44 - Contour plot of GDP-fucose DoE model predictions. Values for pH 9, PolyP 22 mM , and ATP 5 mM are kept constant while the concentration of FKP [g/L] is varied over the y-axis, the concentration of $MnCl_2$ [mM] is varied over the x-axis, and the concentration of the initial substrates GMP & fucose [mM] is varied in three steps. The substrate conversion yield is represented as a heat map. The maximum possible output (100%) is represented with a black line.

A validation experiment was performed with the conditions mentioned earlier to compare the prediction with the actual behaviour of the cascade reaction. The reactions were set up in 1.5 mL centrifuge tubes with a reaction volume of 200 µL, performed in biological triplicate, with sampling conducted after the reaction began, at 3 h, and at the endpoint at 19 h. This was done to interfere as little as possible with the reaction since the DoE screening experiment was also performed with end-point sampling. The samples were analysed and quantified using HPLC-UV as described in the Materials and Methods section, and the resulting concentrations from the various intermediates in the reaction are shown in Figure 45.

It can be observed in the validation experiment that the predicted yield of 99% was not achieved. The final product titer after 19 h was approximately 22 mM, which corresponds to a Yp/s of 81.5 %. Concerning the guanosine-containing species, there are GTP, GDP, and GMP remaining at concentrations of less than 4 mM. It can also be observed that ATP is not fully consumed, with approximately 1 mM remaining measurable after 19 h, indicating that the reaction is not limited by ATP availability. Most of the ATP, however, is converted into AMP and accumulates to about 3.5 mM.

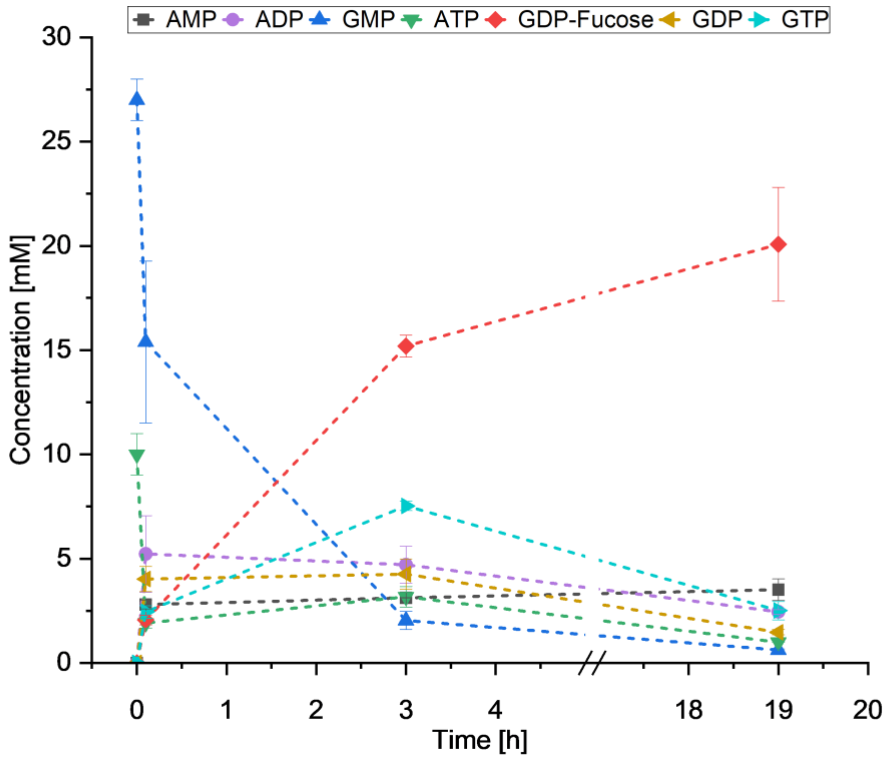


Figure 45 - GDP-fucose DoE validation experiment. Time course of HPLC-UV measurements displaying the concentration of reaction intermediates. The measurement at time point zero was taken as the negative control (no enzymes) at the beginning of the reaction. Error bars represent the standard deviation of biological triplicates.

4.3.2.6 Gram-scale Synthesis

A gram-scale synthesis of GDP-fucose was performed to prove the straightforward scalability of the reaction. The reaction was performed in a 500 mL Erlenmeyer flask with a working volume of 130 mL in a temperature-controlled incubator. The optimised conditions and behaviours elucidated from the DoE experiments were used. Several adjustments to the setup were made. For instance, the substrate load was set to 25 mM GMP because it was observed in previous experiments that GDP-fucose was not further produced over this threshold. A slight excess of fucose was added to further drive the conversion of GMP. Lastly, most components were added to the reaction mix as solids, rather than being added from concentrated liquid stocks. This more closely resembles the appearance of a large-scale industrial process. The reaction was performed under the following conditions: 32°C, Tris/HCl 100 mM (pH 9), 30 mM $MgCl_2$, 0.7 g/L enzymes (strain 6a lysate), 1.7 g/L FKP, 25 mM GMP, 27 mM fucose, 10 mM ATP, 22 mM PolyP, 10 mM $MnCl_2$. The reaction was sampled both immediately after starting and after 20 h of reaction time. The measured initial concentration of the substrate GMP was 23.65 mM, calculated as the average of the total guanosine-containing species in the measurements at T = 0 h and T = 20 h. The reaction mixture was harvested by centrifugation at 12,000 × g for 45 min to separate it from solids, such as precipitated proteins and salts, and stored at -20 °C in small aliquots.

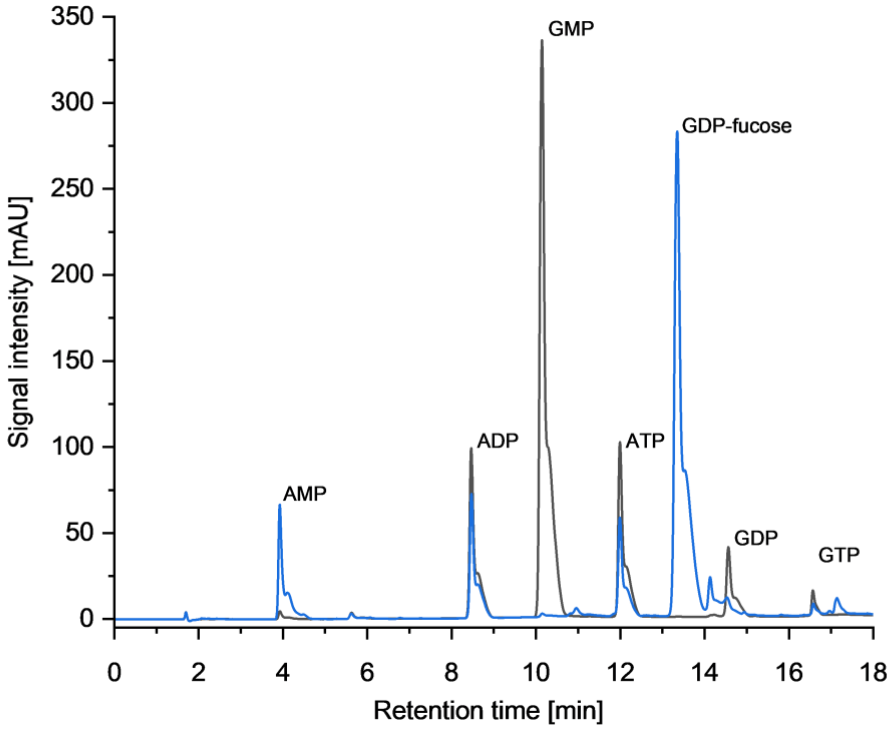


Figure 46 - Gram scale synthesis of GDP-fucose. HPLC-UV chromatogram from measurements of the enzymatic cascade at the beginning of the reaction (black) and after 20 h of incubation (blue).

After 20 h of incubation, the product titer was observed to be 23.13 mM, equivalent to a Yp/s of 97.8 % regarding GMP. The total amount of GDP-fucose synthesised was 1.76 g. The purity relative to all HPLC-UV measurable components was 68.46 %. As observed in Figure 46, ATP is rapidly converted into ADP. The sample T = 0 h was taken approximately one minute after the reaction began. A significant amount of GDP and GTP can also be observed at this early time point. In the sample taken after 20 h of reaction, it can be observed that almost all initial substrate GMP has been converted to GDP-fucose and trace amounts of GTP and GDP. It is also of notice in this measurement that ATP is not depleted (titer 2.8 mM), and accumulated amounts of ADP (3.9 mM) and AMP (3.3 mM) are observed. Finally, a slight shoulder is observed in every peak of this measurement; this was later attributed to a measurement artefact produced by the chromatography column and was resolved by performing a column cleaning as described in the technical manual for the CarboPac PA200® provided by Thermo Fisher Scientific.

4.3.2.7 Cascade Reaction as a Fed-batch

An experiment was conducted to determine if the stepwise addition of substrates would yield higher GDP-fucose titers. The setup had the following conditions: 100 mM Tris/HCl (pH 9), 30 mM $MgCl_2$, 1.57 g/l FKP, 0.767 g/L enzymes (strain 6a lysate), 20 mM GMP, 23 mM fucose, 10 mM ATP, 22 mM PolyP, 5 mM $MnCl_2$. The reaction was carried out in 1.5 mL centrifugal tubes with an initial working volume

of 200 µL, and 5 µL samples were taken at 8 distinct time points. The feeding strategy consisted of adding 27 µL of a concentrated mix with GMP, ATP and fucose after 2 and 4 h of reaction time; the volume and intermediate concentrations over time are shown in Figure 47.

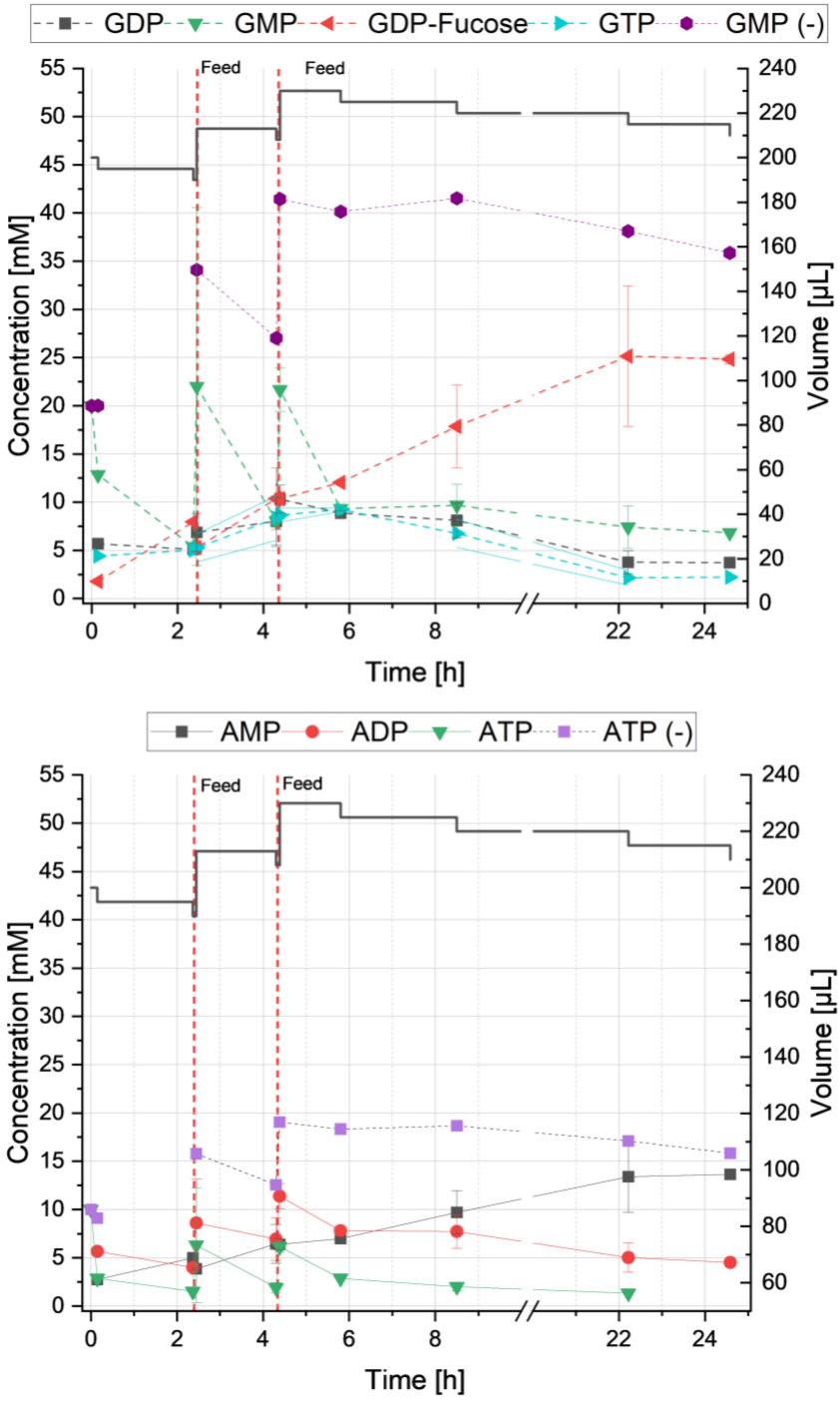


Figure 47 - GDP-fucose fed-batch reaction. The HPLC-UV time course measurements are shown for guanosine-containing components (top) and adenosine-containing components (bottom). Negative controls were performed by adding water instead of the equivalent volume

of biocatalyst, and the results for GMP and ATP are shown in shades of purple. Error bars represent the standard deviation of biological triplicates.

It can be observed that the negative control accumulates up to 41 mM of GMP and 19 mM ATP, while 25 mM of GDP-fucose is produced in the fed-batch reactions with enzymes added. The final substrate conversion yield was taken in this setup as the final titer of GDP-fucose measured after 24 h relative to the final amount of GMP measured in the negative control. With this in mind, the final Yp/s was 60.9 %. Considerable amounts of the intermediates GTP and GDP are accumulated at the end of the reaction. As for the ATP added to the reaction, it could be observed that most of it is converted to AMP and small amounts of ADP. No measurable amounts of ATP were found in the reaction after 24 h.

4.3.2.8 Enzyme Mining for New FKP Candidates

Even if the biocatalyst blending approach proved efficient for synthesising GDP-fucose, it required a separate fermentation and purification step of FKP. The author's hypothesis was that if a better-performing variant of FKP were identified, the co-expression approach would be sufficient, eliminating the need for biocatalyst blending. Enzyme mining was performed to identify alternative FKP enzymes. Higher activity or better solubility were chosen as the desired characteristics. Through database homology search, nine candidates were found from organisms varying from human gut microorganisms to protist parasites. The proteins selected varied widely in coverage and homology to the FKP from *Bacteroides fragilis* (bfFKP). A summary of the selected candidates and their comparison to bfFKP is shown in Table 12. The sequence coverage is often very close to 100% in organisms that are phylogenetically close to *Bacteroides,* such as in *Tannerellaceae bacterium* and *Phocaeicola vulgatus.* The coverage and identity are lower with smaller proteins like those from *Microcystis aeruginosa* and *Caudoviricetes sp..* However, these were still considered candidates because of their small but significant homology with the FKP conserved domains.

Table 12 – Selected FKP candidates. Sequences found through enzyme mining in databases are compared to bfFKP, the widely studied variant originating from *Bacteroides fragilis*. The coverage and identity values relative to bfFKP were calculated using the multiple sequence alignment tool COBALT® from NCBI [250].

FKP	Organism	MW [kDa]	bfFKP coverage	Identity with bfFKP
bfFKP	*Bacteroides fragilis*	105.66	--	--
maFKP	*Microcystis aeruginosa*	37.974	35.62%	22.16%
tgFKP	*Trypanosoma grayi*	116.045	100%	46.58%
cvFKP	*Caudoviricetes sp.*	44.042	42.47%	81.82%
naFKP	*Novipirellula aureliae*	106.999	99.58%	41.38%

FKP	Organism	MW [kDa]	bfFKP coverage	Identity with bfFKP
pcFKP	*Paraprevotella clara*	105.635	99.47%	66.07%
ldFKP	*Leishmania donovani*	126.092	99.68%	35.07%
tbFPK	*Tannerellaceae bacterium*	105.101	100%	72.77%
pjFKP	*Parabacteroides johnsonii*	104.822	99.26%	69.53%
pvFKP	*Phocaeicola vulgatus*	105.157	99.47%	69.19%

*MW – molecular weight.

Each of the nine candidates was cloned into a pET-15b vector by BioCat GmbH and then transformed into an *E. coli* BL21 strain following the protocol described in the Materials and Methods. One fermentation was performed with each of the variants to express the recombinant proteins, and an experiment was performed using cell lysates containing the FKP variants. The experiment´s conditions were the following: 100 mM Tris/HCl (pH 9), 10 mM $MgCl_2$, 0.1 g/L protein from cell lysate, 10 mM GTP, 10 mM fucose, 10 mM ATP, 10 mM $MnCl_2$ at 32°C in 1.5 mL centrifugal tubes with a total reaction volume of 200 µL.

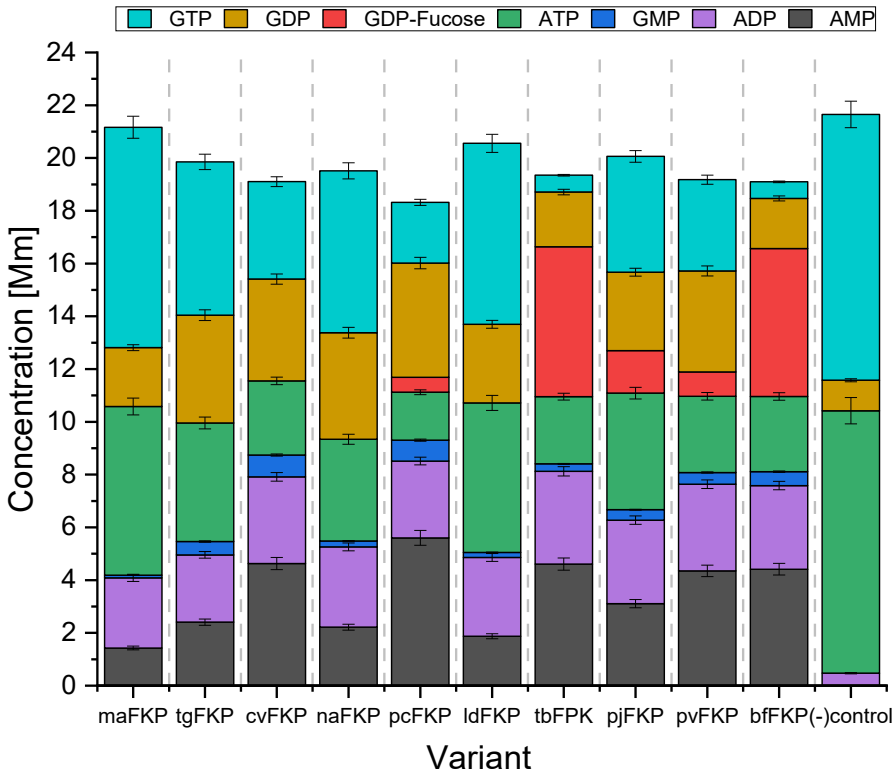


Figure 48 - Enzymatic activity screen for FKP variants. Measurements were performed with HPLC-UV after 24 h of incubation. The original variant (bfFKP) was added to the screen as a benchmark and a reaction mix without enzymes was used as a negative control.

As can be observed in Figure 48, only four of the initial nine FKP variants showed GDP-fucose synthesis activity in our enzyme activity assay. Three of them have a lower activity than the benchmark variant bfFKP, and only one of them had similar activity: tbFKP. An important observation is that all the FKP variants that showed a degree of GDP-fucose synthesis activity belong to the same organism class (*Bacteroidia*) as bfFKP, see Figure 49. The FKP enzyme from the *Tannerellaceae bacterium* is of particular interest because it showed similar activity to the widely used bfFKP but has only 72.77 % sequence identity, which makes it an interesting target for protein engineering in the future.

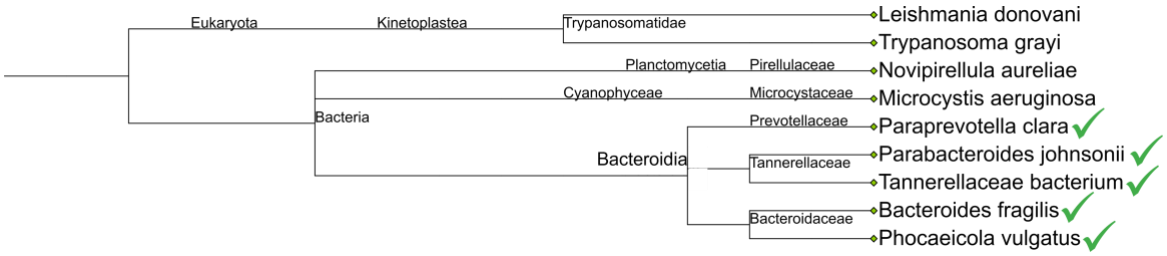


Figure 49 - Phylogenetic tree of selected organisms with FKP enzymes. The super-kingdom, class, family, and species are noted. Green check marks represent the species with identified functional FKP enzymes. The phylogenetic map was created using the PhyloT v2 tool [251].

4.3.3 Discussion

Individually purified enzymes

The enzymatic cascade designed for the synthesis of GDP-fucose consisted of four enzymes: GMK, PPK3, FKP, and PPA. Conveniently, the enzymes PPK3 and PPA are shared with the enzymatic cascades for the production of CDP-glycerol and CMP-Neu5Ac. The remaining two enzymes (GMK and FKP) were included exclusively for the synthesis of GDP-fucose.

The enzyme GMK could be expressed and purified with 68.36 mg of total protein from a 200 mL bacterial culture. The overexpression of *E. coli* GMK was performed back in 1993 by Gentry and colleagues[203], who successfully obtained 14 mg from a 1 L bacterial culture. This amount of protein is significantly lower when compared to the amount that could be purified in this work. However, It is expected that even if it is the same enzyme being expressed in the same host, the advancement in fermentation and purification technology over the last decades would allow for higher yields. Another study reports the overexpression and purification using very similar methods (T7 expression system and affinity chromatography) but for a guanylate kinase from a different organism, *Bruyia Malayi,* the final yield of purified enzyme reported there was 15 mg in a 1 L bacterial culture[252]. With this in mind, the expression and purification method used in this work is well-suited for the production of the GMK enzyme from *E. coli.*

As for FKP, 32.1 mg per 200 mL of bacterial culture could be obtained in this work. The *Bacteroides fragilis* FKP is a thoroughly characterized enzyme with available cryo-EM structure[202] and used in synthesising GDP-fucose[33], its derivatives [133], and other guanosine sugar-nucleotides[253]. It has been reported that a yield of 10 mg/L of the pure enzyme was obtained from an *E. coli* bacterial culture[248], for an FKP originating from a different organism, *Vitis vinifera*. Although from a different organism, this yield is of the same magnitude as the yield obtained in this work. Other process characteristics also align with our experience working with the enzyme, such as the importance of low-temperature induction for a well-folded, soluble enzyme. This characteristic may be most likely due to the large molecular weight of the enzyme (105 kDa).

The purified enzymes GMK, PPK3, PPA and FKP could be used for a baseline experiment, and the successful synthesis of GDP-fucose could be observed. In this baseline experiment, a Yp/s of 52.6 % was obtained, equivalent to a GDP-fucose titer of 7.6 g/L. This titer almost doubles the one obtained in the experiments described by Mahour in 2021[32], showing a clear opportunity for improvement through process intensification strategies.

Kinetic modeling

The amount of each enzyme used is a critical factor for the dynamics of the enzymatic cascade, and there have been many advances in metabolic engineering research to up- or down-regulate genes to control the flux of a particular metabolic network to obtain a particular product[254]. In the same way, working with enzymatic cascades –in vitro or not— differs from biocatalysis using a single enzyme in that the balance between enzyme amounts and activities must be considered. An enzyme kinetic modelling approach was used to address this matter and fine-tune the amounts of each enzyme required for the reaction to solve a specific

optimisation problem, namely, maximising product yield. To this end, a kinetic model based on Michaelis-Menten kinetics was built, and the parameters (K_m, K_{cat}, K_i, and K_{eq}) fit the experimental data obtained from the proof-of-concept experiments shown in section 4.3.2.1. The value of kinetic parameters available in the literature was used as a starting point for the model fit. An interesting question arises when these same parameters end up with very different values after the model fit: where the kinetic parameters found by other researchers wrong? The answer is probably not, but enzyme kinetic parameters are often obtained by using widely different experimental conditions in the various research groups that may characterise an enzyme. This leads to an intrinsic uncertainty when dealing with reaction conditions that differ from the ones used by others. The most common approach to this challenge is to characterise each enzyme in-house under the reaction conditions one intends to use. This approach is, however, very resource and time-intensive. We proposed a methodology for modelling and optimising enzymatic cascades that aim to work with this uncertainty in mind. It was observed that the initial substrate GMP and the final product GDP-fucose had a very good and consistent fit to the experimental data across all 100 predictions. In contrast, the intermediate compounds like GDP and ADP had a less precise fit. The two compounds, fucose and fucose-1-phosphate, had the greatest spread; this was expected as neither of these compounds was measured in the experimental set due to detection limitations (fucose is not UV-active at the wavelength used in our HPLC-UV method).

After one round of kinetic modelling and optimisation, the best-performing solution was selected as enzyme concentrations of 0.057 g/L GMK, 0.32 g/L PPK3, 1.57 g/L FKP and 0.39 g/L PPA. It was notable that all the solutions exhibited a particular pattern, in which FKP was increased at the expense of the other enzymes; this behaviour made it possible to identify FKP as the bottleneck of the enzymatic cascade. The chosen solution was then experimentally validated, and the product titer resulted in 19.86 mM of GDP-fucose, significantly higher than the value of 13.16 mM obtained in the baseline experiment and even the value predicted by the model of 18 mM. The model prediction underestimated the final yield of the enzymatic cascade. When examining substrate conversion yield, an improvement from 51.2% to 90.3% was achieved by utilising the kinetic modelling method. Considering that very few experiments (n=5) were necessary for this outcome, this methodology effectively improved the reaction setup and has the potential to be used in other enzymatic cascades to quickly optimise the enzyme concentrations. That being said, this study was limited to enzyme concentrations and incorporating other factors like temperature, pH and co-factor concentrations might be very challenging. Additionally, individually purified enzymes had to be used to readily control their relative concentration in the reaction. This makes this methodology inconvenient when working with systems where controlling enzyme concentrations cannot be easily done. Examples include fermentative, whole-cell, and co-expression approaches, such as the pDuet vector strategy discussed in this work. Nevertheless, the insight into the cascade dynamics obtained was essential for further steps in the bioprocess intensification, most notably the identification of FKP as the bottleneck of the cascade.

Co-expression with the pDuet platform

To make the synthesis process more economically viable, it was decided to implement a co-expression strategy to produce the biocatalysts, simplify the overall process and reduce costs. The enzymatic cascade for GDP-fucose has the convenient characteristic of comprising only four enzymes, making it possible to fit the entire cascade into just two pDuet plasmids. A total of six different combinations of vectors and genes were created to screen for a strain that could effectively produce the four enzymes and synthesise GDP-fucose. A list with the configuration of each combination is presented in Table 10. The first combination was designed with the insight from the kinetic modelling in mind; FKP was the bottleneck of the enzymatic cascade and was required in higher quantities relative to the other enzymes. Theoretically, placing FKP in the first MCS of a high-copy-number plasmid would yield higher expression of that enzyme. Vector combination one consisted of the high-copy-number plasmid, pDuetRSF, with FKP in the first MCS and PPA in the second, and the enzymes PPK3 and GMK in a vector with a lower copy number, pDuet-CDF. The reasoning behind placing PPA in the same vector as FKP was that FKP is a very large enzyme (105 kDa), and it was decided to pair it with the smallest enzyme in the cascade, PPA (19 kDa); this was hypothesized to have a positive impact at the translation level. Combination 2 was not included in the screening because the required vector could not be synthethised. Combination 3 had a similar arrangement of genes as combination 1, but with a medium copy number, pDuetCDF, instead of the high copy number pDuetRSF. This was done because of the theory that using a high-copy-number plasmid can represent a burden for the cell metabolism and result in non-soluble or inactive protein production. Combination 4 design followed the same reasoning as Combination 3 but uses the plasmid pDuet-pET, which has an even lower copy number. Combinations 5 and 6 were designed without considering enzyme size, so the enzymes that are required in less quantity –GMK and PPA— were cloned into a pDuet-CDF plasmid, and the two most important enzymes –FKP and PPK3— were introduced into a moderate copy number plasmid, pDuet-pET, and a lower copy number plasmid, pDuet-ACYC. As observed in section 4.2.3, accurately predicting the expression of multiple enzymes with this co-expression platform is very challenging. With these different vector combinations, it was thought to account for the main variabilities that could impact efficient soluble and active enzyme expression, namely the copy number of plasmids, gene size, position of MCS and antibiotic resistance of the vector.

Using the high-throughput platform, Biolector® was convenient for screening all the strains resulting from transforming the different pDuet vector combinations into *E.coli.* During the screening of protein expression in the strains created for co-expression of the enzymes, it was observed that two distinct expression patterns emerged among all the strains. One pattern exhibited high levels of GMK and PPA, while showing very low levels of FKP and PPK3. Another pattern displayed lower levels of GMK but significantly higher levels of PPK3. Notably, in strain 6, a non-identified protein was also found at a higher concentration, exceeding that of PPK3 after IMAC purification. (see Figure 39). This protein is likely to be an endogenous protein from *E. coli*, characterised by a high percentage of histidine residues, which can become a contaminant during metal affinity chromatography. The most abundant native contaminant proteins from *E. coli* are SlyD (28 KDa), GlmS (67 KDa), ArnA (74 KDa) and carbonic anhydrase (25 KDa); but none of these match

the molecular weight of approximately 44 kDa that was observed in our protein gel. According to Bolanos-Garcia et al. [255], there is a possibility that this protein is an acetylornithine deacetylase (ArgE – 42.3 KDa) because it contains 4.4% of histidine residues; this is, however, not demonstrated in this work, and it was not further investigated. Two strains were selected for further testing: strain 4a, which showed an expression pattern with high GMK and PPA but low PPK3, and strain 6a, which showed an expression pattern with higher PPK3 but lower PPA and GMK. Purification by IMAC confirmed the presence of FKP, although it was not overexpressed to the same extent as when expressed individually. The author's hypothesis is that other recombinant proteins compete for the cell's translation machinery, and because of its considerably larger size (105 kDa), FKP cannot be overexpressed to higher amounts when co-expressed with smaller proteins. The low expression of FKP when co-expressed represents a challenge because FKP was identified as the bottleneck of the cascade.

The purified enzymes from strain 4a and strain 6a were then tested for enzymatic activity and their capacity to synthesise GDP-fucose from the initial substrates GMP and fucose. This experiment yielded a considerably low substrate conversion rate in both cases, which was expected from transferring the reaction from single-expressed enzymes to co-expression. From a reaction starting with 15 mM substrate load, roughly 26 % substrate conversion yield was obtained in both cases, with 4 mM product titer. Strain 6a achieved 4 mM of GDP-fucose in 24 h, while strain 4a reached this titer in 48 h. Although a lower substrate conversion yield was achieved using co-expression of enzymes when compared to single expression. The strain 6a was chosen to continue with the bioprocess intensification steps to optimise reaction conditions and increase the substrate conversion yield using co-expression.

Reaction engineering

Considerations were made to continue with further process intensification steps. First, the co-expressed enzymes were used as cell lysate, skipping the purification steps. This approach was proven to be viable for the CMP-Neu5Ac synthetizing cascade, and it was also directly applied to GDP-fucose. Next, additional purified FKP was added to the reaction in a biocatalyst blending approach. This approach was performed to compensate for the low expression of this enzyme when co-expressed. This is a step back from the goal of expressing all necessary enzymes with a single fermentation. However, this may ultimately not be feasible since FKP was poorly expressed in the co-expression trials. A compromise must be reached between product yield and the addition of extra FKP. Finally, a high substrate load is desired to offset the cost of the biocatalyst by achieving a high product/enzyme ratio. A substrate load screening (30–100 mM) was performed to investigate the effect of higher substrate loads (Section 0). In this experiment, it was observed that white solids formed when concentrations exceeding 30 mM of GMP were used, and that the substrate conversion yield decreased as the substrate load increased. The conversion achieved from 30 mM GMP/fucose was close to 50%; this substrate conversion yield is considerably lower than the 90% achieved when a substrate load of 22 mM was used in the kinetic modelling validation experiments. The formation of precipitates can be explained by the hydrogel formation of GMP when mixed with $MgCl_2$[247]. Guanosine has the characteristic of self-assembling into higher structures, particularly in the presence of divalent ions (see Figure 50).

This characteristic may explain why increasing the substrate load in the enzymatic cascade is so challenging. The enzymes involved in this cascade depend on ions such as Mg^{2+} and Mn^{2+} as essential cofactors. Consequently, there's an inherent limitation in the cascade reaction that does not occur in cascades designed for synthesising non-guanosine nucleotide sugars. Thus, the process intensification steps implemented for this cascade aimed at achieving a high substrate conversion yield with moderate substrate loads, defined as less than 30 mM of substrate.

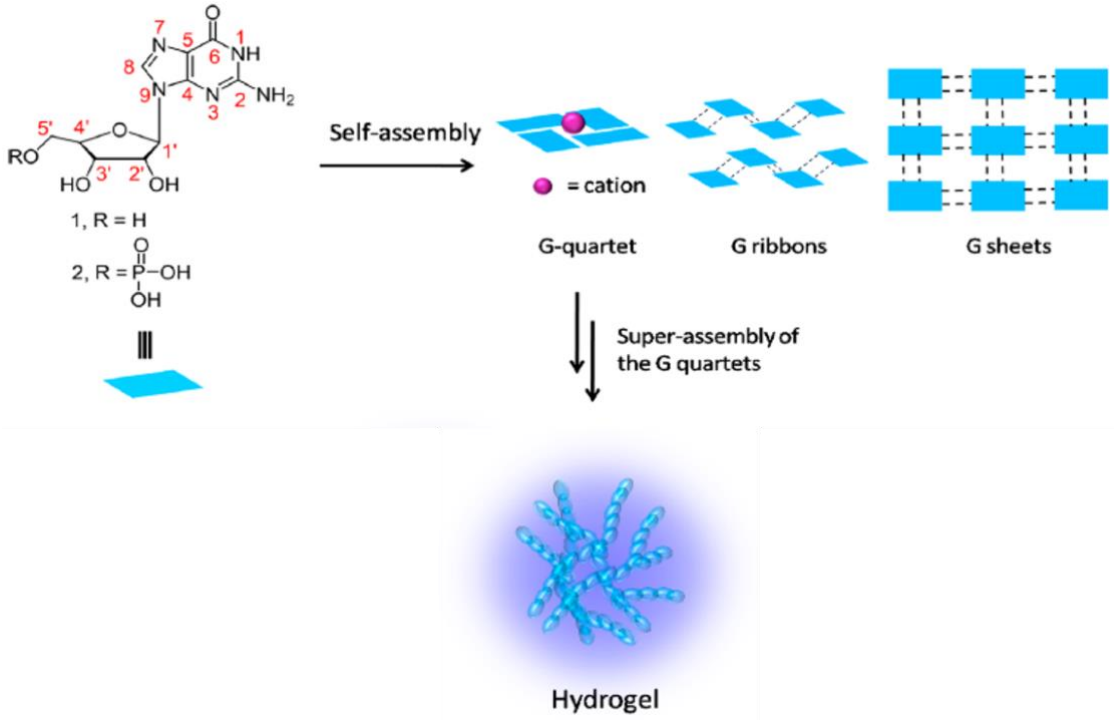


Figure 50 - Guanosine derivatives hydrogel formation. In the presence of ions, guanosine can coordinate with itself and form hydrogels, limiting the availability of enzymes to use guanosine-containing compounds as substrates. This figure was adapted from Bhattacharyya et al. 2018 [247], under the open access ACS AuthorChoice License agreement.

Design of Experiments

A setup was designed as a fractional-factorial IV and to include the nine factors: pH value, PolyP, $MgCl_2$, GMP/fuc, T °C, FKP, ATP, $MnCl_2$ and co-expressed protein in cell lysate. The first iteration of the model fit revealed that the factor with the highest impact was the pH value and a quadratic effect was identified for it. Furthermore, several factor interactions were also observed. Such as the interaction between pH and temperature, which was also identified in the design for the CMP-Neu5Ac enzymatic cascade. This interaction can be explained by the use of TRIS buffer, which is sensitive to temperature. Interactions between pH-PolyP, pH-$MgCl_2$, and PolyP-$MnCl_2$ were identified; however, these findings are confounded. Therefore, it cannot be confidently determined whether the observed effects stem from these combinations or from other factors. A design incorporating additional experimental assays would be necessary to resolve these interactions. A model validity of -0.2 was calculated for this design, indicating potential lack-of-fit issues. For this reason, a second fit was performed by reducing the outlier threshold to 2 σ ; ten outliers were identified and excluded from the design. The

model validation metrics fell within the acceptable range for this model, although it was less effective in identifying the effects of factors and interactions compared to the first model fit. Removing data points from the design results in a better-fitting model, but it deviates from reality, which limits the model's predictive power. The lack of fit issues may stem from a flaw in the experimental design; selecting a more suitable design could lead to a better-fitting model, though this remains uncertain.

Even though the model's prediction power might be limited, an optimisation round using the built-in optimiser tool was performed and validated experimentally. The model predicted a 99 % substrate conversion yield, and the experimental validation yielded 81.5 %. Given the expected limitations of the model's predictive capabilities, the difference between the predicted and measured outcomes was anticipated. When compared to previous experiments using co-expressed enzymes, particularly the strain screening in section 4.3.2.3, where only 26.6 % conversion was achieved with a 15 mM substrate load, there is still an improvement and the reaction using co-expressed enzymes was increased from a titer of 4 mM with 26.6 % Yp/s to a titer of 22 mM with 81.5 % Yp/s.

Gram-scale synthesis

The goal of producing GDP-fucose at the gram scale was achieved by leveraging incremental improvements from kinetic modelling, co-expression platform development, reaction engineering, and DoE screening. GDP-fucose has seldom been produced at the gram scale, and it is one of the most expensive nucleotide sugars. Following the process intensification steps, it became evident that the slow enzymatic activity of FKP, combined with the challenges associated with using guanosine as a substrate for divalent-ion-dependent enzymes, may be the reasons behind the commercial unavailability of this nucleotide sugar. Nevertheless, recent efforts like the repeated batch synthesis by Frohnmayer and colleagues[33] and the high-concentration multi-enzymatic cascade by Li and colleagues[130] have reported a GDP-fucose synthesis of up to 1.6 g. In this work, an amount of 1.76 g could be synthesised with 97.8 % Yp/s, which will allow for the further development of a process with low waste and simplified downstream processing. The reaction was also performed using cell lysates of co-expressed enzymes, a strategy that will significantly reduce costs in the future.

Fed-batch

A fed-batch approach was employed to investigate whether a stepwise addition of substrates could yield higher titers. The production of GDP-fucose above a titer of approximately 25 mM was not observed in any of the batch reactions conducted in this work; in reality, a higher substrate load resulted in lower titers of GDP-fucose (Section 0). It was hypothesised that a fed-batch setup could overcome substrate inhibition and guanosine hydrogel formation by preventing high concentrations of the substrate GMP. The feeding strategy consisted of providing substrate in three steps: at the beginning of the reaction, at two hours of incubation and at four hours of incubation. This timing was chosen based on the observation that most of the conversion of an initial 20 mM GMP amount occurred within the first three hours at the optimised reaction conditions used for the gram-scale synthesis in Section 4.3.2.6.

The GDP-fucose titer obtained at the end of the incubation period of 24 h was 25 mM. The additional substrate fed to the reaction accumulated in other intermediates (GDP, GTP) or side products (ADP-fucose), resulting in a final Yp/s of 60.9%. These results suggest that there is no immediate benefit of a fed-batch reaction over a batch reaction. Furthermore, the final titer of GDP-fucose appears to be capped at 25 mm for the cascade reaction. The author does not discard the idea of a fed-batch or continuous reaction setup in the future, but it has not proven to be beneficial at this stage of development.

Enzyme mining

A noteworthy comparison between the cascade reactions for synthesising CMP-Neu5Ac and GDP-fucose is that the former exhibited highly favourable kinetics when the enzymes were co-expressed. Conversely, the latter demonstrated the opposite behaviour. One theory to explain such a difference is that enzymatic cascades require the equilibrium of the whole system to flow towards the product, and they are often controlled by a bottleneck enzyme[256]. Occasionally, rapid formation of the final product may prevent the accumulation of other intermediates, making the last enzyme in the cascade particularly important. for the CMP-Neu5Ac cascade, the last enzyme, CSS, can be readily produced in soluble form. In contrast, AGE, the first enzyme in the cascade, exhibits poor solubility when over-expressed in E. coli. In contrast, the enzymatic cascade for GDP-fucose shows low expression of the enzyme FKP, whereas the first enzyme, GMK, is readily expressed in soluble form. As it is very challenging to control the relative concentration of each enzyme when co-expressing them with the pDuet vector system, some cascade designs may perform better than others depending on how well the enzymes are expressed in soluble form. This challenge can be approached using enzyme engineering methods to enhance the activity and solubility of enzymes that need it. Unfortunately, enzyme engineering presented a significant challenge in itself and was not within the scope of this thesis. On the other hand, enzyme mining is a resource-efficient approach that can identify better biocatalysts from sources already provided by nature, which have been sequenced by scientists and stored in public databases. During the kinetic modelling experiments, FKP was identified as the reaction bottleneck, and for the cascade to perform well, it was required in a very high concentration relative to the other three enzymes. FKP was required in a 28:1 proportion to GMK, a ratio that can be achieved by expressing each enzyme individually, but not when co-expressed in pDuet vectors. For this reason, FKP was identified as a target for finding an alternative enzyme that might exhibit better activity or expression qualities.

Through sequence alignment with bfFKP in public data repositories (NCBI, Uniprot), nine different potential FKP enzyme candidates were selected. Some of these sequences originated from organisms similar to *Bacteroides fragilis*, such as *Phocaeicola vulgatus*, while others came from quite different organisms, like *Leishmania donovani.* The nine candidates were recombinantly expressed in *E. coli,* and the cell lysates were used to screen for GDP-fucose synthetizing activity. The benchmark enzyme, bfFKP, was tested in parallel for comparison. Among the nine candidates, only four led to successful GDP-fucose synthesis: pcFKP, tbFKP, pjFKP, and pvFKP. Only one exhibited a comparable activity level to the original enzyme tbFKP. Interestingly, only the closer relatives of the original enzyme, specifically those belonging to the Bacteroides genus, showed positive GDP-

fucose synthesis activity. These gram-negative, strict anaerobe bacteria are usually found populating the human digestive tract[257], where they use fucosylated glycans for various functions. The author proposes that tbFKP can be used interchangeably with bfFKP for GDP-fucose synthesis; although it did not exhibit higher enzymatic activity, it shares only 72.7 % identity with tbFKP. This represents an opportunity in terms of commercial uses, such as securing freedom to operate, where other producers are already using bfFKP for a similar purpose.

In summary, the cell-free enzymatic cascade was developed from early-stage µg-scale proof-of-concept experiments using purified single enzymes to gram-scale reactions using crude cell lysates as biocatalysts, achieving a Yp/s up to 97.8% and producing 1.76 g of GDP-fucose. Challenges remain for the development of this enzymatic cascade. In particular, a biocatalyst blending strategy was required to offset the low fucose kinase activity of bfFKP; future research into FKP enzyme engineering may prove valuable in addressing this matter.

4.4 LNFP III

This chapter proposes an enzymatic cascade for synthesising the HMO LNFP III. By extending the enzymatic cascade for synthesising GDP-fucose with one enzyme, this reaction exemplifies the modular potential of HMO synthesis via cell-free enzymatic cascades. The system is first described, followed by the results and discussion on biocatalyst production and enzymatic cascade performance. This chapter was done with the support of Briggith Uribe, during her Master Thesis project.

4.4.1 An Enzymatic Cascade with In-situ Nucleotide Sugar Regeneration

The enzymatic cascade for producing LNFP III uses all the enzymes used for synthesising GDP-fucose, and one enzyme is added. This enzyme is the α1-3/4 fucosyltransferase from *Helicobacter pylori* (Hp3/4FT), which transfers fucose from the donor molecule GDP-fucose to an acceptor; in this case, the acceptor molecule is the tetrasaccharide LNnT, which is added directly to the reaction (Figure 51). It has been reported that Hp3/4FT accepts LNnT as a substrate, but may not be specific for the α1-3 linkage of the GlcNac, a side reaction resulting in the di-fucosylated product can occur when the concentration of GDP-fucose is too high[30]. For this reason, having control over the reaction is a critical factor in obtaining a pure product when using this enzyme. Once the fucose transfer reaction is finished, a GDP molecule is left over, which can be recycled back into the GDP-fucose synthetizing cascade. This allows the cascade to function with amounts of guanosine nucleosides that are lower than stoichiometric. When combined with the PPK3 regeneration system, the result is an enzymatic cascade that avoids using expensive phosphate nucleotides, such as ATP and GTP, by recycling them at the expense of PolyP.

The cascade can be set up in several ways, varying the biocatalysts in their purified or cell-lysate forms. The GDP-fucose synthesising module was used in the form of cell lysates from enzyme co-expressing strains, and the fucosyltransferase was used as a purified enzyme. In theory, through further development of this modular platform, several fucosyltransferases can also be co-expressed in a single strain, enabling the synthesis of even more complex HMOs in a single-pot reaction.

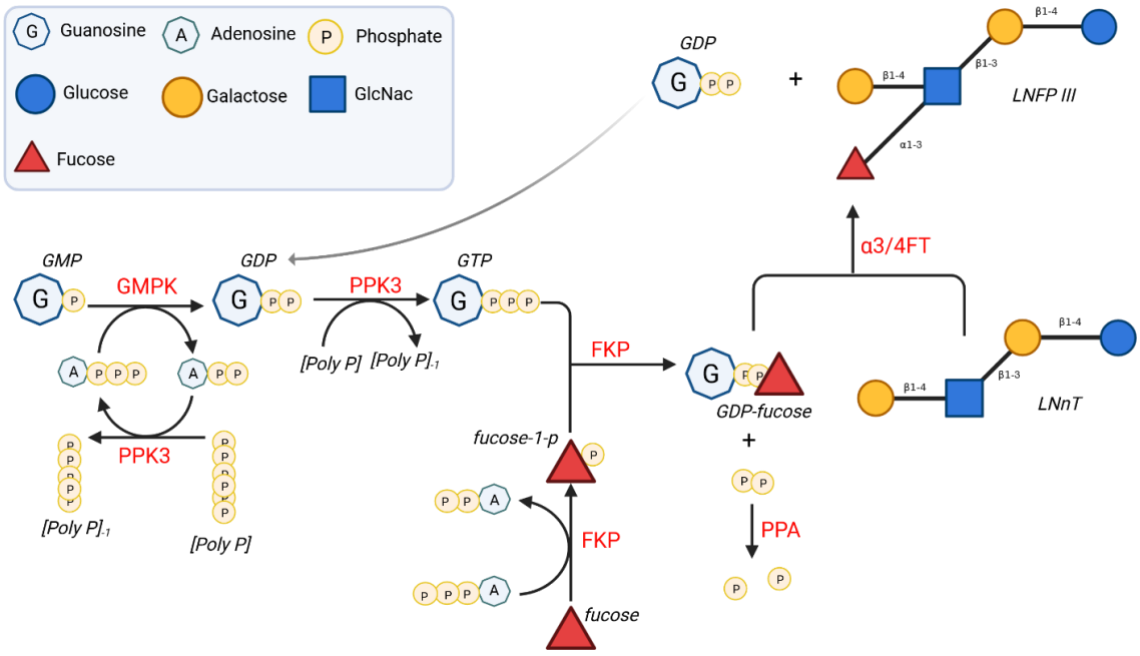


Figure 51 - Enzymatic cascade for the synthesis of LNFP III. Enzymes are depicted in red and components in black; enzymatic reactions are shown in black arrows weighted by size for preferred reaction direction when reversible.

4.4.2 Results

This section shows the recombinant expression and purification of the Hp3/4FT, followed by the cell-free enzymatic cascade results using GDP-fucose regenerating co-expresses enzymes and purified Hp3/4FT as biocatalysts.

4.4.2.1 Biocatalyst Production

The biocatalysts for this cascade were the enzymes involved in the GDP-fucose synthesis and a fucosyltransferase. As the co-expression of the GDP-fucose synthesising enzymes has already been discussed, this section will focus on the expression and purification results for Hp3/4FT.

The enzyme was recombinantly expressed in *E. coli* and purified as described in the Materials and Methods; no modifications to the standard method were necessary to obtain a pure and soluble protein. Hp3/4FT was identified by its theoretical molecular weight of 51.5 kDa by SDS-PAGE, revealing a highly pure protein solution (Figure 52). The purified enzyme was set to a concentration of 5 g/L and stored at -20 °C for later use. The total amount of protein obtained from a 200 mL bacterial culture was 15 mg.

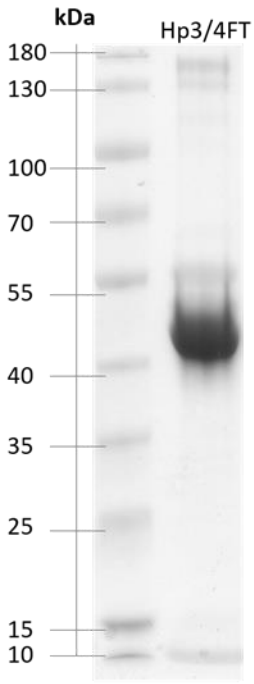


Figure 52 - SDS-PAGE of purified Hp3/4FT. A molecular weight marker was loaded into the first lane, and approximately 2 µg of protein was loaded into the second lane of a 10% polyacrylamide gel well. The gel was stained with Coomassie blue and scanned using a tabletop Epson scanner. The theoretical molecular weight of Hp3/4FT is 51.5 kDa.

4.4.2.2 Synthesis of LNFPIII

A proof-of-concept experiment was set up with the following conditions: 32 °C, 100 mM TRIS/HCl (pH 9), 45 mM $MgCl_2$, 0.1 g/L Hp3/4FT, 0.7 g/L protein of cell lysate from 6a strain, 1 g/L purified FKP, 25 mM GMP, 27 mM fucose, 5 mM ATP, 22 mM PolyP and 10 mM LNnT in 1.5 mL centrifugal tubes with a working volume of 200 µL. The reaction was sampled at several time points and analysed with HPLC-PAD for the detection of oligosaccharides and with HPLC-UV for the detection of UV-active components. To investigate the effect of additional FKP enzyme in the reaction, a second experiment was set up without extra FKP, substituting this volume in the reaction mix with water. Figure 53 shows the time course measurements of LNnT, LNFP III and Lacto-*N*-neodifucohexaose II (LNnDFH II), which is the difucosylated product from LNnT.

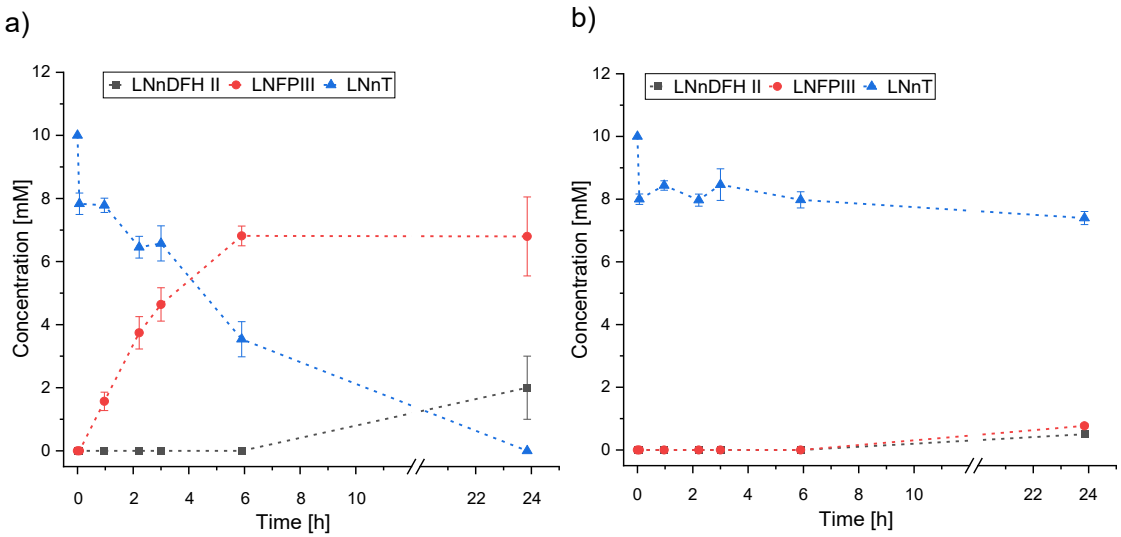


Figure 53 - Synthesis of LNFP III. Time course of HPLC-PAD measurements displaying the concentration of the substrate LNnT and the products LNFP III and LNnDFH II. A) Results from the enzymatic cascade with additional FKP, and b) results from the enzymatic cascade without the addition of extra FKP. The co-product LNnDFH II was identified, and its concentration values were estimated through mass balance analysis. Error bars represent the standard deviation of biological triplicates.

When the cascade was supplemented with additional purified FKP, the production of LNFP III could be confirmed after only 1 h of reaction. After 6 h, it reached a maximum of 6.8 mM and remained until 24 h. At the same time, only a small amount of product was observed after 24 h in the reaction without additional purified FKP. The production of the side product LNnDFH II was observed only after 24 h of reaction in both experiments.

UV-active intermediates were analysed for the reaction that showed LNFP III synthesis; these results are presented in Figure 54. It was observed that GDP-fucose was produced to a titer of 1.5 mM after 1 h of reaction and remained constant until the reaction was stopped after 24 h. At this point, the mass balance of identified components was measured to be considerably lower than that of the previous time points. To the right of Figure 54, the HPLC-UV chromatogram of this sample is shown. There, two unidentified signals can be observed, which can be attributed to nucleotide-containing side products in the reaction. The concentration of ATP is rapidly diminished from the initial 5 mM to roughly 1 mM after only 5 min, and it stays low until 6 h, where it is no longer distinguishable from the signal-to-noise ratio of the measurement.

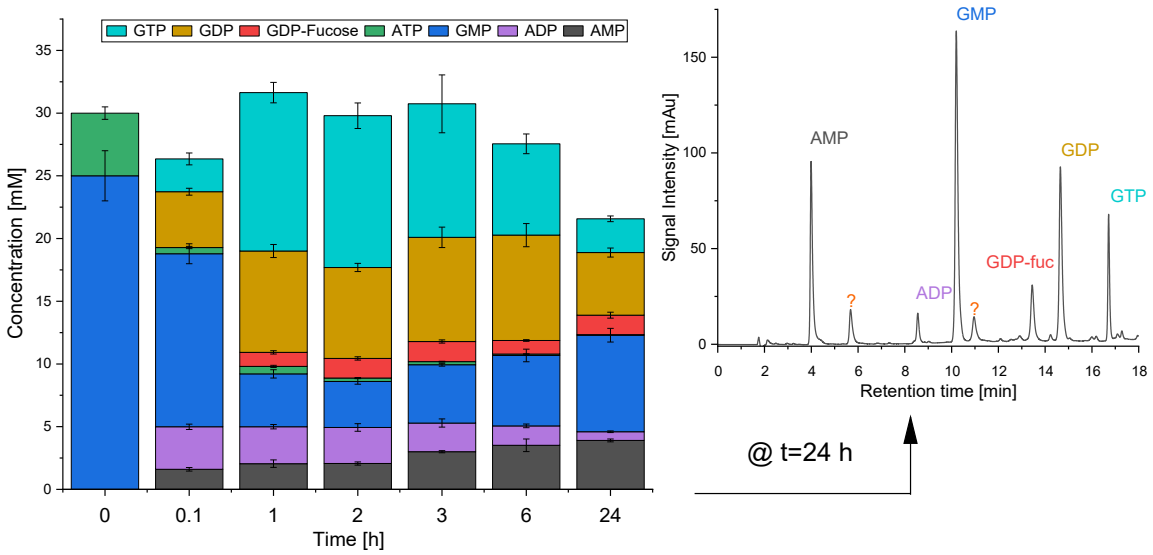


Figure 54 - Nucleotide intermediates in LNFP III synthesis. HPLC-UV measurements of LNFP III synthesis with Hp3/4FT, co-expressed 6a enzymes and extra FKP. Left – stacked column chart to visualise intermediate distribution and change, as well as mass balance. Right – Chromatogram of the sample at the last endpoint showing identified and non-identified signals. Error bars represent the standard deviation of biological triplicates.

The primary objective of this setup was to achieve the regeneration of GTP from GMP in a cyclical process. This characteristic was tested by reducing the initial amount of GMP to less than the stoichiometric requirement; specifically, a 1:2 ratio of GMP to LNnT was used. The experiment was set up with the following conditions 32 °C, 100 mM TRIS/HCl (pH 9), 45 mM $MgCl_2$, 0.1 g/L Hp3/4FT, 0.7 g/L protein of cell lysate from 6a strain, 1 g/L purified FKP, 5 mM GMP, 25 mM fucose, 5 mM ATP, 22 mM PolyP and 10 mM LNnT in 1.5 mL centrifugal tubes with a working volume of 200 µL. The reaction was sampled at several time points and analysed with HPLC-PAD for detecting oligosaccharides and with HPLC-UV for detecting UV-active components. The resulting intermediate concentrations are shown in Figure 55.

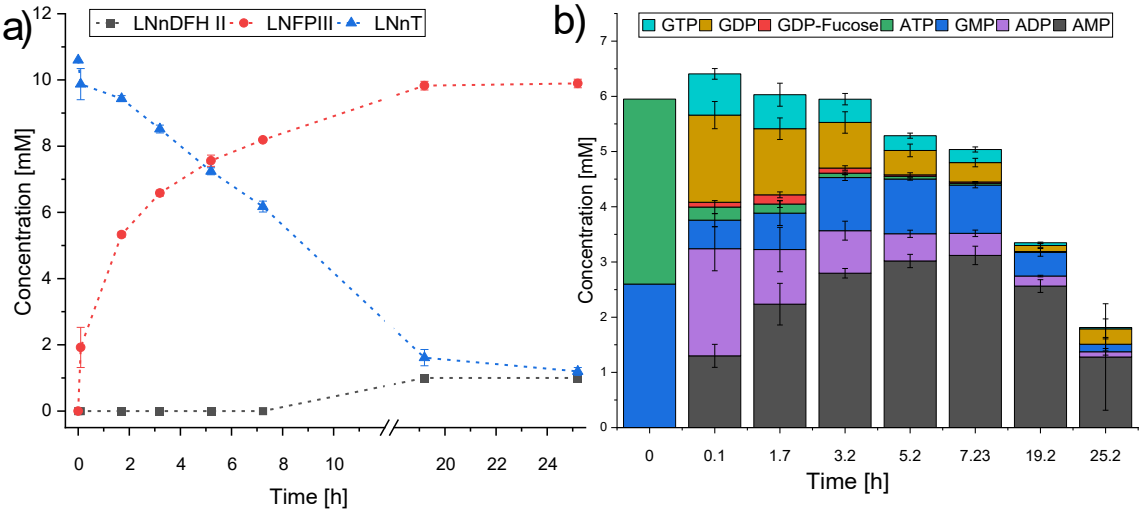


Figure 55 – Synthesis of LNFP III with GMP recycling. a) Time course of HPLC-PAD measurements, showing oligosaccharides LNFP III, LNnT, and LNnDFH II. b) Time course HPLC-UV measurements where the nucleotide-containing species are shown. Error bars represent the standard deviation of biological triplicates.

Considering the measured starting substrate concentration of 10.6 mM LNnT and a final product titer of 9.9 mM, the substrate conversion yield was 93 %. Since the final titer of LNFP III (9.9 mM) is higher than the initial molarity of GMP (5 mM) in the reaction, the recycling of the latter was confirmed. A small amount of the difucosylated product LNnDFH II was observed at the end of the reaction at a titer of 0-1 mM. The measurement of nucleotide sugar-containing species reveals that the concentration of GMP is not constant nor steadily decreasing, but quickly depleted at the start (0-2 h) and then increases, staying seemingly stable at a concentration of around 4.5 mM until 7 h, when the last sample for the day was taken. After 19 h, a concentration of 3 mM of GMP could still be measured but a significant decrease in the detected components was observed, seen as a mass balance deficit. This deficit can be attributed to various factors, such as the precipitation of components over time in the presence of increasing concentrations of molecular phosphate, and the formation of unidentified side products (see Figure 54).

4.4.3 Discussion

Biocatalyst production

The biocatalysts used for this reaction were those used for synthesising GDP-fucose plus one fucosyltransferase. The Hp3/4FT could be successfully recombinantly expressed and purified using standard methods. The total amount of protein that could be produced from a 200 mL bacterial fermentation was 15 mg. In comparison, a recent work by Bai and colleagues report obtaining 27 mg/L of culture of the same enzyme[145]. Considering that similar methods were used, the expression and solubility characteristics of the enzyme are consistent between laboratories.

The wild-type α1-3/4 fucosyltransferase from *Helicobacter pylori* has been previously characterised and studied by other research groups[258-260]; the enzyme is 478 aminoacids long and has a molecular weight of 56.07 kDa. For this work, however, a C-terminal-truncated version was used. The first truncated version was designed by Lin et al. in 2006[261] and is truncated by 66 amino acids, resulting in an enzyme with a molecular weight of 44.2 kDa. Several workgroups have used truncated enzymes to synthesise fucosylated oligosaccharides[133, 262-264]. Notably, it has already been applied in OPME cascade systems, as shown in work by Chen and colleagues in 2015[263]. The truncated version used in this work was designed by Yu et al. in 2017[30] and was truncated by 34 aminoacids for a calculated molecular weight of 51.5 kDa. Furthermore, this enzyme has been successfully used for the selective synthesis of LNFP III, as recently reported by Bai et al. in 2019[145], where it is emphasised that the selectivity of the fucosylation site is heavily dependent on the ratio of GDP-fucose to acceptor molecule in the reaction. In this study, only the single fucosylated product (LNFP III) was produced during the first 6 h of the reaction. At some point between 6-24 h of reaction time, the di-fucosylated product LNnDFH II began to be produced. This may begin to occur when all LNnT has been transformed into LNFP III and the regeneration cascade is still supplying GDP-fucose to the Hp3/4FT. Identifying this time sweet spot will be crucial for the selective production of LNFP III while minimising LNnDFH II as a side product.

Synthesis of LNFPIII

The analysis of UV-active components during the reaction revealed the presence of side products in small but significant concentrations. Based on previous experience with nucleotide-sugar synthesis cascades, the author suggests that these side products may include species such as ADP-fucose, AATP, or GGTP (Guanosine tetraphosphate). These unidentified species appear when the reaction is run for an extended period; for this reason, future efforts to develop this cascade should consider reaction time as a crucial factor to minimise the formation of such side products.

A key characteristic of the setup is that additional purified FKP is necessary to produce sufficient GDP-fucose to sustain the reaction. When the cell lysate of co-expressed enzymes was used alone, only a very small amount of LNFP III could be synthesised in 24 h. The need to add extra FKP is an additional cost that was intended to be avoided by applying the co-expression strategy. This represents an opportunity for process improvement that can be explored in future research.

Enzyme engineering may address this issue by offering a more active or better-expressed FKP variant.

The available literature on the gram-scale synthesis of LNFPIII is limited, as are the methods for its purification. Companies such as Nestlé, DSM, BASF, Kirin, and DuPont are in competition to develop efficient and cost-effective processes for synthesising a variety of HMOs. The growing commercial interest in the large-scale production of HMOs may explain why such research has not been made public.

In summary, the synthesis of LNFP III using an enzymatic cascade that regenerates ATP and GTP was described and validated. This opens up the possibility of using substrate concentrations lower than those required stoichiometrically. The author believes that this strategy will be used in the future for the inexpensive production of various HMOs, including complex fucosylated and sialylated structures. The modular characteristics of nucleotide sugar synthesising enzymatic cascades allow for adaptable design of extended cascades that can produce additional HMOs. Future developments will focus on creating efficient and cost-effective processes, particularly by utilising process intensification strategies as discussed in this study.

5 Conclusion and Outlook

This work originated from the need to develop processes that can reliably and cost-effectively source nucleotide sugars. This need arose as the Synthetic Biotechnology team at the MPI for DCTS explored the possibility of using glycosyltransferases in the synthesis of HMOs and in vitro glycosylation of therapeutic proteins, and the availability of nucleotide sugars was found to be a bottleneck. The objective of this work was to implement selected process intensification strategies —enzyme co-expression, DoE, kinetic modelling — in the OPME cascade reactions capable of synthesising CDP-glycerol, GDP-fucose, CMP-Neu5Ac and LNFP III. One of the objectives was to reduce process costs, e.g., by reducing the total amount of fermentations through enzyme co-expression, and to enhance the overall performance of each cascade, e.g., by optimising product titer for a higher product-to-biocatalyst ratio.

A novel enzymatic cascade was designed for the synthesis of CDP-glycerol, a valuable substrate for the enzymatic synthesis of teichoic acids in glycoconjugate vaccines. The product was identified by MALDI-TOF MS, making this the first report of an enzymatic cascade capable of synthesising CDP-glycerol. The Yp/s of the cascade was increased from 10 to 89 % through the screening of cascade reaction conditions using a DoE approach. The final product titer after a batch time of 24 h was 31.2 mM of CDP-glycerol. Considering that CDP-glycerol is not commercially available, even as a reference standard, the cascade presented here will be very useful for researchers and glycoconjugate vaccine producers alike to source it in a reliable manner. Further work should focus on developing a process, based on this cascade, that includes downstream processing to comply with regulations for use in vaccine manufacturing.

The nucleotide sugar CMP-Neu5Ac is of high interest because it is a substrate for synthesising acidic HMOs like DSNLT, whose absence in the diet of infants has been correlated to the devastating disease necrotising enterocolitis. Additionally, sialylation of mAbs is a post-translational modification that can enhance their therapeutic properties. The enzymatic cascade consisted of six enzymes that could be successfully co-expressed in a single *E. coli* strain, and the reaction setup was optimised through a DoE methodology. The process could be advanced from purified enzymes synthesising single-digit milligrams of product to a method using cell lysates, which can produce 6.4 g with a Yp/s close to 100%.

The fact that over 60 % of HMOs are fucosylated in mothers with a secretory phenotype makes the availability of GDP-fucose an essential factor in developing new baby formulas that are closer to natural milk. The most promising way to produce complex fucosylated HMOs is enzymatically, and this will never be possible on a large scale without a reliable method for producing GDP-fucose. The enzymatic cascade proposed in this dissertation consisted of four enzymes that could be successfully co-expressed in a single *E. coli* strain. Still, the bottleneck enzyme of the cascade could not be over-expressed in the necessary amounts to ensure a high substrate conversion yield. The complexity of controlling enzyme co-expression when using pDuet vectors made it difficult to achieve the optimal enzyme ratios that were discerned through kinetic modelling. To compensate for this, the co-expressed enzymes were blended with purified FKP. After applying

DoE methods and reaction engineering, a process capable of producing 1.76 g of GDP-fucose, with a Yp/s of 97.8 %, was achieved. It is important to note that the substrate load could not be increased past a concentration of 25 mM. Considering that the enzymatic cascade for the synthesis of CMP-Neu5Ac could be run with a starting concentration of CMP of 100 mM, there is reason to believe that the complex interactions between guanosine, $MgCl_2$ and polyP may be responsible for the challenges encountered when increasing the substrate load. The work presented here exemplifies how similar enzymatic cascades can have intrinsic and critical differences that pose challenges during process intensification. Future work on this cascade reaction may employ a different strategy to (1) eliminate the need for biocatalyst blending with extra FKP and (2) increase the substrate load or design an in situ product removal scheme.

The final chapter of this thesis demonstrates the use of the GDP-fucose enzymatic cascade as a nucleotide regeneration module. The fucosylated HMO LNFP III could be successfully synthesised by adding a glycosyltransferase to the nucleotide sugar generation cascade. Further work in this direction is developing this strategy by optimising the reaction conditions and including a suitable downstream processing method. In theory, the cascade reactions described here can be used to produce a variety of fucosylated and sialylated HMOs.

In conclusion, the steps for process intensification discussed in this dissertation are largely absent from the academic literature regarding their application to OPME cascade reactions. Strategies such as co-expression, kinetic modeling, and DoE optimisation are useful within the limitations discussed in this work and can be adapted for other enzymatic cascade projects. It can be said that the development of production processes is never truly complete, particularly when discussing specialty chemicals such as nucleotide sugars and HMOs. Nevertheless, this work signifies a step toward achieving a reliable and cost-effective process.

Further work on the enzymatic cascades described here is being done at eversyn GmbH, a spin-off company from the MPI, which aims to be a leading provider of nucleotide sugars and HMOs in the near future.

Bibliography

1. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3-49 (2017).
2. Lairson, L.L., Henrissat, B., Davies, G.J. & Withers, S.G. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555 (2008).
3. Palcic, M.M. Glycosyltransferases as biocatalysts. *Curr Opin Chem Biol* **15**, 226-233 (2011).
4. Holst, A.Q. et al. Infant Formula Supplemented with Five Human Milk Oligosaccharides Shifts the Fecal Microbiome of Formula-Fed Infants Closer to That of Breastfed Infants. *Nutrients* **15** (2023).
5. Gokhale, U.B., Hindsgaul, O. & Palcic, M.M. Chemical synthesis of GDP-fucose analogs and their utilization by the Lewis* A (1→ 4) fucosyltransferase. *Canadian Journal of Chemistry* **68**, 1063-1071 (1990).
6. Bych, K. et al. Production of HMOs using microbial hosts - from cell engineering to large scale production. *Curr Opin Biotechnol* **56**, 130-137 (2019).
7. Schelch, S., Zhong, C., Petschacher, B. & Nidetzky, B. Bacterial sialyltransferases and their use in biocatalytic cascades for sialo-oligosaccharide production. *Biotechnology Advances* **44**, 107613 (2020).
8. Jaroentomeechai, T. et al. Cell-Free Synthetic Glycobiology: Designing and Engineering Glycomolecules Outside of Living Cells. *Front Chem* **8**, 645 (2020).
9. Priem, B., Gilbert, M., Wakarchuk, W.W., Heyraud, A. & Samain, E. A new fermentation process allows large-scale production of human milk oligosaccharides by metabolically engineered bacteria. *Glycobiology* **12**, 235-240 (2002).
10. Zhai, Y. et al. Enhancing GDP-fucose production in recombinant Escherichia coli by metabolic pathway engineering. *Enzyme Microb Technol* **69**, 38-45 (2015).
11. Kim, S., Park, J., Jeong, H. & Park, Y.-S. Optimization and semi-continuous fermentation of gluco-oligosaccharide production with Weissella cibaria YRK005. *Food Science and Biotechnology*, 1-10 (2024).
12. Hodgman, C.E. & Jewett, M.C. Cell-free synthetic biology: thinking outside the cell. *Metab Eng* **14**, 261-269 (2012).
13. Alissandratos, A. In vitro multi-enzymatic cascades using recombinant lysates of E. coli: an emerging biocatalysis platform. *Biophys Rev* **12**, 175-182 (2020).
14. Christie, R.M. Why is indigo blue? *Biotech Histochem* **82**, 51-56 (2007).
15. Ladisch, M.R. & Kohlmann, K.L. Recombinant human insulin. *Biotechnol Prog* **8**, 469-478 (1992).
16. Li, C.J. Organic reactions in aqueous media with a focus on carbon-carbon bond formations: a decade update. *Chem Rev* **105**, 3095-3165 (2005).
17. Prelog, V. Chirality in chemistry. *Science* **193**, 17-24 (1976).
18. Martinez, R.F., Cuccia, L.A., Viedma, C. & Cintas, P. On the Origin of Sugar Handedness: Facts, Hypotheses and Missing Links-A Review. *Orig Life Evol Biosph* **52**, 21-56 (2022).
19. Bommarious, A. Biocatalysis 569-592 (2004).
20. Park, S.H. et al. Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnology and bioengineering* **102**, 988-994 (2009).
21. Zhao, M. et al. An Overview of Sugar Nucleotide-Dependent Glycosyltransferases for Human Milk Oligosaccharide Synthesis. *J Agric Food Chem* **71**, 12390-12402 (2023).
22. Figueroa, C.M., Lunn, J.E. & Iglesias, A.A. Nucleotide-sugar metabolism in plants: the legacy of Luis F. Leloir. *J Exp Bot* **72**, 4053-4067 (2021).
23. Prohaska, R. & Schenkel-Brunner, H. A simple and efficient method for the preparation of GDP-fucose. *Anal Biochem* **69**, 536-544 (1975).
24. Yamamoto, K. et al. Preparation of GDP-L-fucose by using microbial enzymes. *Agricultural and Biological Chemistry* **48**, 823-824 (1984).
25. Ichikawa, Y., Wang, R. & Wong, C.H. Regeneration of sugar nucleotide for enzymatic oligosaccharide synthesis. *Methods Enzymol* **247**, 107-127 (1994).
26. Ichikawa, Y., Sim, M.M. & Wong, C.H. Efficient Chemical Synthesis of Gdp-Fucose. *J Org Chem* **57**, 2943-2946 (1992).
27. Zhao, G., Guan, W., Cai, L. & Wang, P.G. Enzymatic route to preparative-scale synthesis of UDP-GlcNAc/GalNAc, their analogues and GDP-fucose. *Nat Protoc* **5**, 636-646 (2010).

28. Lee, W.H., Chin, Y.W., Han, N.S., Kim, M.D. & Seo, J.H. Enhanced production of GDP-L-fucose by overexpression of NADPH regenerator in recombinant Escherichia coli. *Appl Microbiol Biotechnol* **91**, 967-976 (2011).
29. Petschacher, B. & Nidetzky, B. Biotechnological production of fucosylated human milk oligosaccharides: Prokaryotic fucosyltransferases and their use in biocatalytic cascades or whole cell conversion systems. *J Biotechnol* **235**, 61-83 (2016).
30. Yu, H. et al. H. pylori alpha1-3/4-fucosyltransferase (Hp3/4FT)-catalyzed one-pot multienzyme (OPME) synthesis of Lewis antigens and human milk fucosides. *Chem Commun (Camb)* **53**, 11012-11015 (2017).
31. Wan, L., Zhu, Y., Li, W., Zhang, W. & Mu, W. Combinatorial Modular Pathway Engineering for Guanosine 5'-Diphosphate-l-fucose Production in Recombinant Escherichia coli. *J Agric Food Chem* **68**, 5668-5675 (2020).
32. Mahour, R., Marichal-Gallardo, P.A., Rexer, T.F.T. & Reichl, U. Multi-enzyme Cascades for the In Vitro Synthesis of Guanosine Diphosphate L-Fucose. *ChemCatChem* **13**, 1981-1989 (2021).
33. Frohnmeyer, H., Rueben, S. & Elling, L. Gram-Scale Production of GDP-β-l-fucose with Multi-Enzyme Cascades in a Repetitive-Batch Mode. *ChemCatChem* **14** (2022).
34. Peracchi, A. The Limits of Enzyme Specificity and the Evolution of Metabolism. *Trends Biochem Sci* **43**, 984-996 (2018).
35. McDonald, A.G., Mariethoz, J., Davey, G.P. & Lisacek, F. In silico analysis of the human milk oligosaccharide glycome reveals key enzymes of their biosynthesis. *Sci Rep* **12**, 10846 (2022).
36. Zeuner, B., Teze, D., Muschiol, J. & Meyer, A.S. Synthesis of Human Milk Oligosaccharides: Protein Engineering Strategies for Improved Enzymatic Transglycosylation. *Molecules* **24** (2019).
37. Frohnmeyer, H. & Elling, L. Enzyme cascades for the synthesis of nucleotide sugars: Updates to recent production strategies. *Carbohydr Res* **523**, 108727 (2023).
38. Mikkola, S. Nucleotide Sugars in Chemistry and Biology. *Molecules* **25** (2020).
39. Bode, L. Human milk oligosaccharides: every baby needs a sugar mama. *Glycobiology* **22**, 1147-1162 (2012).
40. Kuettel, S. et al. The de novo and salvage pathways of GDP-mannose biosynthesis are both sufficient for the growth of bloodstream-form Trypanosoma brucei. *Mol Microbiol* **84**, 340-351 (2012).
41. Wang, S., Zhang, J., Wei, F., Li, W. & Wen, L. Facile Synthesis of Sugar Nucleotides from Common Sugars by the Cascade Conversion Strategy. *J Am Chem Soc* **144**, 9980-9989 (2022).
42. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
43. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* **49**, D545-D551 (2021).
44. Corfield, A.P. & Berry, M. Glycan variation and evolution in the eukaryotes. *Trends Biochem Sci* **40**, 351-359 (2015).
45. Kightlinger, W., Warfel, K.F., DeLisa, M.P. & Jewett, M.C. Synthetic Glycobiology: Parts, Systems, and Applications. *ACS Synth Biol* **9**, 1534-1562 (2020).
46. Kunz, C. Historical aspects of human milk oligosaccharides. *Adv Nutr* **3**, 430S-439S (2012).
47. Dinleyici, M., Barbieur, J., Dinleyici, E.C. & Vandenplas, Y. Functional effects of human milk oligosaccharides (HMOs). *Gut Microbes* **15**, 2186115 (2023).
48. Bode, L. & Jantscher-Krenn, E. Structure-function relationships of human milk oligosaccharides. *Advances in Nutrition* **3**, 383S-391S (2012).
49. Manthey, C.F., Autran, C.A., Eckmann, L. & Bode, L. Human milk oligosaccharides protect against enteropathogenic Escherichia coli attachment in vitro and EPEC colonization in suckling mice. *J Pediatr Gastroenterol Nutr* **58**, 165-168 (2014).
50. Plaza-Diaz, J., Fontana, L. & Gil, A. Human Milk Oligosaccharides and Immune System Development. *Nutrients* **10** (2018).
51. Hauser, J. et al. Sialylated human milk oligosaccharides program cognitive development through a non-genomic transmission mode. *Mol Psychiatry* **26**, 2854-2871 (2021).
52. Masi, A.C. et al. Human milk oligosaccharide DSLNT and gut microbiome in preterm infants predicts necrotising enterocolitis. *Gut* **70**, 2273-2282 (2021).
53. Nidetzky, B., Gutmann, A. & Zhong, C. Leloir Glycosyltransferases as Biocatalysts for Chemical Production. *ACS Catalysis* **8**, 6283-6300 (2018).

54. Pressley, S.R., McGill, A.S., Luu, B. & Atsumi, S. Recent advances in the microbial production of human milk oligosaccharides. *Current Opinion in Food Science* **57** (2024).
55. Hill, D.R., Chow, J.M. & Buck, R.H. Multifunctional Benefits of Prevalent HMOs: Implications for Infant Health. *Nutrients* **13,** 3364 (2021).
56. Soyyilmaz, B. et al. The Mean of Milk: A Review of Human Milk Oligosaccharide Concentrations throughout Lactation. *Nutrients* **13,** 2737 (2021).
57. Bergquist, P.L., Siddiqui, S. & Sunna, A. Cell-Free Biocatalysis for the Production of Platform Chemicals. *Frontiers in Energy Research* **8,** 193 (2020).
58. Teshima, M., Willers, V.P. & Sieber, V. Cell-free enzyme cascades - application and transition from development to industrial implementation. *Curr Opin Biotechnol* **79**, 102868 (2023).
59. Dudley, Q.M., Karim, A.S. & Jewett, M.C. Cell-free metabolic engineering: biomanufacturing beyond the cell. *Biotechnology journal* **10**, 69-82 (2015).
60. Claassens, N.J., Burgener, S., Vogeli, B., Erb, T.J. & Bar-Even, A. A critical comparison of cellular and cell-free bioproduction systems. *Curr Opin Biotechnol* **60**, 221-229 (2019).
61. Pyser, J.B., Chakrabarty, S., Romero, E.O. & Narayan, A.R.H. State-of-the-Art Biocatalysis. *ACS Cent Sci* **7**, 1105-1116 (2021).
62. Zhang, Y.H.P. Production of biocommodities and bioelectricity by cell-free synthetic enzymatic pathway biotransformations: Challenges and opportunities. *Biotechnology and bioengineering* **105**, 663-677 (2010).
63. Lowell, A.N. et al. Chemoenzymatic Total Synthesis and Structural Diversification of Tylactone-Based Macrolide Antibiotics through Late-Stage Polyketide Assembly, Tailoring, and C-H Functionalization. *J Am Chem Soc* **139**, 7913-7920 (2017).
64. Zetzsche, L.E. & Narayan, A.R.H. Broadening the scope of biocatalytic C-C bond formation. *Nat Rev Chem* **4**, 334-346 (2020).
65. Yu, H., Yu, H., Karpel, R. & Chen, X. Chemoenzymatic synthesis of CMP-sialic acid derivatives by a one-pot two-enzyme system: comparison of substrate flexibility of three microbial CMP-sialic acid synthetases. *Bioorg Med Chem* **12**, 6427-6435 (2004).
66. Andexer, J.N. & Richter, M. Emerging enzymes for ATP regeneration in biocatalytic processes. *Chembiochem* **16**, 380-386 (2015).
67. Gilormini, P.A. et al. Improved workflow for the efficient preparation of ready to use CMP-activated sialic acids. *Glycobiology* **26**, 1151-1156 (2016).
68. Mahour, R. et al. Cell-Free Multi-Enzyme Synthesis and Purification of Uridine Diphosphate Galactose. *Chembiochem* **23**, e202100361 (2022).
69. Muthana, M.M. et al. Improved one-pot multienzyme (OPME) systems for synthesizing UDP-uronic acids and glucuronides. *Chem Commun (Camb)* **51**, 4595-4598 (2015).
70. Chen, Y. et al. One-pot three-enzyme synthesis of UDP-GlcNAc derivatives. *Chem Commun (Camb)* **47**, 10815-10817 (2011).
71. Mahour, R. et al. Establishment of a five-enzyme cell-free cascade for the synthesis of uridine diphosphate N-acetylglucosamine. *J Biotechnol* **283**, 120-129 (2018).
72. Errey, J.C., Mukhopadhyay, B., Kartha, K.P. & Field, R.A. Flexible enzymatic and chemo-enzymatic approaches to a broad range of uridine-diphospho-sugars. *Chem Commun (Camb)*, 2706-2707 (2004).
73. Rexer, T.F.T. et al. One pot synthesis of GDP-mannose by a multi-enzyme cascade for enzymatic assembly of lipid-linked oligosaccharides. *Biotechnol Bioeng* **115**, 192-205 (2018).
74. Li, L. et al. Efficient enzymatic synthesis of guanosine 5'-diphosphate-sugars and derivatives. *Org Lett* **15**, 5528-5530 (2013).
75. Mizanur, R.M. & Pohl, N.L. Phosphomannose isomerase/GDP-mannose pyrophosphorylase from Pyrococcus furiosus: a thermostable biocatalyst for the synthesis of guanidinediphosphate-activated and mannose-containing sugar nucleotides. *Org Biomol Chem* **7**, 2135-2139 (2009).
76. Partha, S.K. et al. Chemoenzymatic synthesis, inhibition studies, and X-ray crystallographic analysis of the phosphono analog of UDP-Galp as an inhibitor and mechanistic probe for UDP-galactopyranose mutase. *J Mol Biol* **403**, 578-590 (2010).
77. Caputi, L. et al. A one-pot enzymatic approach to the O-fluoroglucoside of N-methylanthranilate. *Bioorg Med Chem* **21**, 4762-4767 (2013).
78. Wagstaff, B.A. et al. Enzymatic synthesis of nucleobase-modified UDP-sugars: scope and limitations. *Carbohydr Res* **404**, 17-25 (2015).

79. Schultz, V.L. et al. Chemoenzymatic Synthesis of 4-Fluoro-N-Acetylhexosamine Uridine Diphosphate Donors: Chain Terminators in Glycosaminoglycan Synthesis. *J Org Chem* **82**, 2243-2248 (2017).
80. Kay, E., Cuccui, J. & Wren, B.W. Recent advances in the production of recombinant glycoconjugate vaccines. *NPJ Vaccines* **4**, 16 (2019).
81. Hilleman, M.R. Vaccines in historic evolution and perspective: a narrative of vaccine discoveries. *Vaccine* **18**, 1436-1447 (2000).
82. Astronomo, R.D. & Burton, D.R. Carbohydrate vaccines: developing sweet solutions to sticky situations? *Nat Rev Drug Discov* **9**, 308-324 (2010).
83. Duke, J.A. & Avci, F.Y. Immunological Mechanisms of Glycoconjugate Vaccines. *Acs Sym Ser* **1290**, 61-74 (2018).
84. Litschko, C. et al. A New Family of Capsule Polymerases Generates Teichoic Acid-Like Capsule Polymers in Gram-Negative Pathogens. *mBio* **9** (2018).
85. Micoli, F. et al. Glycoconjugate vaccines: current approaches towards faster vaccine design. *Expert Rev Vaccines* **18**, 881-895 (2019).
86. Oldrini, D. et al. Combined Chemical Synthesis and Tailored Enzymatic Elongation Provide Fully Synthetic and Conjugation-Ready Neisseria meningitidis Serogroup X Vaccine Antigens. *ACS Chem Biol* **13**, 984-994 (2018).
87. Findlow, H. & Borrow, R. Immunogenicity and safety of a meningococcal serogroup A, C, Y and W glycoconjugate vaccine, ACWY-TT. *Adv Ther* **30**, 431-458 (2013).
88. Litschko, C. et al. Mix-and-Match System for the Enzymatic Synthesis of Enantiopure Glycerol-3-Phosphate-Containing Capsule Polymer Backbones from Actinobacillus pleuropneumoniae, Neisseria meningitidis, and Bibersteinia trehalosi. *mBio* **12**, e0089721 (2021).
89. Baddiley, J., Buchanan, J.G. & Sanderson, A.R. Synthesis of Cytidine Diphosphate Glycerol. *J Chem Soc*, 3107-3110 (1958).
90. Badurina, D.S., Zolli-Juran, M. & Brown, E.D. CTP:glycerol 3-phosphate cytidylyltransferase (TarD) from Staphylococcus aureus catalyzes the cytidylyl transfer via an ordered Bi-Bi reaction mechanism with micromolar K(m) values. *Biochim Biophys Acta* **1646**, 196-206 (2003).
91. Brown, S., Santa Maria, J.P., Jr. & Walker, S. Wall teichoic acids of gram-positive bacteria. *Annu Rev Microbiol* **67**, 313-336 (2013).
92. Roseman, S., Distler, J.J., Moffatt, J.G. & Khorana, H.G. Nucleoside Polyphosphates. XI.1 An Improved General Method for the Synthesis of Nucleotide Coenzymes. Syntheses of Uridine-5', Cytidine-5' and Guanosine-5' Diphosphate Derivatives. *Journal of the American Chemical Society* **83**, 659-663 (1961).
93. Shaw, D.R. Pyrophosphorolysis and enzymic synthesis of cytidine diphosphate glycerol and cytidine diphosphate ribitol. *Biochem J* **82**, 297-312 (1962).
94. Alberto Alcala-Orozco, E. et al. A Cell-Free Multi-enzyme Cascade Reaction for the Synthesis of CDP-Glycerol. *Chembiochem* **24**, e202300463 (2023).
95. Gilbert, M. et al. The synthesis of sialylated oligosaccharides using a CMP-Neu5Ac synthetase/sialyltransferase fusion. *Nature Biotechnology* **16**, 769-772 (1998).
96. Pilatte, Y., Bignon, J. & Lambré, C.R. Sialic acids as important molecules in the regulation of the immune system: pathophysiological implications of sialidases in immunity. *Glycobiology* **3**, 201-218 (1993).
97. Gianchecchi, E., Arena, A. & Fierabracci, A. Sialic acid-siglec axis in human immune regulation, involvement in autoimmunity and cancer and potential therapeutic treatments. *International Journal of Molecular Sciences* **22**, 5774 (2021).
98. Zhang, C., Chen, J., Liu, Y. & Xu, D. Sialic acid metabolism as a potential therapeutic target of atherosclerosis. *Lipids in health and disease* **18**, 1-11 (2019).
99. Chia, S. et al. Enhancing pharmacokinetic and pharmacodynamic properties of recombinant therapeutic proteins by manipulation of sialic acid content. *Biomedicine & Pharmacotherapy* **163**, 114757 (2023).
100. Büll, C., Stoel, M.A., den Brok, M.H. & Adema, G.J. Sialic acids sweeten a tumor's life. *Cancer research* **74**, 3199-3204 (2014).
101. Jastrząb, P., Narejko, K., Car, H. & Wielgat, P. Cell Membrane Sialome: Sialic Acids as Therapeutic Targets and Regulators of Drug Resistance in Human Cancer Management. *Cancers* **15**, 5103 (2023).
102. Thurl, S., Munzert, M., Boehm, G., Matthews, C. & Stahl, B. Systematic review of the concentrations of oligosaccharides in human milk. *Nutr Rev* **75**, 920-933 (2017).

103. ten Bruggencate, S.J., Bovee-Oudenhoven, I.M., Feitsma, A.L., van Hoffen, E. & Schoterman, M.H. Functional role and mechanisms of sialyllactose and other sialylated milk oligosaccharides. *Nutrition reviews* **72**, 377-389 (2014).
104. Zhu, Y., Zhang, J., Zhang, W. & Mu, W. Recent progress on health effects and biosynthesis of two key sialylated human milk oligosaccharides, 3′-sialyllactose and 6′-sialyllactose. *Biotechnology Advances* **62**, 108058 (2023).
105. Zhu, Y. et al. Recent advances on 2'-fucosyllactose: physiological properties, applications, and production approaches. *Crit Rev Food Sci Nutr* **62**, 2083-2092 (2022).
106. Chen, G. et al. Glycosyltransferase from Bacteroides gallinaceum Is a Novel alpha-1,3-Fucosyltransferase that Can Be Used for 3-Fucosyllactose Production In Vivo by Metabolically Engineered Escherichia coli. *J Agric Food Chem* **70**, 1934-1942 (2022).
107. Zhu, Y., Cao, H., Wang, H. & Mu, W. Biosynthesis of human milk oligosaccharides via metabolic engineering approaches: current advances and challenges. *Current Opinion in Biotechnology* **78**, 102841 (2022).
108. Liu, J.L.C. et al. Overproduction of CMP-sialic acid synthetase for organic synthesis. *Journal of the American Chemical Society* **114**, 3901-3910 (1992).
109. Endo, T., Koizumi, S., Tabata, K. & Ozaki, A. Large-scale production of CMP-NeuAc and sialylated oligosaccharides through bacterial coupling. *Appl Microbiol Biotechnol* **53**, 257-261 (2000).
110. Ishige, K., Hamamoto, T., Shiba, T. & Noguchi, T. Novel method for enzymatic synthesis of CMP-NeuAc. *Biosci Biotechnol Biochem* **65**, 1736-1740 (2001).
111. Lee, S.G., Lee, J.O., Yi, J.K. & Kim, B.G. Production of cytidine 5'-monophosphate N-acetylneuraminic acid using recombinant Escherichia coli as a biocatalyst. *Biotechnol Bioeng* **80**, 516-524 (2002).
112. Song, J. et al. Reassembled biosynthetic pathway for a large-scale synthesis of CMP-Neu5Ac. *Marine Drugs* **1**, 34-45 (2003).
113. Nahalka, J., Wu, B., Shao, J., Gemeiner, P. & Wang, P.G. Production of cytidine 5'-monophospho-N-acetyl-beta-D-neuraminic acid (CMP-sialic acid) using enzymes or whole cells entrapped in calcium pectate-silica-gel beads. *Biotechnol Appl Biochem* **40**, 101-106 (2004).
114. Hamamoto, T., Takeda, S. & Noguchi, T. Enzymatic synthesis of cytidine 5'-monophospho-N-acetylneuraminic acid. *Biosci Biotechnol Biochem* **69**, 1944-1950 (2005).
115. Nahalka, J. & Patoprsty, V. Enzymatic synthesis of sialylation substrates powered by a novel polyphosphate kinase (PPK3). *Org Biomol Chem* **7**, 1778-1780 (2009).
116. Mahour, R. (2022).
117. de Vries, T., Knegtel, R.M., Holmes, E.H. & Macher, B.A. Fucosyltransferases: structure/function studies. *Glycobiology* **11**, 119R-128R (2001).
118. Miyoshi, E. et al. Fucosylation Is a Promising Target for Cancer Diagnosis and Therapy. *Biomolecules* **2**, 34-45 (2012).
119. Chung, S. et al. in MAbs, Vol. 4 326-340 (Taylor & Francis, 2012).
120. Shields, R.L. et al. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fc γ RIII and antibody-dependent cellular toxicity. *Journal of Biological Chemistry* **277**, 26733-26740 (2002).
121. Suzuki, E. et al. A nonfucosylated anti-HER2 antibody augments antibody-dependent cellular cytotoxicity in breast cancer patients. *Clinical cancer research* **13**, 1875-1882 (2007).
122. Yamane-Ohnuki, N. & Satoh, M. in MAbs, Vol. 1 230-236 (Taylor & Francis, 2009).
123. Ma, B., Simala-Grant, J.L. & Taylor, D.E. Fucosylation in prokaryotes and eukaryotes. *Glycobiology* **16**, 158R-184R (2006).
124. Orczyk-Pawiłowicz, M. & Lis-Kuberka, J. The Impact of Dietary Fucosylated Oligosaccharides and Glycoproteins of Human Milk on Infant Well-Being. *Nutrients* **12**, 1105 (2020).
125. Verkhnyatskaya, S.A. et al. Digestion, fermentation, and pathogen anti-adhesive properties of the hMO-mimic di-fucosyl- β -cyclodextrin. *Food & Function* **12**, 5018-5026 (2021).
126. Yu, Z.-T. et al. The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. *Glycobiology* **23**, 169-177 (2012).
127. Tonetti, M. et al. The metabolism of 6-deoxyhexoses in bacterial and animal cells. *Biochimie* **80**, 923-931 (1998).

128. Turnock, D.C., Izquierdo, L. & Ferguson, M.A. The de novo synthesis of GDP-fucose is essential for flagellar adhesion and cell growth in Trypanosoma brucei. *Journal of Biological Chemistry* **282**, 28853-28863 (2007).
129. Niittymäki, J. et al. Cloning and expression of murine enzymes involved in the salvage pathway of GDP-l-fucose: l-fucokinase and GDP-l-fucose pyrophosphorylase. *European journal of biochemistry* **271**, 78-86 (2004).
130. Li, S. et al. Gram-scale production of sugar nucleotides and their derivatives. *Green Chemistry* **23**, 2628-2633 (2021).
131. Koizumi, S. et al. Large-scale production of GDP-fucose and Lewis X by bacterial coupling. *Journal of Industrial Microbiology and Biotechnology* **25**, 213-217 (2000).
132. Albermann, C., Distler, J. & Piepersberg, W. Preparative synthesis of GDP-beta-L-fucose by recombinant enzymes from enterobacterial sources. *Glycobiology* **10**, 875-881 (2000).
133. Wang, W. et al. Chemoenzymatic synthesis of GDP-L-fucose and the Lewis X glycan derivatives. *Proc Natl Acad Sci U S A* **106**, 16096-16101 (2009).
134. Liu, T.W. et al. Functional expression of L-fucokinase/guanosine 5'-diphosphate-L-fucose pyrophosphorylase from Bacteroides fragilis in Saccharomyces cerevisiae for the production of nucleotide sugars from exogenous monosaccharides. *Glycobiology* **21**, 1228-1236 (2011).
135. Wang, W. et al. Cell-free enzymatic synthesis of GDP-L-fucose from mannose. *AMB Express* **9**, 74 (2019).
136. Gan, J., Cao, C., Stahl, B., Zhao, X. & Yan, J. Advances and challenges for obtaining human milk oligosaccharides: Extraction from natural sources and synthesis by intentional design. *Trends in Food Science & Technology*, 104203 (2023).
137. Tsai, T.-W. et al. Exploring the Synthetic Application of Helicobacter pylori α1,3/4-Fucosyltransferase FucTIII toward the Syntheses of Fucosylated Human Milk Glycans and Lewis Antigens. *ACS Catalysis* **9**, 10712-10720 (2019).
138. Lee, W.-H. et al. Whole cell biosynthesis of a functional oligosaccharide, 2′-fucosyllactose, using engineered Escherichia coli. *Microbial Cell Factories* **11**, 1-9 (2012).
139. Sugita, T., Sampei, S. & Koketsu, K. Efficient production of lacto-N-fucopentaose III in engineered Escherichia coli using α1, 3-fucosyltransferase from Parabacteroides goldsteinii. *Journal of Biotechnology* **361**, 110-118 (2023).
140. Totani, K., Shimizu, K., Harada, Y., Murata, T. & Usui, T. Enzymatic synthesis of oligosaccharide containing Lex unit by using partially purified chicken serum. *Bioscience, biotechnology, and biochemistry* **66**, 636-640 (2002).
141. Dumon, C., Samain, E. & Priem, B. Assessment of the two Helicobacter pylori alpha-1,3-fucosyltransferase ortholog genes for the large-scale synthesis of LewisX human milk oligosaccharides by metabolically engineered Escherichia coli. *Biotechnol Prog* **20**, 412-419 (2004).
142. Saumonneau, A. et al. Design of an α-L-transfucosidase for the synthesis of fucosylated HMOs. *Glycobiology* **26**, 261-269 (2016).
143. Zeuner, B., Vuillemin, M., Holck, J., Muschiol, J. & Meyer, A.S. Loop engineering of an α-1, 3/4-l-fucosidase for improved synthesis of human milk oligosaccharides. *Enzyme and microbial technology* **115**, 37-44 (2018).
144. Huang, H.-H. et al. Substrate characterization of Bacteroides fragilis α1, 3/4-fucosyltransferase enabling access to programmable one-pot enzymatic synthesis of KH-1 antigen. *Acs Catalysis* **9**, 11794-11800 (2019).
145. Bai, J. et al. Biochemical characterization of Helicobacter pylori alpha1-3-fucosyltransferase and its application in the synthesis of fucosylated human milk oligosaccharides. *Carbohydr Res* **480**, 1-6 (2019).
146. Whitford, Bioprocess intensification: aspirations and achievements. Biotechnology., Vol. 69 84-87 (Taylor & Francis, 2020).
147. Boodhoo, K.V.K., Flickinger, M.C., Woodley, J.M. & Emanuelsson, E.A.C. Bioprocess intensification: A route to efficient and sustainable biocatalytic transformations for the future. *Chemical Engineering and Processing - Process Intensification* **172**, 108793 (2022).
148. Finnigan, W., Hepworth, L.J., Flitsch, S.L. & Turner, N.J. RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades. *Nat Catal* **4**, 98-104 (2021).
149. Godoy, C.M.d., Machado, D.L. & Costa, A.C.d. Batch and fed-batch enzymatic hydrolysis of pretreated sugarcane bagasse – Assays and modeling. *Fuel* **253**, 392-399 (2019).

150. Ferrer, M., Golyshina, O., Beloqui, A. & Golyshin, P.N. Mining enzymes from extreme environments. *Current opinion in microbiology* **10**, 207-214 (2007).
151. Coudert, E. et al. Annotation of biologically relevant ligands in UniProtKB using ChEBI. *Bioinformatics* **39** (2022).
152. Chang, A. et al. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res* **49**, D498-D508 (2021).
153. Arnold, F.H. & Volkov, A.A. Directed evolution of biocatalysts. *Current opinion in chemical biology* **3**, 54-59 (1999).
154. Tishkov, V. et al. Rational design of practically important enzymes. *Moscow University Chemistry Bulletin* **73**, 1-6 (2018).
155. Song, Z., Zhang, Q., Wu, W., Pu, Z. & Yu, H. Rational design of enzyme activity and enantioselectivity. *Frontiers in Bioengineering and Biotechnology* **11**, 1129149 (2023).
156. Otten, L.G., Hollmann, F. & Arends, I.W. Enzyme engineering for enantioselectivity: from trial-and-error to rational design? *Trends in biotechnology* **28**, 46-54 (2010).
157. Zhang, Y. & Hess, H., Vol. 7 6018-6027 (ACS Publications, 2017).
158. Novy, R., Yaeger, K., Held, D., & Mierendorf, R. Coexpression of multiple target proteins in E. coli. *Innovations* **15**, 2 (2002).
159. Siedentop, R., Claassen, C., Rother, D., Luetz, S. & Rosenthal, K. Getting the Most Out of Enzyme Cascades: Strategies to Optimize Multi-Enzymatic Reactions. *Catalysts* **11** (2021).
160. Ricca, E., Brucher, B. & Schrittwieser, J.H. Multi-Enzymatic Cascade Reactions: Overview and Perspectives. *Advanced Synthesis & Catalysis* **353**, 2239-2262 (2011).
161. Taha, F., Ibrahim, M. & El-Zanaty, E. Optimum conditions for enzymatic degradation of some oilseed proteins. *Grasas y aceites* **53**, 267-272 (2002).
162. Zimmer, J. & Doyle, D.A. Phospholipid requirement and pH optimum for the in vitro enzymatic activity of the E. coli P-type ATPase ZntA. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1758**, 645-652 (2006).
163. Paloyan, A. et al. Molecular Cloning and Biochemical Characterization of a novel thermostable α-amylase of Bacillus subtilis MDC 3500 isolated from acidic soils in Armenia. (2024).
164. Ilzarbe, L., Álvarez, M.J., Viles, E. & Tanco, M. Practical applications of design of experiments in the field of engineering: a bibliographical review. *Quality and Reliability Engineering International* **24**, 417-428 (2008).
165. Gmelch, T.J., Sperl, J.M. & Sieber, V. Optimization of a reduced enzymatic reaction cascade for the production of L-alanine. *Sci Rep* **9**, 11754 (2019).
166. Hold, C., Billerbeck, S. & Panke, S. Forward design of a complex enzyme cascade reaction. *Nat Commun* **7**, 12971 (2016).
167. Niedz, R.P. & Evens, T.J. Design of experiments (DOE)—history, concepts, and relevance to in vitro culture. *In Vitro Cellular & Developmental Biology-Plant* **52**, 547-562 (2016).
168. Fisher, R.A. Statistical methods for research workers. (1934).
169. Hotelling, H. The design of experiments (JSTOR, 1937).
170. Kuehl, R.O. Design of Experiments: Statistical Principles of Research Design and Analysis. (Duxbury/Thomson Learning, 2000).
171. Eichhorn, E. et al. Biocatalytic Process for (−)-Ambrox Production Using Squalene Hopene Cyclase. *Advanced Synthesis & Catalysis* **360**, 2339-2351 (2018).
172. Bornadel, A. et al. Technical Considerations for Scale-Up of Imine-Reductase-Catalyzed Reductive Amination: A Case Study. *Organic Process Research & Development* **23**, 1262-1268 (2019).
173. Bornadel, A., Hatti-Kaul, R., Hollmann, F. & Kara, S. Enhancing the productivity of the bi-enzymatic convergent cascade for ε-caprolactone synthesis through design of experiments and a biphasic system. *Tetrahedron* **72**, 7222-7228 (2016).
174. Martín del Campo, J.S. et al. High-yield production of dihydrogen from xylose by using a synthetic enzyme cascade in a cell-free system. *Angewandte Chemie* **125**, 4685-4688 (2013).
175. Michaelis, L. & Menten, M. (Biochem, 1913).
176. Horn, F. & Jackson, R. General mass action kinetics. *Archive for Rational Mechanics and Analysis* **47**, 81-116 (1972).
177. O'Sullivan, C. & Tompson, F.W. LX. - Invertase: A contribution to the history of an enzyme or unorganised ferment. *Journal of the Chemical Society, Transactions* **57**, 834-931 (1890).
178. Harry, C.J. *Science* **9**, 717-717 (1899).
179. Cornish-Bowden, A. The origins of enzyme kinetics. *FEBS Lett* **587**, 2725-2730 (2013).

180. Nam, K., Shao, Y., Major, D.T. & Wolf-Watz, M. Perspectives on Computational Enzyme Modeling: From Mechanisms to Design and Drug Development. *ACS Omega* **9**, 7393-7412 (2024).
181. Jumper, J. et al. AlphaFold 2. *Fourteenth Critical Assessment of Techniques for Protein Structure Prediction* (2020).
182. Umbarger, H.E. Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. *Science* **123**, 848-848 (1956).
183. Cleland, W.W. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: I. Nomenclature and rate equations. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Enzymological Subjects* **67**, 104-137 (1963).
184. Fersht, A.R. Catalysis, binding and enzyme-substrate complementarity. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **187**, 397-407 (1974).
185. Johnson, K.A. in The Enzymes, Vol. 54. (eds. L.S. Kaguni & F. Tamanoi) 107-134 (Academic Press, 2023).
186. Islam, M.M., Schroeder, W.L. & Saha, R. Kinetic modeling of metabolism: Present and future. *Current Opinion in Systems Biology* **26**, 72-78 (2021).
187. Sukhova, E., Vodeneev, V. & Sukhov, V. Mathematical modeling of photosynthesis and analysis of plant productivity. *Biochemistry (Moscow), Supplement Series A: Membrane and Cell Biology* **15**, 52-72 (2021).
188. Costa, R.S., Hartmann, A. & Vinga, S. Kinetic modeling of cell metabolism for microbial production. *Journal of Biotechnology* **219**, 126-141 (2016).
189. Sudar, M., Vasić-Rački, Đ., Müller, M., Walter, A. & Blažević, Z.F. Mathematical model of the MenD-catalyzed 1, 4-addition (Stetter reaction) of α-ketoglutaric acid to acrylonitrile. *Journal of biotechnology* **268**, 71-80 (2018).
190. Sudar, M. & Blažević, Z.F. Enzyme cascade kinetic modelling. *Enzyme Cascade Design and Modelling*, 91-108 (2021).
191. Engel, J., Bornscheuer, U.T. & Kara, S. Kinetics modeling of a convergent cascade catalyzed by monooxygenase–alcohol dehydrogenase coupled enzymes. *Organic Process Research & Development* **25**, 411-420 (2020).
192. Johannsen, J. et al. Multi-enzyme cascade reaction in a miniplant two-phase-system: Model validation and mathematical optimization. *AIChE journal* **67**, e17158 (2021).
193. Studier, F.W. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J Mol Biol* **219**, 37-44 (1991).
194. Park, Y.S., Sweitzer, T.D., Dixon, J.E. & Kent, C. Expression, purification, and characterization of CTP:glycerol-3-phosphate cytidylyltransferase from Bacillus subtilis. *Journal of Biological Chemistry* **268**, 16648-16654 (1993).
195. Kaneko∗, T. & Yamamoto, H. Cytidine Kinase: A Novel Pyrimidine Ribonucleoside Phosphorylating Enzyme inEscherzchza Coli. *Nucleosides and Nucleotides* **16**, 2111-2122 (2006).
196. Zhou, L. & Thornburg, R. Site-specific mutations of conserved residues in the phosphate-binding loop of the Arabidopsis UMP/CMP kinase alter ATP and UMP binding. *Arch Biochem Biophys* **358**, 297-302 (1998).
197. Achbergerova, L. & Nahalka, J. Degradation of polyphosphates by polyphosphate kinases from Ruegeria pomeroyi. *Biotechnol Lett* **36**, 2029-2035 (2014).
198. Koga, Y. et al. Thermostable glycerol kinase from a hyperthermophilic archaeon: gene cloning and characterization of the recombinant enzyme. *Protein Eng* **11**, 1219-1227 (1998).
199. Halsor, M.J.H., Rothweiler, U., Altermark, B. & Raeder, I.L.U. The crystal structure of the N-acetylglucosamine 2-epimerase from Nostoc sp. KVJ10 reveals the true dimer. *Acta Crystallogr D Struct Biol* **75**, 90-100 (2019).
200. Huynh, N. et al. Structural basis for substrate specificity and mechanism of N-acetyl-D-neuraminic acid lyase from Pasteurella multocida. *Biochemistry* **52**, 8570-8579 (2013).
201. Edwards, U. & Frosch, M. Sequence and functional analysis of the cloned Neisseria meningitidis CMP-NeuNAc synthetase. *FEMS Microbiol Lett* **75**, 161-166 (1992).
202. Liu, Y. et al. Cryo-EM structure of L-fucokinase/GDP-fucose pyrophosphorylase (FKP) in Bacteroides fragilis. *Protein Cell* **10**, 365-369 (2019).
203. Gentry, D., Bengra, C., Ikehara, K. & Cashel, M. Guanylate kinase of Escherichia coli K-12. *Journal of Biological Chemistry* **268**, 14316-14321 (1993).
204. Biolabs, N.E. High Efficiency Transformation Protocol (C2987I) V.1. *protocols.io* (2014).

205. Chung, C., Niemela, S.L. & Miller, R.H. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences* **86**, 2172-2175 (1989).
206. Hoops, S. et al. COPASI--a COmplex PAthway SImulator. *Bioinformatics* **22**, 3067-3074 (2006).
207. Huber, N., Alcalá-Orozco, E.A., Rexer, T., Reichl, U. & Klamt, S. Model-based optimization of cell-free enzyme cascades exemplified for the production of GDP-fucose. *Metabolic Engineering* **81**, 10-25 (2024).
208. Nelder, J.A. & Mead, R. A simplex method for function minimization. *The computer journal* **7**, 308-313 (1965).
209. Park, Y.S., Sweitzer, T.D., Dixon, J.E. & Kent, C. Expression, purification, and characterization of CTP:glycerol-3-phosphate cytidylyltransferase from Bacillus subtilis. *J Biol Chem* **268**, 16648-16654 (1993).
210. Sanker, S., Campbell, H.A. & Kent, C. Negative cooperativity of substrate binding but not enzyme activity in wild-type and mutant forms of CTP:glycerol-3-phosphate cytidylyltransferase. *J Biol Chem* **276**, 37922-37928 (2001).
211. Katsumi, R. et al. Crystallization and preliminary X-ray diffraction study of glycerol kinase from the hyperthermophilic archaeon Thermococcus kodakaraensis. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**, 126-129 (2007).
212. Koga, Y. et al. Crystal structure of highly thermostable glycerol kinase from a hyperthermophilic archaeon in a dimeric form. *FEBS J* **275**, 2632-2643 (2008).
213. Valentin-Hansen, P. Uridine-cytidine kinase from Escherichia coli. *Methods Enzymol* **51**, 308-314 (1978).
214. Akiyama, M., Crooke, E. & Kornberg, A. The polyphosphate kinase gene of Escherichia coli. Isolation and sequence of the ppk gene and membrane location of the protein. *Journal of Biological Chemistry* **267**, 22556-22561 (1992).
215. Applebee, M.K., Joyce, A.R., Conrad, T.M., Pettigrew, D.W. & Palsson, B.O. Functional and metabolic effects of adaptive glycerol kinase (GLPK) mutants in Escherichia coli. *J Biol Chem* **286**, 23150-23159 (2011).
216. Okesli-Armlovich, A. et al. Discovery of small molecule inhibitors of human uridine-cytidine kinase 2 by high-throughput screening. *Bioorg Med Chem Lett* **29**, 2559-2564 (2019).
217. Iyer, P.V. & Ananthanarayan, L. Enzyme stability and stabilization—Aqueous and non-aqueous environment. *Process Biochemistry* **43**, 1019-1032 (2008).
218. Neville, N., Roberge, N. & Jia, Z. Polyphosphate Kinase 2 (PPK2) Enzymes: Structure, Function, and Roles in Bacterial Physiology and Virulence. *Int J Mol Sci* **23** (2022).
219. Kuroda, A. & Kornberg, A. Polyphosphate kinase as a nucleoside diphosphate kinase in Escherichia coli and Pseudomonas aeruginosa. *Proc Natl Acad Sci U S A* **94**, 439-442 (1997).
220. Klermund, L., Poschenrieder, S.T. & Castiglione, K. Biocatalysis in Polymersomes: Improving Multienzyme Cascades with Incompatible Reaction Steps by Compartmentalization. *ACS Catalysis* **7**, 3900-3904 (2017).
221. Li, Y. et al. Pasteurella multocida sialic acid aldolase: a promising biocatalyst. *Appl Microbiol Biotechnol* **79**, 963-970 (2008).
222. Ganguli, S. et al. Molecular cloning and analysis of genes for sialic acid synthesis in Neisseria meningitidis group B and purification of the meningococcal CMP-NeuNAc synthetase enzyme. *J Bacteriol* **176**, 4583-4589 (1994).
223. Li, Y., Yu, H., Cao, H., Muthana, S. & Chen, X. Pasteurella multocida CMP-sialic acid synthetase and mutants of Neisseria meningitidis CMP-sialic acid synthetase with improved substrate promiscuity. *Appl Microbiol Biotechnol* **93**, 2411-2423 (2012).
224. Matthews, M.M. et al. Catalytic Cycle of Neisseria meningitidis CMP-Sialic Acid Synthetase Illustrated by High-Resolution Protein Crystallography. *Biochemistry* **59**, 3157-3168 (2020).
225. Arcus, V.L. & Mulholland, A.J. Temperature, dynamics, and enzyme-catalyzed reaction rates. *Annual review of biophysics* **49**, 163-180 (2020).
226. Klermund, L., Groher, A. & Castiglione, K. New N-acyl-D-glucosamine 2-epimerases from cyanobacteria with high activity in the absence of ATP and low inhibition by pyruvate. *J Biotechnol* **168**, 256-263 (2013).
227. Schelch, S. et al. Engineering analysis of multienzyme cascade reactions for 3'-sialyllactose synthesis. *Biotechnol Bioeng* **118**, 4290-4304 (2021).
228. Wang, S.L. et al. Molecular Characterization of a Novel N-Acetylneuraminate Lyase from a Deep-Sea Symbiotic Mycoplasma. *Mar Drugs* **16** (2018).

229. Ferreira, R.D.G., Azzoni, A.R. & Freitas, S. Techno-economic analysis of the industrial production of a low-cost enzyme using E. coli: the case of recombinant beta-glucosidase. *Biotechnol Biofuels* **11**, 81 (2018).
230. Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K.S., Chowdhury, A. & Hossain, M.N. Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant E. coli. *ISRN Biotechnol* **2013**, 590587 (2013).
231. Xiong, Y.L. in Quality in frozen food 111-140 (Springer, 1997).
232. Takeshita, Y. et al. The effects of pressure on pH of Tris buffer in synthetic seawater. *Marine Chemistry* **188**, 1-5 (2017).
233. Onaka, T., Miyajima, T. & Ohashi, S. On the binding of magnesium to long-chain polyphosphate ions. *Journal of Inorganic and Nuclear Chemistry* **43**, 3323-3327 (1981).
234. Liu, J.L.C. et al. Overproduction of CMP-sialic acid synthetase for organic synthesis. *Journal of the American Chemical Society* **114**, 3901-3910 (2002).
235. Dharmadi, Y., Murarka, A. & Gonzalez, R. Anaerobic fermentation of glycerol by Escherichia coli: a new platform for metabolic engineering. *Biotechnol Bioeng* **94**, 821-829 (2006).
236. Friehs, K. Plasmid copy number and plasmid stability. *Adv Biochem Eng Biotechnol* **86**, 47-82 (2004).
237. Vieira, E., Brandao, T. & Ferreira, I.M. Evaluation of Brewer's spent yeast to produce flavor enhancer nucleotides: influence of serial repitching. *J Agric Food Chem* **61**, 8724-8729 (2013).
238. Rocha, R.A.R. et al. Temporal profile of flavor enhancers MAG, MSG, GMP, and IMP, and their ability to enhance salty taste, in different reductions of sodium chloride. *J Food Sci* **85**, 1565-1575 (2020).
239. Feichtinger, R.G. et al. A spoonful of L-fucose-an efficient therapy for GFUS-CDG, a new glycosylation disorder. *EMBO Mol Med* **13**, e14332 (2021).
240. Liu, J.J. et al. L-Fucose production by engineered Escherichia coli. *Biotechnol Bioeng* **116**, 904-911 (2019).
241. Meng, J. et al. Microbial Synthesis of l-Fucose with High Productivity by a Metabolically Engineered Escherichia coli. *J Agric Food Chem* **71**, 2464-2471 (2023).
242. Meng, J. et al. High-Level Production of l-Fucose by Plasmid-Free or Antibiotic-Independent Metabolically Engineered Escherichia coli Strains. *ACS Sustainable Chemistry & Engineering* **12**, 6619-6628 (2024).
243. Ren, Y. et al. Biochemical characterization of GDP-L-fucose de novo synthesis pathway in fungus Mortierella alpina. *Biochem Biophys Res Commun* **391**, 1663-1669 (2010).
244. Kotake, T. et al. A bifunctional enzyme with L-fucokinase and GDP-L-fucose pyrophosphorylase activities salvages free L-fucose in Arabidopsis. *J Biol Chem* **283**, 8125-8135 (2008).
245. Coyne, M.J., Reinap, B., Lee, M.M. & Comstock, L.E. Human symbionts use a host-like pathway for surface fucosylation. *Science* **307**, 1778-1781 (2005).
246. Tao, Y.H. et al. Hydration water drives the self-assembly of guanosine monophosphate. *Biophys J* **123**, 931-939 (2024).
247. Bhattacharyya, T., Saha, P. & Dash, J. Guanosine-Derived Supramolecular Hydrogels: Recent Developments and Future Opportunities. *ACS Omega* **3**, 2230-2241 (2018).
248. Quirk, S. Crystallization and preliminary X-ray characterization of the Vitis vinifera fucokinase:GDP-fucose pyrophosphorylase. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 1109-1112 (2012).
249. Cornish-Bowden, A. One hundred years of Michaelis–Menten kinetics. *Perspectives in Science* **4**, 3-9 (2015).
250. Papadopoulos, J.S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073-1079 (2007).
251. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**, W475-478 (2011).
252. Gupta, S. et al. Purification and characterization of guanylate kinase, a nucleoside monophosphate kinase of Brugia malayi. *Parasitology* **141**, 1341-1352 (2014).
253. Ohashi, H., Wahl, C., Ohashi, T., Elling, L. & Fujiyama, K. Effective Synthesis of Guanosine 5′-Diphospho-β-l-galactose Using Bacterial l-Fucokinase/Guanosine 5′-Diphosphate-l-fucose Pyrophosphorylase. *Advanced Synthesis & Catalysis* **359**, 4227-4234 (2017).
254. Beguerisse-Diaz, M., Bosque, G., Oyarzun, D., Pico, J. & Barahona, M. Flux-dependent graphs for metabolic networks. *NPJ Syst Biol Appl* **4**, 32 (2018).

255. Bolanos-Garcia, V.M. & Davies, O.R. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta* **1760**, 1304-1313 (2006).
256. Shiraishi, F. & Suzuki, Y. Method for determination of the main bottleneck enzyme in a metabolic reaction network by dynamic sensitivity analysis. *Industrial & Engineering Chemistry Research* **48**, 415-423 (2009).
257. Patrick, S. in Molecular Medical Microbiology 1921-1948 (Elsevier, 2002).
258. Ge, Z., Chan, N.W., Palcic, M.M. & Taylor, D.E. Cloning and heterologous expression of an alpha1,3-fucosyltransferase gene from the gastric pathogen Helicobacter pylori. *J Biol Chem* **272**, 21357-21363 (1997).
259. Ma, B., Lau, L.H., Palcic, M.M., Hazes, B. & Taylor, D.E. A single aromatic amino acid at the carboxyl terminus of Helicobacter pylori alpha1,3/4 fucosyltransferase determines substrate specificity. *J Biol Chem* **280**, 36848-36856 (2005).
260. Sun, H.-Y. et al. Structure and Mechanism of Helicobacter pylori Fucosyltransferase. *Journal of Biological Chemistry* **282**, 9973-9982 (2007).
261. Lin, S.-W., Yuan, T.-M., Li, J.-R. & Lin, C.-H. Carboxyl terminus of Helicobacter pylori α 1, 3-fucosyltransferase determines the structure and stability. *Biochemistry* **45**, 8108-8116 (2006).
262. Prudden, A.R., Chinoy, Z.S., Wolfert, M.A. & Boons, G.-J. A multifunctional anomeric linker for the chemoenzymatic synthesis of complex oligosaccharides. *Chemical Communications* **50**, 7132-7135 (2014).
263. Chen, C. et al. Sequential one-pot multienzyme (OPME) synthesis of lacto-N-neotetraose and its sialyl and fucosyl derivatives. *Chemical Communications* **51**, 7689-7692 (2015).
264. Wang, S. et al. Facile chemoenzymatic synthesis of O-mannosyl glycans. *Angewandte Chemie International Edition* **57**, 9268-9273 (2018).

List of Figures[i]

[i] Figures 2, 3, 5, 7, 19, 31, 34 and 54 were created using Biorender.com

List of Tables

Appendix

Table 13 - Recombinant *E.coli* strains used in this work. All bacterial expression strains were transformed with vectors produced by Biocat GmbH. Competent cells from New England Biolabs were used for all the different expression strains.

Gene	Vector	Strain	Restriction sites	Cascade
tagD	pET28a (+)	NiCo21 (DE3)	NdeI-XhoI	CDP-glycerol
UDK	pET28a (+)	NiCo21 (DE3)	NdeI-XhoI	CDP-glycerol
URA6	pET28a (+)	Bl21(DE3)	NdeI-XhoI	CDP-glycerol / CMP-Neu5Ac
PPK3	pET28a (+)	Bl21(DE3)	NdeI-XhoI	CDP-glycerol / CMP-Neu5Ac / GDP-fucose
glpK	pET28a(+)	NiCo21 (DE3)	NdeI-XhoI	CDP-glycerol
PPA	pET28a (+)	Bl21(DE3)	NdeI-XhoI	CDP-glycerol / CMP-Neu5Ac / GDP-fucose
AGE	-	-	-	CMP-Neu5Ac
NANA	-	-	-	CMP-Neu5Ac
CSS	-	-	-	CMP-Neu5Ac
FKP	pET100/D-TOPO	Bl21(DE3)	NdeI-XhoI	GDP-fucose
GMK	pET28a (+)	Bl21(DE3)	NdeI-XhoI	GDP-fucose
A3/4 FT	pET28a (+)	Bl21(DE3)	NdeI-XhoI	LNFP III
UDK+AGE	-	-	-	CMP-Neu5Ac
UDK+AGE	-	-	-	CMP-Neu5Ac
NANA+CSS	-	-	-	CMP-Neu5Ac
URA6+PPK3	-	-	-	CMP-Neu5Ac
URA6+PPA	-	-	-	CMP-Neu5Ac
CSS+NANA	-	-	-	CMP-Neu5Ac
AGE+PPK3	-	-	-	CMP-Neu5Ac
PPK3+CSS	-	-	-	CMP-Neu5Ac
NANA+AGE	-	-	-	CMP-Neu5Ac
URA6+PPA	-	-	-	CMP-Neu5Ac

Gene	Vector	Strain	Restriction sites	Cascade
FKP+PPA	pDuet-RSF	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
PPK3+GMK	pDuet-CDF	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
FKP+PPA	pDuet-CYC	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
PPK3+GMK	pDuet-CDF	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
FKP+PPA	pDuet-pET	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
PPK3+GMK	pDuet-CDF	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
FKP+PPK3	pDuet-CYC	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
GMK+PPA	pDuet-CDF	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
FKP+PKK3	pDuet-pET	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose
GMK+PPA	pDuet-CDF	Bl21(DE3)	BamHI-NotI / NdeI-PacI	GDP-fucose

*Vectors for the CMP-Neu5Ac enzymatic cascade were removed at request of eversyn GmbH.

Table 14 - Worksheet of CDP-glycerol DoE. Central composite orthogonal design. (*) excluded outliers (>3 σ); (°) Centre points.

Experiment Name	pH value	T [°C]	$MgCl_2$ [mM]	CDP-glycerol Peak area [mAU*min]
N1	7	21	10	0.929
N2	8.5	21	10	27.862
N3	7	45	10	0.275
N4	8.5	45	10	20.44
N5	7	21	100	2.109
N6	8.5	21	100	48.924
N7	7	45	100	7.131
N8	8.5	45	100	14.598
N9	6.7	33	55	0.548
N10	8.7	33	55	40.595
N11	7.75	16	55	9.303
N12	7.75	49	55	2.957
N13	7.75	33	0	0
N14	7.75	33	115	19.897
N15°	7.75	33	55	14.283
N16°	7.75	33	55	25.857
N17°	7.75	33	55	25.934
N18	7	21	10	0.446
N19	8.5	21	10	17.862

Experiment Name	pH value	T [°C]	$MgCl_2$ [mM]	CDP-glycerol Peak area [mAU*min]
N20	7	45	10	1.473
N21	8.5	45	10	12.538
N22	7	21	100	0.998
N23	8.5	21	100	7.227
N24	7	45	100	2.586
N25	8.5	45	100	8.028
N26	6.7	33	55	2.123
N27	8.7	33	55	17.192
N28	7.75	16	55	5.965
N29	7.75	49	55	0.455
N30	7.75	33	0	0
N31	7.75	33	115	20.314
N32°	7.75	33	55	14.384
N33°	7.75	33	55	23.439
N34°*	7.75	33	55	53.541
N35	7	21	10	0
N36	8.5	21	10	22.29
N37	7	45	10	1.514
N38	8.5	45	10	16.946
N39	7	21	100	9.164
N40	8.5	21	100	9.317
N41	7	45	100	2.264
N42	8.5	45	100	16.029
N43	6.7	33	55	1.224
N44	8.76	33	55	15.825
N45	7.75	16	55	3.026
N46	7.75	49	55	1.622
N47	7.75	33	0	0.356
N48	7.75	33	115.9	16.139
N49°	7.75	33	55	14.246
N50°	7.75	33	55	44.95
N51°	7.75	33	55	13.072

Table 15 - Worksheet of CMP-Neu5Ac DoE. D-optimal design. (*) excluded outliers (>2 σ). (°) Centre points.

Experiment No.	T [°C]	pH	$MgCl_2$ [mM]	CMP-Neu5Ac [mM]
1	45	8	70	4.8
2	25	8	40	6.5
3	25	9	10	0

Experiment No.	T [°C]	pH	$MgCl_2$ [mM]	CMP-Neu5Ac [mM]
4	40	7.5	100	6.1
5	25	7.5	10	0.5
6	25	9	10	0.9
7	25	7.5	70	9.9
8	40	8	10	1.1
9*	25	7.5	100	11.9
10	45	7.5	100	2.8
11	25	8.5	100	8.5
12	45	7.5	40	3.9
13	40	9	40	7.3
14	25	7.5	10	0.4
15	40	9	40	5.8
16	40	8.5	70	7
17	25	7.5	70	9.8
18	25	9	100	9.8
19	45	8.5	10	0
20°*	35	8.5	70	11
21	25	9	10	0.5
22	25	9	10	0.7
23	45	8.5	10	0
24°*	35	8.5	70	10.9
25	25	7.5	10	0.5
26	40	8.5	70	7.1
27	30	7.5	10	0
28*	30	9	70	6.6
29	45	7.5	100	2.8
30	45	8.5	10	0
31	45	7.5	10	0
32°*	35	8.5	70	10.8
33	30	8	100	7.5
34	30	8	100	7.4
35*	25	7.5	100	10.6
36°*	35	8.5	70	12.9
37	25	9	10	0.6
38*	30	9	70	6.3
39	25	8	40	6.4
40	40	8.5	70	6.9
41	40	8	10	1.1
42	30	8	100	7.7
43	40	8	10	0.9
44	45	9	10	0

Experiment No.	T [°C]	pH	$MgCl_2$ [mM]	CMP-Neu5Ac [mM]
45	25	8.5	100	9.7
46	30	7.5	10	0
47	45	8	70	4.8
48	30	8.5	40	6.5
49°*	35	8.5	70	12
50	40	7.5	100	6.2
51°*	35	8.5	70	12.3
52*	25	7.5	100	11.9
53	30	8.5	40	6.5
54	45	9	100	3.2
55	45	7.5	40	3.8
56	30	7.5	10	0
57	25	7.5	70	10
58	25	9	10	0
59	25	8	40	6.6
60	45	7.5	40	3.7
61°*	35	8.5	70	11.5
62	25	9	100	8.5
63	45	9	10	0
64	45	9	100	2.6
65°*	35	8.5	70	12.1
66°*	35	8.5	70	11.1
67*	30	9	70	6.5
68	45	8	70	4.9
69	30	8.5	40	6.7
70*	40	9	40	7.7
71	45	9	100	3.1
72*	25	9	100	5.6
73	45	7.5	100	2.8
74	40	7.5	100	4.4
75	45	9	10	0
76	25	8.5	100	9.4
77	45	7.5	10	0
78	45	7.5	10	0

Table 16 - Worksheet of GDP-fucose DoE. Fractional Factorial resolution IV design; (*) Outliers (>2 σ); (°) Centre points.

Exp	pH	polyP [mM]	$MgCl_2$ [mM]	GMP/ fucose [mM]	T° C	FKP [g/L]	ATP [mM]	Duets [g/L]	$MnCl_2$ [mM]	GDP-fucose [Yp/s]
1	7	5	5	5	23	2	10	1.5	10	42.51
2	9	5	5	5	23	2	2	0.3	0	6.03
3	7	50	5	5	23	0.5	10	0.3	0	0
4	9	50	5	5	23	0.5	2	1.5	10	8.93
5	7	5	50	5	23	0.5	2	1.5	0	0
6	9	5	50	5	23	0.5	10	0.3	10	41.86
7	7	50	50	5	23	2	2	0.3	10	1.24
8	9	50	50	5	23	2	10	1.5	0	60.09
9	7	5	5	30	23	0.5	2	0.3	10	6.31
10	9	5	5	30	23	0.5	10	1.5	0	71.67
11	7	50	5	30	23	2	2	1.5	0	0
12*	9	50	5	30	23	2	10	0.3	10	0.94
13	7	5	50	30	23	2	10	0.3	0	8.57
14	9	5	50	30	23	2	2	1.5	10	0.83
15*	7	50	50	30	23	0.5	10	1.5	10	11.7
16	9	50	50	30	23	0.5	2	0.3	0	0
17	7	5	5	5	50	0.5	2	0.3	0	0
18*	9	5	5	5	50	0.5	10	1.5	10	56.49
19	7	50	5	5	50	2	2	1.5	10	6.438
20	9	50	5	5	50	2	10	0.3	0	60.44
21*	7	5	50	5	50	2	10	0.3	10	60.83
22*	9	5	50	5	50	2	2	1.5	0	32.68
23	7	50	50	5	50	0.5	10	1.5	0	0
24	9	50	50	5	50	0.5	2	0.3	10	28.26
25	7	5	5	30	50	2	10	1.5	0	20.12
26	9	5	5	30	50	2	2	0.3	10	11.86
27	7	50	5	30	50	0.5	10	0.3	10	0
28	9	50	5	30	50	0.5	2	1.5	0	10.55
29	7	5	50	30	50	0.5	2	1.5	10	15.81
30	9	5	50	30	50	0.5	10	0.3	0	21.79
31	7	50	50	30	50	2	2	0.3	0	0
32	9	50	50	30	50	2	10	1.5	10	45.48
33°	8	27.5	25.5	17.5	36	1.25	6	0.9	5	27.20
34°	8	27.5	25.5	17.5	36	1.25	6	0.9	5	27.20
35°	8	27.5	25.5	17.5	36	1.25	6	0.9	5	27.20
36*	7	5	5	5	23	0.5	2	0.3	10	25.20
37*	9	5	5	5	50	0.5	10	1.5	0	0

Exp	pH	polyP [mM]	$MgCl_2$ [mM]	GMP/ fucose [mM]	T° C	FKP [g/L]	ATP [mM]	Duets [g/L]	$MnCl_2$ [mM]	GDP-fucose [Yp/s]
38	7	50	5	5	50	2	2	1.5	0	0
39*	9	50	5	5	23	2	10	1.5	10	16.24
40*	7	5	50	5	50	2	10	0.3	0	0.73
41	9	5	50	5	23	2	2	1.5	10	29.04
42	7	50	50	5	23	0.5	10	1.5	10	0.88
43	9	50	50	5	50	0.5	2	0.3	0	1.88
44	7	5	5	30	23	2	10	1.5	0	13.43
45	9	5	5	30	50	2	2	0.3	10	11.86
46	7	50	5	30	50	0.5	10	0.3	10	0
47	9	50	5	30	23	0.5	2	1.5	0	0
48	7	5	50	30	50	0.5	2	1.5	10	15.81
49	9	5	50	30	23	0.5	10	0.3	0	22.95
50	7	50	50	30	23	2	2	0.3	0	0.33
51	9	50	50	30	50	2	10	1.5	10	45.48
52°	8	27.5	25.5	17.5	36	1.25	6	0.9	5	27.20
53°	8	27.5	25.5	17.5	36	1.25	6	0.9	5	35.48
54°	8	27.5	25.5	17.5	36	1.25	6	0.9	5	32.3°

Sequences of FKP candidates:

>MBN2660834.1 MAG: bifunctional fucokinase/L-fucose-1-P-guanylyltransferase [Tannerellaceae bacterium] ENV 24-FEB-2021
MRKLLSLPPNLVDCFHELENASREEWFCTSDPLGSKLGSGGGTTWLLEACKENEDNTTESTEWLAREKRIL
LHAGGQSRRLPGYAPSGKILTPIPVFRWERGQRLGQKLLDLQLPLYERIMNLAPDSMHTLIASGDVYIRSEKA
LQPIPEADVVCYGLWVEPSLATHHGVFVSDRKHPDALDFMLQKPSLEELGSLAKTHLFLMDIGIWLLSDRAV
ELLMKRSEGDGANGLKYYDLYSDFGRALGTNPKVKDPAINQLSVAILPLPGGEFYHYGTSRELISSTLSVQNL
VRDQREIMHRKVKPNPAMFVQNAIMKGPLTAENENVWIENSYIGNRWKLSQGNIVTGVPANDWELSVPAGI
CVDVVPVGNTSWVARPYGLNDAFKGELSNKHTLWMGKPVSEWFIARGIESEKELKGNTDDLQSAAIFPVCE
SIEDLGVLLRWMINEPGLAEGKALWLTAEKLSADEISATANLSRLYAQRTLFRQDNWTQLADNYERSVFYQL
NLSDAANEFARGGLAVPPALPEDAPLLSRMHNHMFRARTLRLKGDSNYTDDEKASFALLREGLAGSVSDNK
QEPHLSVFSDQIVWGRSPVRIDLAGGWTDTPPFCLYSGGNVVNMAIELNGQPPLQVYIKPSKTFQIVLRSIDL
GAMEVINTYEELKEFTKVGSPFSIPKAALSLAGFAPQFGTGCYPTLEERLKEFGSGIEVTLLAAIPAGSGLGTS
SILAATVLGAVNDFCGLAWDKNEICRRTLVLEQLLTTGGGWQDQYGGVLQGVKLLQTGRGFDQLPLVRWLP
DFIFNRAEYKACHLLYYTGITRTAKGILAEIVSGMFLNSTEHLELLQEMKLHALDTYEAILRGNFAETAALVGKS
WRLNKALDSGTNPPSIEAIIAQIKDYTLGYKLPGAGGGGYLYMIAKDPLAAIRIREILTQNPPNANARFVDMNL
SDKGLQISRS

>WP_117853596.1 bifunctional fucokinase/fucose-1-phosphate guanylyltransferase [Phocaeicola vulgatus] BCT 01-MAR-2022
MKKLLSLPPNLVECFHDIEKADQTEWFCTSDPIGSKLGSGGGTAWLLEACCQKVAPDSDFLTWLGKEKRILL
HAGGQSRRLPGYAPSGKILTPIPVFRWARGQRLSQNLLSLQLPLYEQIMEKAPSSLHTLIASGDVYIRAGQPL
QTIPDADVVCYGLWVDPNLAKNHGVFVSSRATPDKLDFMLQKPSVEELGKLMQTHLFLMDIGIWLLSDRAVS
LLVKRSYKEGKLSYYDMYSDFGLTLGEHPRMMDDELNKLSVAILPLPGGEFYHYGTSRELISSTLAVQNLVN
DQREIMHKKVKPHPAMFVQNAEVGYQLTSQNSEIWIENSYVGAGWNIHHQTIITGVPANNWNLEVPSGVCID
VVPFGESGYVARPYGFNDTFKGALAKEETYYQGMSVGEWCAVRGISVEEIENGHDLQAARLFPVCSSVEEL
GAVMRWMVSEPALQQGKEIWQRCRKLSADDISAYSNLYRLAEQREAFRIKNWLALAHNYERSVFYQLNLEN
AAGEFARYDLSLPEPLSESAPLMTRISDNMFRARVQQLKGLAYREYENEAFRLMRDGLTASALAKRQQPHL
SVYSDQIVWGRSPVRIDLAGGWTDTPPYCLNEGGNVVNIAIELNGQPPLQVYVKPCREYKIILRSIDLGAMEV
VTTYGEVRDFMQVGSPFSIPKAALVLAGFQPGFSTESYVSLEEQLKAFGSGMEITLLSAIPAGSGLGTSSILAS
TVLGAISDFCGLNWDKNEICNRTLILEQLLTTGGGWQDQYGGVLRGVKLLQTHAGMDQSPLVRWLPDYLFT
GGEYQKCHLLYYTGITRTAKGILAEIVRSMFLNSTEHLSILGGMKGHALDLYEAIQRGNFDEMGRLVGKSWKL

NQALDPGTNPEAVETIIRRIDDYCLGYKLPGAGGGGYLYMVAKDPEAAIRIRSILTQNPPNSCARFVDMALSD
KGLQISRS

>MBD9167429.1 MAG: bifunctional fucokinase/L-fucose-1-P-guanylyltransferase [Parabacteroides johnsonii]
ENV 05-OCT-2020
MKKLLSLPPNLVGSFHEIANADPADWFCTSDPVGARLGSGGGTTWLLEACRRDDDAAGTLTTGEWLACEK
RILLHAGGQSRRLPGYAPSGKILTPIPVFRWARGQKLSQNLLSLQLPLYEEIMRKAPDSLHTLIASGDVYLRNS
EPLQEIPEADVVCYGLWVDPALATRHGVFVSDRKFPDQLDFMLQKPTLDELGRLAGTHLFLMDIGVWLLSDR
AVELLMKHSYDPDGKQMKEYDLYSEFGLALGAHPRIADEELNALTVAILPLPGGEFYHYGTSRELISSTLSVQ
NLVRDQRAIMQRKVKPHPAMFIQNAEVCCPLTSGNSELWIENSFVGKRWILADRHVITGVPVNDWELHIPSG
VCIDVVPAGDEGWVARPYGFNDPFKGNTMDESTLFMGRPVVEWAAERGVCLPECADIQNAALFPVCRSMD
ELGIVLRWMVSEPHLDEGRRVWEAAGKMSANRLSDCADLRRLFAQREAFRKKNWPMLAANHDKSIFYQLD
LADAASEFVTGGIALPEALPAEAPLMKRVHDHMFRARVLQLSGDDRGMVEQQQAFSLLREGLISTIADEKQA
PRLNVYRDQIVWGRSPVRIDLAGGWTDTPPYCMYAGGNVVNVAIELNGQPPLQVYVKPAREYRVILRSIDIG
AMESISTWDELRNFNKVGSPFSIPKAALALAGFAPEFSAGRYASLEEQLKAFGCGLEVTLLAAIPAGSGLGTS
SILAATVLGALSDFCGLAWDKNEIGNRTLILEQLLTTGGGWQDQYGGVLHGLKLLQTGEGFHQNPSVRWLP
EYLFTEPEYRACHLLYYTGITRTAKDILAEIVRGMFLNSSTHLRLLSEMKAHALDMYEAILRGDFASYGRLVGK
SWEQNKALDAGTNPPAVERLISRIKDYTLGYKLPGAGGGGYLYIVAKDPEASLQIRRLLTADPQNGNARFVE
MSLSDKGLQVSRS

>WP_278455875.1 bifunctional fucokinase/fucose-1-phosphate guanylyltransferase [Paraprevotella clara]
BCT 13-APR-2023
MKKLLSLPPNLVACFHEVERADRKEWFCTSDPIGHKLGSGGGTAWLLQACKASETDEGETFAAWLSREKRI
LLHAGGQSRRLPAYAPSGKILTPVPVFRWKRGQRLDQNLLSLQLPLYEQIMERAPQSLHTLIASGDVYIRTDR
PLQDIPEADVVCYGMWVDPSLAQNHGVFVSDRRSPERLAFMLQKPSVAELGALMSKHLFLMDIGIWLLSDR
AVELMVKRSYRDGRLSYYDMYSEFGLALGDRPTLDDPELNSLTVAILPLEGGSFYHYGTSREMISSTLAVQNI
VTDQREIMHRKAKPHPAMFVQNAEVEVKLEAANSELWIENSFVGRDWTLACRNIITGVPENRWPLKLADGL
CIDVVPVGEEAFVARPYGFNDAFKGNLSDGAVLYQGMPVTEWLAGRGLKPEDIEENHDLQAARLFPLCDNV
EDLGRAMRWMTTEPELEDGRKVWFSARKMSADELSAYANLHRLTCQREAFRTKNLPLLAAHYERSVFYQL
NLDDVAHEYAEEDLPLPDALPEMADGLTRISDAMFRARVADLKGEDGRKYEEQAFGLMRKMLTGTACAAR
QSPRLSVYADQIVWGRSPVRIDLAGGWTDTPPYSLMEGGNVVNLSIELNGQPPLQVYVKPCRERHIVMRSID
LGAMEVVRTYDELAAFNKVGSPFSIPKAALVLAGFHPDFSAEVHASLEAQLEAFGAGIEITLLSAIPAGSGLGT
SSILASTVLGAVNDFCGLGWDRYETGNRTLVLEQLLTTGGGWQDQYGGILQGVKLLQTQPGACQQPLVRW
LPDYVFTAPEYRKCHLLYYTGITRTAKNILAEIVKGMFLNETGRLELLGRMKTHALDMYDAIQRNCFEETGRL
VRRSWMQNCRLDAGTNPAAVRAIIEKIDDLCLGYKLPGAGGGGFLYMMAKDEEAAARIRKILVQEAVNDRAR
FVEMSLSEKGLEVSRS

>XP_009314224.1 putative fucose kinase [Trypanosoma grayi]
MRVLLSVPLSASTAQAALQKGLLSQHHNVFFTSDPQGSHLGSGAGTAWLLEACYRADQNDEENHNNNNTG
SHETASFLRWLAREPRIIVHSGGQSRRLPAYGPCGKVLAPIPVFRWSRGQTCEQTLLDLQMPLYKEVMRRA
PPHLRTLVSCGDVLILRGQPLPPVPHDADVVCYGIQKETDMLRRHGVFFMNRETPEELEVMLQKPLLQEINE
FSAKRQCVMDIGLWLLSDRAVEVLRQRSMERSGVDQPQSACGWREYDLYSDFGAALGHRPLRDDPLVRQ
LRVKVVTLHEARFFHYGTTTELIDSTFVIQNTVRDDDSLLQAMVSCHPSVFCQNACIRAPIIEDQRQLWIENSE
VSAGWSLQHQSVITGVPRNEWKLHVPARTCVDVVPVRQGGVRGWVARPYGFNDVFHGLTNDEHTEFLGL
PLTQWLQERGLSLDALAAGKPAVDIQYAALFPWCASVEDLGLVLRWMIAANVACDMEPGDAAHAKRIWEAS
TRYSANDLNDVADVAAVLESREAFRREVLPVMAAHAHRNPFYQLDLNHTAQKYAAAHLPLPERLSADEAPM
MQRIHHHMFCARVLQYLLALLPKDACPARGTRLAEVAMTTSDEAEGGGNCLPPLLQGATMTPELLLSVYGIA
SKEELVARAGEEEAAAFMLLQKAILKHLIASTVRPSPRLSVYSDQIVWGRGPARIDLSGGWTDTPPYSILCGG
NVVNIGINLNGQPPLHVYVKSSPTPSITLRSIDLGAAETLSTYEELRRYDMVGSPFSIPKAALALAGFLPEFGA
MAYGTLEEQLQTSFCGKGLEITLLVAIPAGSGLGTSSLVAATVLSALSDYCGLGWDAHEVGRRTLALEQLLTT
GGGWQDQYGGLFRGLKLLQSTPGFSQSPTARWLPEHLMEDARYAQCHLLYYTGITRTAKVILADIVRGMFL
NSASHLGVLAEMRQQALDLHDAMAHDDFERYGKLIAAAWDQNKRLDAGTCPPPIAEIISRVEKYVWGLKLAG
AGGGGYLYMVAKDTEAARCIREELTEHPPNATARFVEMSVSHEGIQVSRS

>CAJ1987631.1 pyrophosphorylase | fucose kinase | FKP [Leishmania donovani]
MNVSFLFSLPRSLAEFKGAGKLLADAAALPLGERGAKGDVAVPGLFCGCDPEDRPLGSGGGTVHLLHACYV
DEQRHAPASAKRSFLSWLRLADNDGRVIVHAGGLSRRLPAYAAMGKALLPLPPCRWHRGSQLSRTLLSTQV
PMYRGIVTTAPARLRTLIACGDVCVDARGALPSLAPFADSDVLCFGIRADADLLQNHGVLFMARERPLDLDY
MLQKPSLDEVRRRVAEGRHSVLLDVGMWMLSDRAVEVLARKCLGGALDAEDATGGGHDGHRCTAVRTAY
DLYSEFGSALGNHAPSADPEIAGLKASVVELRDAEFLHYGTNRQLISAAATLQGRESPTPPWSWAAPFDGVL
APAGGATEELDFAAGDAMPLLATPSSVIVQNSVVLAPLDGSAAAAAADAPAGAPLSPVRHLWVESSWVGAR
WVLRDQHILTGIPRNDWALALPPGACVSVVPVLPAALGGCGGGAAPHAARPYHMDDAFRGDVRSGSTMYM
GVRLHDWLSSRGFSVAELYPAARAADGGAAQPLDINEAALFPVCATEQQLGDFLYVATLPLEAATGGCAGA

DAARLAAGQAVWRTQPRVSAMDLLRLTDVPAMLRNSEYYERRMLTLMLALVASDPAVLCARSLVLSGPMIA
LLQQQFFLLDLNRVAQRVVELGVPLPRRSGPTEWWGGVDAAGNAVALSRQAAPAAPLAKGAAASAGSWA
QMFSDEGSPSGGGRGAPLCTAPGRIVAAHYHAFASRLMELALEAMDAAEPPVSAAEKARRLALLRAPSAEE
ARSLVALHDAAALAELRAAIMDSFPDERHDPEMGVHLDQIVWGRCPIRIDVAGAWTDTPPYAILSGGSVINVA
VELNGQPPVQVYVRVREDPIIVMHSIDSGEQLSVSSFDEIRTYATMQNPFSIPKAALALCGFLPEFCATEYATL
REQLAARFGGHGLEISLFVAIPTGSGLGTSSIVAGTVLRSLAEFCKLPWDNHDVCRRVLLIEQMLTAGGGWQ
DQYGGLFEGLKLVQCVPGLPCLPTVRWMPDSVYTDPRFAACHLLYYTGITRMAKSILGEIVRDVFLNNGATL
QLLREMGGPTTAAMYDAITAGNYKGYARLVHRTWEQKKRLDDGVCNPAVQSIVDVVEPYVWGLTLPGAGG
GGYMYMCAKDEACARRIRELLTANPPNGNARFVEMSVSASGLQISRS

>DAL00536.1 MAG TPA: bifunctional fucokinase/L-fucose-1-P-guanylyltransferase, partial [Caudoviricetes sp.]
MKDCAKEEQAAFQLLRDGLLGVMSERKSHPILNVYSDQIVWGRSPVRIDVAGGWTDTPPYSLYSGGSVVNL
AIELNGQPPLQVYVKPCKEYHITLRSIDMGAMEVIRNYEELQDYKKVGSPFSIPKAALTLAGFAPAFSTESYPS
LAKQLEDFGSGIEITLLAAIPAGSGLGTSSILASTVLGAINDFCGLAWDKNDICSYTLVLEQLLTTGGGWQDQY
GGVFSGIKLLQSEAGFEQNPLVRWLPDQLFVHPDYRDCHLLYYTGITRTAKSILAEIVSSMFLNSGPHLSLLA
EMKAHAMDMSEAILRSNFESFGRLVGKTWIQNQALDCGTNPPAVAAIIEKIKDYTLGYKLPGAGGGGYLYMV
AKDPQAAGQIRRILTEQAPNPRARFVEMTLSDKGLQVSRS

>WP_197171522.1 bifunctional fucokinase/fucose-1-phosphate guanylyltransferase [Novipirellula aureliae]
MSLDPKLLLSLPPQMVSQLEICHPDVARRSFSTSDPAETQLGSGGGTAHVLHQAWLHSDSDSLSQWVADH
RQIMIHGGGESRRLPAYAAAGKLFIPIPTLRWSRGQRLGQTLLDFNEPFLRSVFEQAGSKARLLIASGDVLLR
STRPLPNLPDADVVLLGMWADPEMAANFGVMFVEKAQPERLQTFLQKPDPDEIRDRSRDSAFLIDVGVWLL
SARAIECLMTQCGWDAAEAKFDRSEHPHPCDLYGHWALRLGENPQAFDKTISELSVAVAPLSQGEFYHFGK
TGDVIDSMYDLQTIVRDHTELGLVPSLAQPRQFVQDAHFGVPLRRQQNESLWVENCHVPACWSLSRRHML
TGVPENEWSLNLPEGICLDFVPIEDTQVAIRVYGFSDAFRGKISHLDTHWLERSAMQWFLDRGLDFDATEISA
DDDLQESALFPVFEIESLDGDFVQWMVDVDSSDQPTKQRSVWQKARKLSARELAHHARLDRVYAERLAKR
EAILPIMASHGNRSVFYNLDLAHVASTYANSDAAVPEYWDESEDLILAVHNRMFRSEVLRNRGDQAWLAEQ
AACFGLLERSIVEPYHSERVVPECRLAEDQIVWARSPARVDLAGGWTDTPPYCLEHGGSVVNIALDLNGQP
PVQVFARRCEKPEITIRSIDLGSMETLTSYEEVGGFRGIGSGFSVAKAAVALAGFHPHFNGDAFVSLQRQLER
FGGGLDVSMLAAIPKGSGLGTSSILAGTVLGALNELCSLGWDAHQIVARVSAVEQMLGSGGGWQDQFGGIL
PGAKLIQTTPGMSQAAAVRWLPSDFFRTPEFTSRSLLYYTGITRVAHNVLSEIVRGMFLNDPARLSVLDRIGE
NSLECFDAAQRSDMSVFNRSIRRSWELNQRLDEGTNPAEVSQLVERVESMCSSLKLSGAGGGGFLYMIAR
DAESAERIRRDLENDPPNAGARFVDMAISTTGLRVTRS

>EPF19498.1 bifunctional fucokinase/L-fucose-1-P-guanylyltransferase [Microcystis aeruginosa SPC777]
MIIRTKAPLRLGLGGGGTDVEPYCSLYGGYVLNATIDLYAYCTVTEITDKQVIFVAADRGEQQIETLQFPLPYN
GVLDLHKAVYNRLIKDFNQGQPLPLQITTFSDAPAGSGLGSSSTLAVAMVKAFVELLKLPLGEYDIAHLAYEIE
RIDLSWLGGKQDQYAATFGGFNFMEFYEQDRVIVNPLRIKNWVINELEVSLILYYTGISRYSSQVIEDQIQNVQ
EKNEQAIAATHQLKKEAILFKEALLKSDFMGIAEILRTSWEAKKKLSKLISNPQIDRIYRVARETGAYSGKISGA
GGGGFIIFMVDPTKKIEVTKALKSEGGQVINFHFTKYGTQAWTL

Scientific Contributions

The work presented here has led to several contributions to scientific meetings, theses, and peer-reviewed journals, all listed in this section.

Peer-Reviewed Journal Articles[ii]

- Nicolas Huber† & **Edgar Alberto Alcalá-Orozco**†; Thomas Rexer; Udo Reichl; Steffen Klamt (2024). Model-based Optimization of Cell-free Enzymatic cascades Exemplified for the Production of GDP-fucose. Metabolic Engineering, doi: 10.1016/j.ymben.2023.10.007

- **Alberto Alcala-Orozco**, E., Grote, V., Fiebig, T., Klamt, S., Reichl, U., & Rexer, T. (2023). A Cell-Free Multi-enzymatic cascade for the Synthesis of CDP-glycerol. Chembiochem, doi:10.1002/cbic.2023.00.463

- *Espinel-Ríos, S., Huber, N., **Alcalá-Orozco, E. A.,** Morabito, B., Rexer, T. F. T., Reichl, U., . . . Findeisen, R. (2022). Cell-free biosynthesis meets dynamic optimization and control: a fed-batch framework. IFAC-PapersOnLine, 55(23), 92-97. doi:10.1016/j.ifacol.2023.01.021

- *Litschko, C., Budde, I., Berger, M., Bethe, A., Schulze, J., **Alcala Orozco, E. A.**, . . . Fiebig, T. (2021). Mix-and-Match System for the Enzymatic Synthesis of Enantiopure Glycerol-3-Phosphate-Containing Capsule Polymer Backbones from Actinobacillus pleuropneumoniae, Neisseria meningitidis, and Bibersteinia trehalosi. mBio, 12(3), e0089721.

Supervision of Theses

- Briggith Uribe (2024). Development of an enzymatic cascade for the synthesis of the human milk oligosaccharide LNFP III. Master Thesis. Universidad Técnica de Manabí, Ecuador.

- Jesus Pardo (2023). Process development for the biocatalytic synthesis of CMP-Neu5Ac. Master Thesis. Aalborg University, Denmark.

- Jasmin Vogel (2022). Development of a cell-free multi-enzyme system for the synthesis of CMP-Neu5Ac. Bachelor Thesis. Hochschule Flensburg, Germany.

Conference Speaker Contributions

- **Edgar Alberto Alcalá-Orozco**, Nicolas Huber, Steffen Klamt, Thomas Rexer, Udo Reichl. Development of a cell-free enzymatic cascade for the synthesis of GDP-fucose - modelling and optimization. IV Spanish Biocatalysis Conference 2024. San Sebastian, Spain.

---

ii

*Related research collaborations which content is not directly used in this dissertation.
†equal contribution

- **Edgar Alberto Alcalá-Orozco**, Nicolas Huber, Steffen Klamt, Thomas Rexer, Udo Reichl. Development of a cell-free enzymatic cascade for the synthesis of GDP-fucose - modelling and optimization. Multi-Enzyme Catalyzed Processes (MECP) 2024. Vienna, Austria.
- **Edgar Alberto Alcalá-Orozco**, Nicolas Huber, Steffen Klamt, Thomas Rexer, Udo Reichl. Development of a cell-free enzymatic cascade for the synthesis of GDP-fucose - modelling and optimization. 3rd Japan-Switzerland-Germany Workshop on Biocatalysis and Bioprocess Development (JSG) 2023. Nagoya, Japan.
- **Alberto Alcalá**, Nicolas Huber, Steffen Klamt, Thomas Rexer, Udo Reichl. Development of a cell-free enzymatic cascade for the synthesis of GDP-fucose - modelling and optimization. Glycobiotech 2022. Berlin, Germany.
- **Alberto Alcala**, Timm Fiebig, Reza Mahour, Valerian Grote, Tuan Son, Thomas Rexer, Udo Reichl. A cell-free multi-enzymatic cascade for the synthesis of CDP-glycerol. Multi-Enzymatic cascade Processes (MECP) 2021. Aachen, Germany (online).

Conference Poster Contributions

- T.F.T. Rexer, **A. Alcala**, S. Kleeberg, T.H. Son, U. Reichl. Platform for the Scalable Synthesis of Activated Sugars and Complex Human Milk oligosaccharides. The Origins and Benefits of Biologically Active Components in Human Milk Conference (FASEB) 2023. Melbourne, USA.

- **Alberto Alcalá**, Tuan Hoang Son, Thomas Rexer and Udo Reichl. In vitro multi-enzymatic cascades for the synthesis of nucleotide sugars. 11th International CeBiTec Research Conference (IBPRO) 2022. Bielefeld, Germany (online).

- **Alberto Alcalá**, Nicolas Huber, Steffen Klamt, Thomas Rexer, Udo Reichl. Development of a cell-free enzymatic cascade for the synthesis of GDP-fucose - modelling and optimization. Glycobiotech 2022. Berlin, Germany.

- **Alberto Alcala**, Timm Fiebig, Reza Mahour, Valerian Grote, Tuan Son, Thomas Rexer, Udo Reichl. A cell-free multi-enzymatic cascade for the synthesis of CDP-glycerol. 10th International Congress on Biocatalysis (BIOCAT) 2022. Hamburg, Germany.

- **Alberto Alcala**, Timm Fiebig, Reza Mahour, Valerian Grote, Tuan Son, Thomas Rexer, Udo Reichl. A cell-free multi-enzymatic cascade for the synthesis of CDP-glycerol. Multi-Enzymatic cascade Processes (MECP) 2021. Aachen, Germany (online).

Declaration of Honor

I hereby declare that I produced this thesis without prohibited external assistance and that none other than the listed references and tools have been used.

In the case of co-authorship, especially in the context of a cumulative dissertation, the own contribution is correctly and completely stated. I did not make use of any commercial consult-ant concerning graduation. A third party did not receive any nonmonetary perquisites neither directly nor indirectly for activities which are connected with the contents of the presented thesis. All sources of information are clearly marked, including my own publications.

In particular, I have not consciously:

- Fabricated data or rejected undesired results.
- Misused statistical methods with the aim of drawing other conclusions than those warranted by the available data.
- Plagiarized data or publications.
- Presented the results of other researchers in a distorted way.

I do know that violations of copyright may lead to injunction and damage claims of the author and also to prosecution by the law enforcement authorities.

I hereby agree that the thesis may need to be reviewed with an electronic data processing for plagiarism.

This work has not yet been submitted as a doctoral thesis in the same or a similar form in
Germany or in any other country. It has not yet been published as a whole.

Belgium, Dec. 5th 2025

Disclosure of scientific publications

- Nicolas Huber† & **Edgar Alberto Alcalá-Orozco**†; Thomas Rexer; Udo Reichl; Steffen Klamt (2024). Model-based Optimization of Cell-free Enzymatic cascades Exemplified for the Production of GDP-fucose. Metabolic Engineering, doi:10.1016/j.ymben.2023.10.007

  -The methods and results presented for the kinetic modelling have been taken from the original publication.
  -The following figures were reproduced from the original publication: Figure 33, Figure 34, Figure 35 and Figure 36.

- **Alberto Alcala-Orozco**, E., Grote, V., Fiebig, T., Klamt, S., Reichl, U., & Rexer, T. (2023). A Cell-Free Multi-enzymatic cascade for the Synthesis of CDP-glycerol. Chembiochem, doi:10.1002/cbic.2023.00.463

  -The methods and results presented for the CDP-glycerol cascade development been taken from the original publication.
  -The following figures were reproduced from the publication: Figure 7, Figure 9, Figure 10, Figure 11, Figure 13, Figure 14, Figure 15 and Figure 16.

†equal contribution