國立臺灣大學生命科學院生化科學研究所

碩士論文

Institute of Biochemical Sciences

College of Life Science

National Taiwan University

Master's Thesis

分枝桿菌噬菌體的 O-醣基化之酵素特性描述嘗試

Attempted Characterization of O-Glycosylation Enzymes in Mycobacteriophages

韓東均

DongKyun Han

指導教授：呂桐睿 博士

Advisor: Todd L. Lowary, Ph.D.

中華民國 113 年 7 月

July 2024

doi:10.6342/NTU202402746

國立臺灣大學碩士學位論文
口試委員會審定書
MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

分枝桿菌噬菌體的 O-醣基化酵素之特性描述嘗試

Attempted Characterization of O-Glycosylation Enzymes in Mycobacteriophages

本論文係韓東均君（R11B46026）在國立臺灣大學生化科學研究所完成之碩士學位論文，於民國 113 年 7 月 23 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Institute of Biochemical Science on 23rd July 2024 have examined a Master's thesis entitled above presented by Han DongKyun (R11B46026) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

（指導教授 Advisor）

系主任/所長 Director:

doi:10.6342/NTU202402746

Acknowledgement

I would like to extend my sincere gratitude to all those who have supported me throughout the course of my master's thesis. This work would not have been possible without the encouragement, guidance, and assistance from many individuals and organizations. First and foremost, I would like to express my deepest appreciation to my advisor, Dr. Todd L. Lowary, whose expertise, patience, and unwavering support were invaluable throughout this research. Your insightful feedback and encouragement guided me through the complexities of this project. Special thanks to Institute of Biological Chemistry, Academia Sinica for providing the necessary resources, environment for my research and for their financial support which made this study possible. I am immensely thankful to my colleagues and fellow researchers, Anita Wu for the experimental support on sugar nucleotide synthesis, Min-Yu Chiang and Lourriel Macale for the technical assistance with protein cloning and expression, and all of the lab members for their stimulating discussions, and for making the research journey enjoyable. Your cooperation and willingness to share your knowledge were vital in the completion of this thesis. I would like to extend my gratitude to the staff of Institute of Biological Chemistry Synthesis Core Facility (IBCSCF), particularly Ms. Yu-Ling Hwang, for their assistance with acceptor peptide synthesis. To my lovely girlfriend, Shang Yi Jiang, thank you for your emotional support and for providing love and laughter during this journey. Finally, I owe my deepest gratitude to my family for their unconditional love, encouragement, and sacrifices. Your belief in me was a constant source of strength and motivation. This thesis is dedicated to all those who supported me and believed in my potential. Thank you.

doi:10.6342/NTU202402746

摘要

分枝桿菌噬菌體是一群多樣化的噬菌體，他們能夠選擇性地感染分枝桿菌，包括結核分枝桿菌。最近已發表的研究揭示了分枝桿菌噬菌體生物學一個獨特的特性—部分噬菌體頭部或尾管具有 O-醣基化。特殊的是這些噬菌體表的醣基化是由噬菌體自體編碼的醣基轉移酶所轉移的，與一般對病毒的認知不同，這些噬菌體並不會通過利用宿主的機制進行修飾，也因此導致不同分枝桿菌噬菌體之間具有不同的表面糖類之修飾。這些醣鏈通常很長、結構複雜且在代謝扮演著重要的角色。這些長醣鏈可以作為噬菌體的醣護盾，使病毒顆粒不易受抗體結合的影響。在小鼠接種後的初始階段，IgM 和 IgG 抗體對未醣基化的噬菌體顆粒的親和力比對醣基化的病毒顆粒更強。雖然已經有論文發表了分枝桿菌噬菌體表面糖類的存在以及負責對病毒結構蛋白進行醣基化的噬菌體編碼醣基轉移酶，但仍有幾個方面尚未清楚，包括這些酵素的功能、基質特異性以及這些噬菌體表面醣類的化學組成。在這篇論文中，我試圖證明分枝桿菌噬菌體中醣基轉移酶的功能。首先，我通過使用各式生物訊息及結構工具與軟體鑑定出多個醣基轉移酶候選基因，再將所鑑定的醣基轉移酶在預測出的三維結構模型中找出保守的醣基轉移酶家族結構胺基酸序列，這些序列代表了潛在的醣基轉移酶功能位點。進一步的結構分析揭曉了這些醣基轉移酶具有類 Rossman 折疊，這是 GT-A 或 GT-B 家族的醣基轉移酶家族所擁有的特徵。其中的一組醣基轉移酶來自於分枝桿菌噬菌體 Corndog（gp36、gp37 和 gp38），這組基因被挑選出來對其進行了克隆並且成功表達出這三個酵素。再使用合成的核苷酸醣和胜肽作為受體對 Corndog 的醣基轉移酶進行功能測試。儘管在不同的受體設計和反應條件下進行了多次嘗試，但 Corndog 醣基轉移酶未能將醣基轉移到受體上。這項結果表明幾種可能性：這些酵素的供體和受體特異性可能比最初預期的更複雜，或者用於測試的體外條件未能模擬酵素活性所需的環境。供體與受體的專一配對在本研究中是一個重大的挑戰，也因此成為了這些未知功能的醣基轉移酶功能測試的瓶頸。

關鍵詞：醣生物學、分枝桿菌噬菌體、O-醣基化、醣基轉移酶、酵素合成。

doi:10.6342/NTU202402746

Abstract

Mycobacteriophages are a diverse group of bacteriophages that infect mycobacteria, including *Mycobacterium tuberculosis*. Recent investigations have revealed a fascinating aspect of mycobacteriophage biology, the O-glycosylation of capsid and/or tail tube subunits. This viral surface glycosylation is orchestrated by phage-encoded glycosyltransferases, bypassing the need to exploit the host's biosynthetic machinery, resulting in a distinct surface glycans across different mycobacteriophages. These glycans, often large and intricately structured, are proposed to play a pivotal role as a glycan shield, rendering viral particles less susceptible to antibody binding. Both IgM and IgG antibodies exhibit a stronger affinity for non-glycosylated phage particles than to glycosylated virions during the initial stages following mouse inoculation. While the presence of mycobacteriophage surface glycans and the phage-encoded glycosyltransferases responsible for glycosylating viral structural proteins have been discovered, several aspects remain fairly unknown. These include the precise function of these enzymes, their substrate specificity, and the chemical composition of the surface glycans. In this study, the primary focus is to demonstrate the function of these glycosyltransferases. Multiple glycosyltransferase candidates were identified by using of a number of online bioinformatic and structural tools and software. The identified glycosyltransferases exhibited conserved glycosyltransferase family structural motifs in high-confidence three-dimensional models showing potential functional sites. Further structural analysis suggests that these glycosyltransferases are likely to be soluble enzymes belonging to the GT-A or GT-B families with Rossman-like folds, which are characteristic of these glycosyltransferase families. A set of glycosyltransferases from

doi:10.6342/NTU202402746

these candidates belonging to the mycobacteriophage Corndog (gp36, gp37, and gp38) was successfully cloned with relatively high efficiency. The functional assays of the Corndog glycosyltransferases using synthesized nucleotide sugars and peptide as acceptors, although initially promising, did not result in detectable glycosylation. Despite multiple attempts with different acceptor designs and reaction conditions, the glycosyltransferases were unable to transfer sugar residues to the acceptors. This outcome suggests several possibilities: the donor and acceptor specificities of these enzymes may be more complex than initially anticipated, or the *in vitro* conditions used in the assays failed to simulate or fully replicate the native environment required for enzymatic activity. The identification of specific donor-acceptor pairs remains a significant challenge and a potential bottleneck in demonstrating the precise functionalities of these novel enzymes.

Keywords: Glycobiology, Mycobacteriophage, O-glycosylation, Glycosyltransferase, Enzymatic synthesis.

doi:10.6342/NTU202402746

Table of Contents

doi:10.6342/NTU202402746

doi:10.6342/NTU202402746

doi:10.6342/NTU202402746

List of Figures

doi:10.6342/NTU202402746

doi:10.6342/NTU202402746

doi:10.6342/NTU202402746

List of Schemes

doi:10.6342/NTU202402746

List of Tables

doi:10.6342/NTU202402746

1. Introduction

1.1 Mycobacteria and Mycobacteriophages

1.1.1 Mycobacteria

Mycobacterium is a genus of bacteria that includes a wide range of species. There are more than 195 different species of *Mycobacterium* and about 30 of them are pathogenic to humans [1,2]. One of the most notable members of this genus is *Mycobacterium tuberculosis*, the bacterial species responsible for tuberculosis (TB), a disease that has haunted mankind for centuries. *Mycobacterium tuberculosis* was first described as the causative agent of tuberculosis by Robert Koch in 1882 [3]. Since its discovery over 140 years ago, TB still remains a significant global health concern, more specifically in developing countries where access to healthcare and resources for medical treatment may be greatly limited [4,5]. According to the *2023 Global Tuberculosis Report* from the World Health Organization (WHO), TB was the second leading infectious disease worldwide post COVID-19, with an estimated 10.6 million people infected and 1.3 million deaths worldwide including 167,000 people with HIV in 2022 [4]. To add onto that, it was shown that individuals who have recovered from COVID-19 have higher risk of developing TB [6].

In addition to tuberculosis, other mycobacterial species can cause serious human disease. One example is *Mycobacterium leprae*, which causes leprosy (also known as Hansen's Disease), a chronic infectious disease that affects the skin, nerves, and mucous membranes [7]. Despite our greatest efforts to control and eliminate these diseases, *Mycobacterium*-related diseases continue to pose challenges to humanity worldwide. Therefore, developing new drugs and treatments against mycobacterial diseases remains an urgent global research priority.

doi:10.6342/NTU202402746

1.1.2 Bacteriophages

Bacteriophages, or phages, discovered by Frederick Twort in 1915 [7] and Félix d'Hérelless in 1917 [8], are a diverse group of viruses that specifically infect bacterial hosts to replicate. Their dynamic interactions with bacteria make them crucial members of microbial ecosystems and an important driver of bacterial evolution [9]. The most common way of classifying bacteriophages is by looking at their biological cycle, i.e., lytic or virulent and lysogenic or temperate [10,11]. In the lytic cycle, phages infect and kill the bacterial hosts by releasing endolysin and holin proteins after particle replication [12]. In the lysogenic cycle, the phages fuse its genome into the host's genome and replicate without making any viral particles [13]. For many phages, the lytic cycle is the only mode of replication. However, some temperate phages can enter the lytic cycle just like the lytic phages, while others go through lysogenic cycle upon infection (Figure 1-1) [14].

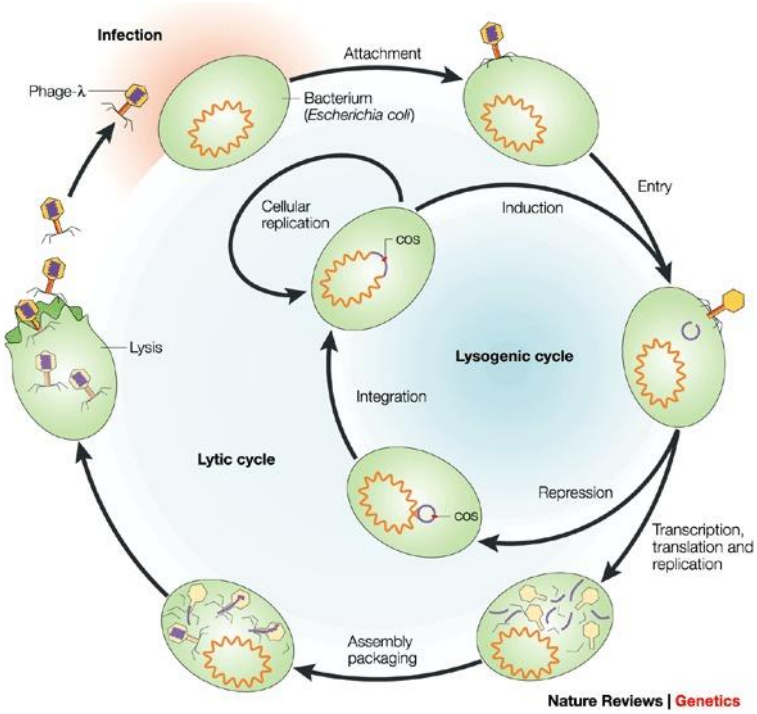


Figure 1-1 Bacteriophage biological cycles.

Reprinted with permission from Springer Nature[14].

doi:10.6342/NTU202402746

One of the key characteristics of bacteriophages is their ability to rapidly evolve in response to host mutations including changes in host susceptibility due to host mutation, environmental conditions, and co-evolutionary interactions with other phages and bacteria [9]. This adaptive ability promotes the ongoing arms race between phages and their bacterial hosts, thus increasing the diversity of phage populations over time and making them one of the most abundant viral particles in the world at around 10^{31} total phage particles in the entire biosphere [15].

1.1.3 Mycobacteriophages

Mycobacteriophages are phages that specifically infect mycobacteria, including the species responsible for tuberculosis and leprosy. While relatively small (50–200 nm), mycobacteriophages can be visualized by performing electron microscopy. The majority of these phages are classified into the *Siphovirdae* family (long, noncontractile tails), a few in *myovirdae* family (contractile tails) and none in *Podoviridae* family (short, stubby tails) (Figure 1-2) [16].

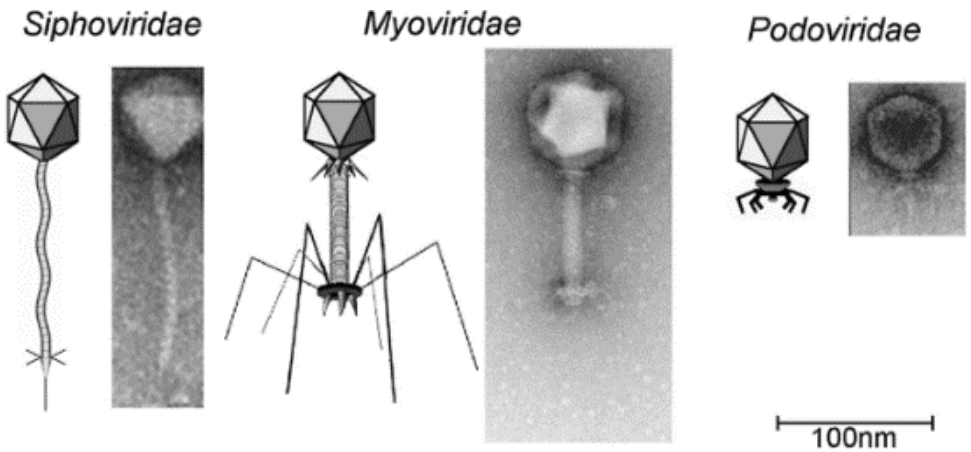


Figure 1-2 Tailed bacteriophage classifications.

Reprinted with permission from Springer Nature[16].

doi:10.6342/NTU202402746

Mycobacteriophages can also be further classified by their genomes into different cluster and subclusters. As of now, there are 34 known clusters, 73 known subclusters and 73 singletons on the largest mycobacteriophage database (PhagesDB available at https://phagesdb.org/) [17]. Phages that are grouped in the same clusters have higher genomic content similarity, while new phages identified without any similarity to other existing phage clusters, they are classified as singletons [18]. Only a selected few clusters of these phages, for instance phages in cluster A2, F1, and K2, are so far used medically as treatments mycobacterial diseases [19].

1.2 Phage therapy

1.2.1 Introduction to phage therapy

Phage therapy is an alternative treatment to bacterial diseases that revolves around using bacteriophages to selectively target and kill pathogenic bacteria while leaving human cells unharmed. This concept has been around for over a century. The earliest documented case of phage therapy was in 1919 when Félix d'Herelle successfully treated a patient suffering from dysentery using bacteriophages [8]. The discovery of phage therapy demonstrated the potential of phages to target and kill pathogenic bacteria within the human body. In the early to mid-20th century, phage therapy gained wider attention worldwide and showed many early successes with *in vivo* experiments. These include the study conducted by William Smith and his coworkers reporting the successful use of phages against *Escherichia.coli* in mice [20]. However, phage therapy fell out of favour in around 1940 due to the rising usage of antibiotics and the usually unreliable early trials of phage therapy. The only country that persisted in using phage as treatment was the former Soviet Union [21,22].

doi:10.6342/NTU202402746

In recent years, phage therapy has been brought back under the spotlight due to the rising number of antibiotic resistant strains of pathogenic bacteria being identified. With the use of phages, we could potentially eliminate pathogenic bacteria without exacerbating the already existing issue of multiple drug resistance. However, there are still limitations regarding the therapeutic use of phages including the patient immune response to phages and phage resistance in patients.

1.2.2 Mycobacterial disease phage therapy

It is estimated that around 410,000 people were infected with multidrug-resistant or rifampicin-resistant tuberculosis (MDR/RR-TB) globally in 2022 [5]. Rifampicin, discovered in 1965 and formerly one of the most used and effective antibiotics against TB is now deemed less ideal for TB treatment due to resistance. The rising number of MDR/RR-TB infections possess a serious problem to TB treatment and has made the current treatment regimens much more complicated (Figure 1-3) [23]. While some research groups are focused on developing novel small molecule drugs or inhibitors against MDR/RR-TB [24], others have turned their attention to phage therapy [25]. Mycobacteriophage therapy involves using mycobacteriophages to specifically treat *Mycobacterium*-induced diseases. This treatment usually requires using multiple phages in a mycobacteriophage cocktail, to minimize phage resistance in patients [15]. There are only a handful of mycobacteriophages (including DS6A, TM4, D29, BTCU-1, SWU1, and Ms6) that have been investigated *in vitro* for TB treatment [26], and just two *in vivo* studies with guinea pigs have shown encouraging results [27,28]. Although there is still limited information on using mycobacteriophages to treat TB infections in humans, mycobacteriophages have been used in 20 compassionate cases in treating non-

doi:10.6342/NTU202402746

tuberculous mycobacterium (NTM) *Mycobacterium absessus* infections and with many showing positive results [29] which shows that mycobacteriophage therapy appears to be a promising therapeutic method in treating mycobacterial diseases.

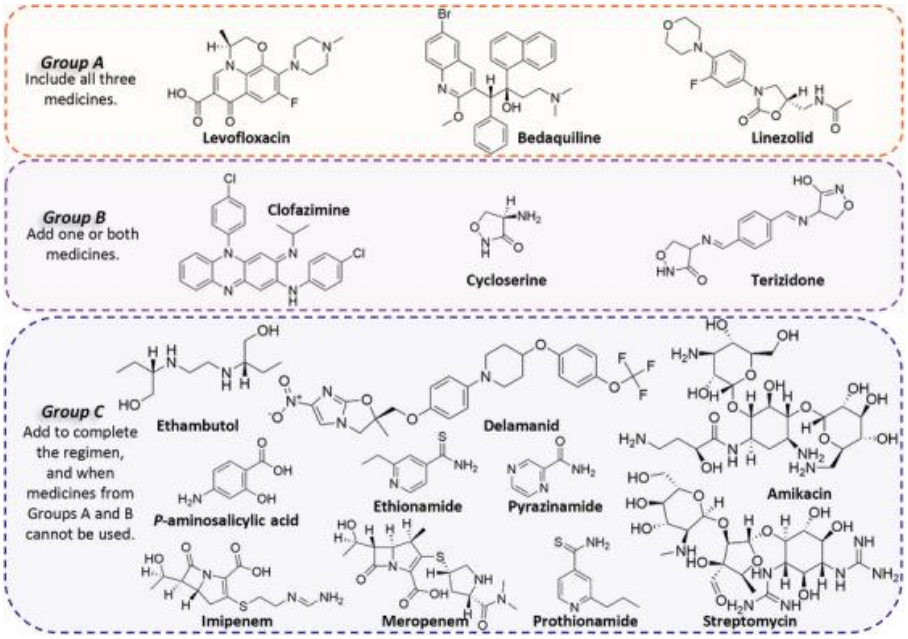


Figure 1-3 Recommended drug use regimen against TB by the WHO. Group A drugs are highly advised for inclusion in all MDR-TB treatment plans, and it is suggested to add one or two Group B drugs to enhance the treatment outcomes. Group C drugs are recommended as a backup choice to Group A and B drugs.

1.2.3 Immune responses in patients

As stated in the previous section, there are some bottlenecks regarding phage therapy in treating mycobacterial diseases. The interaction between phages and the patient immune system is a critical aspect that influences the success of phage therapy. Upon administration, phages may trigger immune responses from the patient, including the production of neutralizing antibodies and activation of the innate immune system [15,30].

doi:10.6342/NTU202402746

While not always the case, these immune responses can heavily impact the pharmacokinetics and efficacy of phage therapy by clearing phage particles or preventing their access to target bacteria [15,30]. Strategies to mitigate immune responses and enhance phage delivery and longevity within the host are crucial areas of research to overcome these challenges. In a study conducted in 2019, a patient was treated with a cocktail of three different phages in order to treat a pulmonary *Mycobacterium absessus* infection [30]. In the beginning, the therapy was quite successful as the bacterial load was lowered after one month of treatment. However, IgG, IgM and IgA antibody responses in the patient to all three phages increased shortly after this period, resulting in the failure of the treatment. (Figure 1-4) [30].

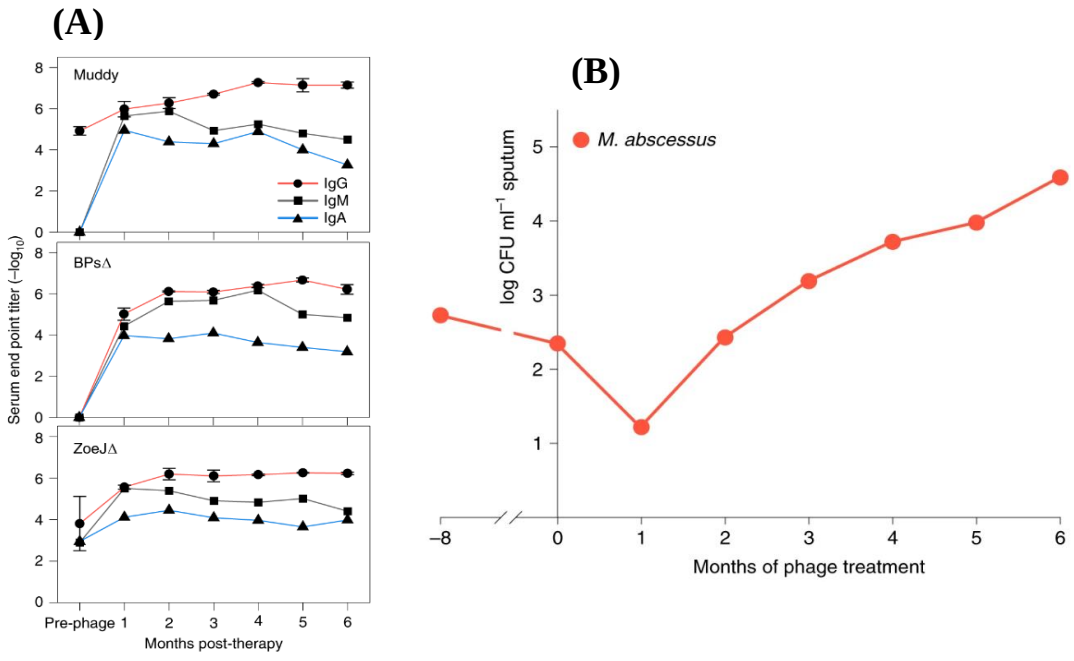


Figure 1-4 (A) Antibody response after phage therapy and (B) *M. abscessus* bacterial load after phage therapy. Reprinted with permission from Springer Nature[30].

doi:10.6342/NTU202402746

1.3 Glycosylation

1.3.1 Glycosyltransferase functions

Glycosylation is a post-translational modification process in which carbohydrate molecules (sugars or glycans), are covalently attached to proteins, lipids, or other biomolecules. This process plays critical roles in various biological functions, including protein folding, stability, cell–cell recognition, immune response modulation, and signal transduction [31,32,33,34]. Glycosylations in bacteriophages were thought to be mediated by the host's own machinery, as bacteriophages are a type of virus, which generally lack the enzymes needed to add sugars to other biomolecules. However, it was recently shown otherwise for some mycobacteriophages as they can also encode glycosyltransferases in their own genome [35].

Glycosyltransferases (GTs) are enzymes responsible for catalyzing the transfer of sugar moieties from activated nucleotide sugar donors to specific acceptor molecules [36]. Glycosyltransferases are involved in diverse biological processes, including the biosynthesis of glycoproteins, glycolipids, and proteoglycans, as well as modification of cell surface molecules and extracellular matrix components. Glycosyltransferases can be grouped into multiple families, each possessing unique three-dimensional structural characteristics and currently already categorized into more than 110 families in the Carbohydrate-Active Enzymes (CAZy) database (available at http://afmb.cnrs-mrs.fr/CAZY). With each having different substrate specificities, these enzymes contribute to the huge diversity of glycan modifications [37,38].

doi:10.6342/NTU202402746

1.3.2 Glycosyltransferase structures

Glycosyltransferases exhibit diverse structural motifs and catalytic mechanisms which contribute to the large functional diversity of the glycans they produced. Based on their structural characteristics, glycosyltransferases are classified into three families, GT-A, GT-B, and GT-C (Figure 1-5) [39].

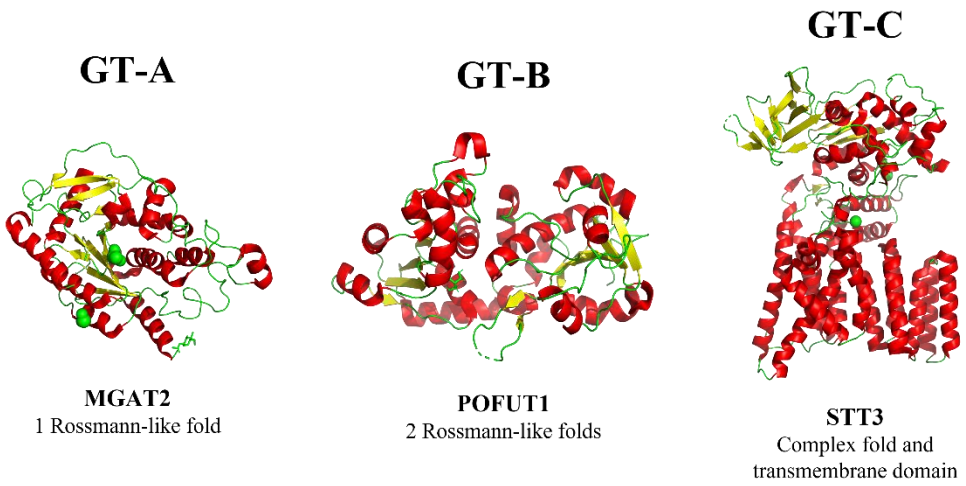


Figure 1-5 General structures of GT-A, GT-B, and GT-C GTs.

1.3.2.1 GT-A

GT-A glycosyltransferases typically are found with a single Rossmann-like fold and contain conserved a 'DxD' amino-acid motif involved in nucleotide sugar binding and catalysis. Most, but not every, member of GT-A requires a divalent cation for activity [40]. These enzymes catalyze the transfer of sugar moieties from nucleotide sugar donors to acceptor molecules such as peptides with a single sugar added each time. GT-A glycosyltransferases are involved in various glycosylation processes including N-glycosylation, glycosylphosphatidylinositol (GPI) anchor biosynthesis, and O-glycosylation [39,40.41].

doi:10.6342/NTU202402746

1.3.2.2 GT-B

GT-B glycosyltransferases are characterized structurally by containing two Rossmann-like folds. Unlike GT-A glycosyltransferases, the GT-B do not require divalent cations for activity nor do they contain DxD motifs [40]. These enzymes are involved in the biosynthesis of diverse glycoconjugates and are responsible for much of the glycosylation of proteins and lipids in membranes including glycolipids, glycosaminoglycans, and bacterial cell wall polysaccharides. GT-B glycosyltransferases often exhibit flexibility in substrate specificity and can accommodate a wide range of acceptor substrates [39,40,41].

1.3.2.3 GT-C

GT-C glycosyltransferases represent a group of enzymes with varying structural architectures and often found with complex folds [40]. These enzymes contain a bundle of transmembrane α-helices and recognize lipid-linked sugar donors. These enzymes can be found in endoplasmic reticulum (ER) or in cell membranes [39,40,41].

1.3.3 O-glycosylation

Protein O-glycosylation is the process of the attachment of sugar moieties to amino acids with a hydroxyl functional group, most commonly serine or threonine residues, in a peptide or a protein [42]. This type of glycosylation exists in all domains of life [43,44,45]. In organisms with complex cells (eukaryotes), this process happens in different places like the endoplasmic reticulum and Golgi apparatus, and sometimes in cytoplasm. But for simpler cells (prokaryotes), it only happens in the cytoplasm [42]. This type of glycosylation significantly impacts protein stability, trafficking, and function [46].

doi:10.6342/NTU202402746

O-glycosylation pathways involve a diverse arsenal of glycosyltransferases and glycosidases responsible for the addition and removal of sugar moieties from proteins. The focus of this thesis is glycosyltransferases.

1.3.4 O-glycosylation in mycobacteriophages

O-glycosylation is a common post-translational modification in viral proteins. Viral proteins with such sugar additions can undergo protein conformational changes, antigenicity, and change in interaction with host cell receptors [47]. Recently some mycobacteriophages were shown to modify their capsid and/or tail proteins using their own glycosyltransferases (Figure 1-6) [35]. This was quite a unique discovery since viruses typically perform protein post translational modification via hijacking the modification machinery of the host. MS analysis results revealed that the mycobacterial surface O-glycosylation happened on the C-terminal serine residue of either the capsid or major/minor tail proteins. Furthermore, MS analysis also showed the numbers of HexNAcs and Hexoses that were attached on the surface of each mycobacteriophages (Che8, Myrna and Corndog). However, due to the insufficient amounts of the materials, performing Nuclear magnetic resonance spectroscopy (NMR) was not possible, leaving the exact compositions of these surface glycans remain undetermined. Interestingly, *in vivo* mice experiments, resulting from injecting phages with or without surface glycosylation showed both IgM and IgG antibodies exhibit a lower affinity for O-glycosylated phage particles (Figure 1-7) [35]. It was proposed that these surface glycosylations work as a glycan shield in preventing efficient binding of antibodies to the phage particles.

doi:10.6342/NTU202402746

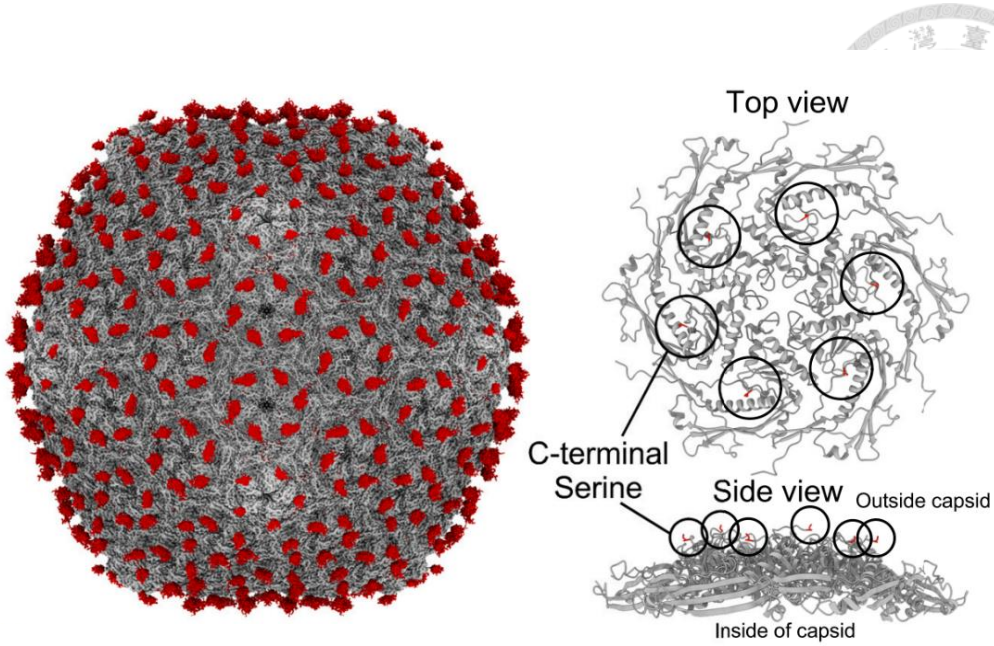


Figure 1-6 Cryo-EM structure of C-terminal serine O-glycosylated mycobacteriophage Che8 capsid. Red densities in the figure indicates the surface sugar modification on the capsid. Reprinted with permission from Cell Press [35].

These findings suggest that mycobacteriophage surface O-glycosylation may help to circumvent the unwanted immune system response in phage therapy. With this new-found knowledge, we could strategically design mycobacteriophages that have better ability to mitigate the immune system response and antibody affinity. However, with over three thousand mycobacteriophages sequenced, only a handful of mycobacteriophage proteins are functionally characterized, leaving most of the proteins, including those glycosyltransferases' functions relatively unknown. There is still much to be explored regarding to the path for a complete understanding of mycobacteriophage surface O-glycosylation.

doi:10.6342/NTU202402746

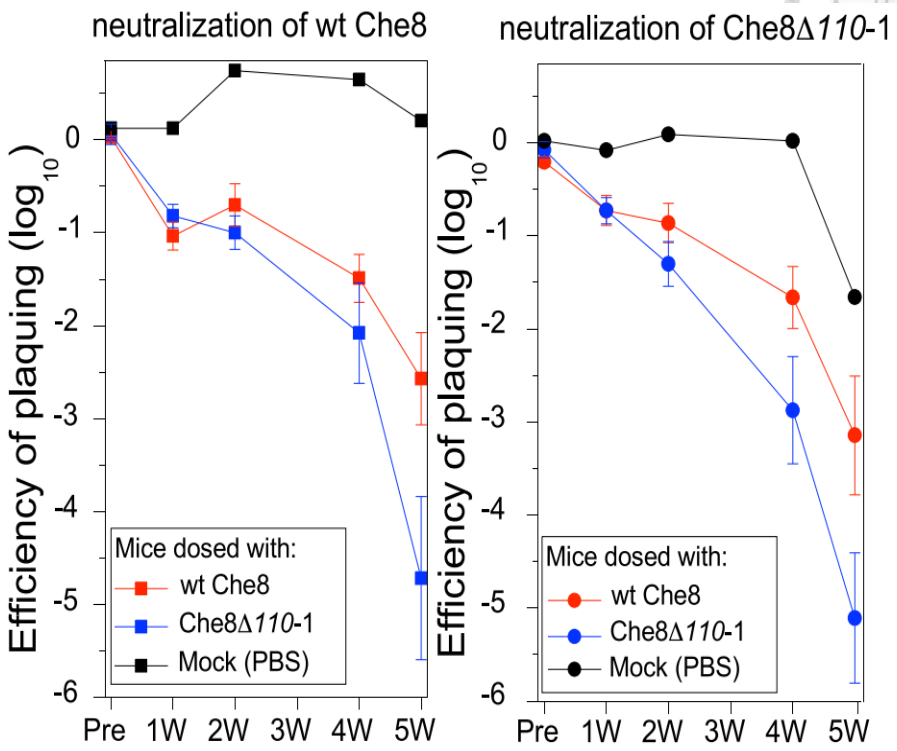


Figure 1-7 Phage neutralization of wildtype and mycobacteriophage Che8 with glycosyltransferase deletion (Che8Δ110-1) in mice serum upon phage injection. Reprinted with permission from Cell Press [35].

doi:10.6342/NTU202402746

1.4 Aims of this study

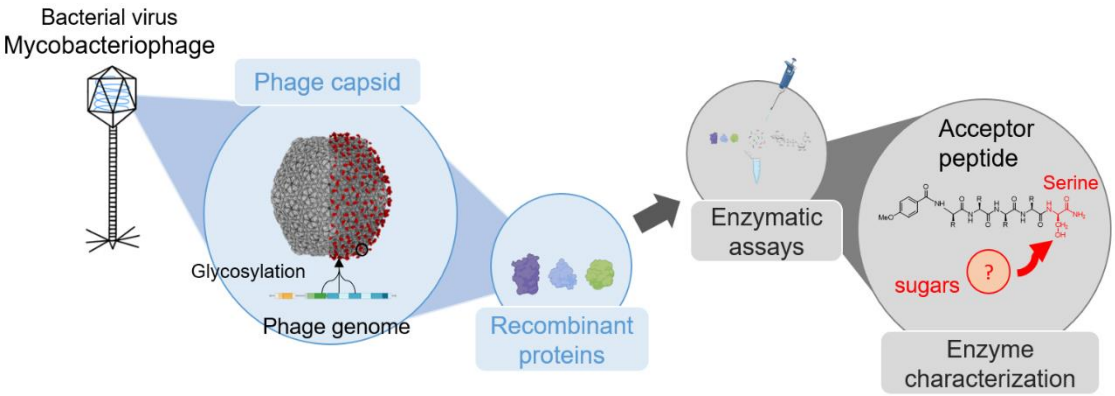
Phage therapy is a potential treatment that could effectively help the pressing issue of multi-drug resistant TB infections. To further optimize the utility of this treatment strategy, mitigating the unwanted immune response to phages in patients is foremost. In order to do so, studying the potential use of glycosylated phages as new "shielded" phage therapy agents is of great interest. However, there are still number of important questions left unanswered in this research area which, include:

I. Precise function of these phage encoded glycosyltransferases.

II. The substrate specificity of these glycosyltransferases.

III. Chemical composition of the surface glycans.

My research primarily aims to show the function of the glycosyltransferases, which in turn will allow us to better understand the glycan structure. These discoveries will help us learn more about O-glycosylation on the surface of mycobacteriophages and create new opportunities to explore the role of glycosylation in mycobacteriophage.



Scheme 1-1 Aims of this study

doi:10.6342/NTU202402746

2. Results and Discussion

2.1 Genome analysis and structure predictions

To accomplish the aim of this study, I initially compared all of the reported potential mycobacteriophage glycosyltransferase genes to generate a list of mycobacteriophage glycosyltransferase candidates for the functional assay (Table 2-1).

Table 2-1 The list of mycobacteriophages with potential GT genes.

Name	Cluster	Host	GT genes
Phrappuccino	AA	*Mycobacterium*	*142, 143, 154*
Myrna	C1	*Mycobacterium*	*234*
Che8	F1	*Mycobacterium*	*108, 109, 110*
Sbash	I2	*Mycobacterium*	*87, 88*
Omega	J	*Mycobacterium*	*16, 23*
Fionnbharth	K4	*Mycobacterium*	*6*
Corndog	O	*Mycobacterium*	*37, 38*
Marvin	S	*Mycobacterium*	*81, 83*
Nebkiss	X	*Mycobacterium*	*166*
MooMoo	Singleton	*Mycobacterium*	*94, 96*
Sparky	Singleton	*Mycobacterium*	*88, 89*

The genome analysis of mycobacteriophages was conducted using a suite of bioinformatics tools, leading to significant insights into the genetic structure and functional potentials of these viruses. By using the PhagesDB database and Phamerator, I identified and compared the potential GT encoding genes within the complete

doi:10.6342/NTU202402746

Actinobacteriophage genome database. I also extracted conserved domains from the NCBI Conserved Domain Database using RPS-BLAST and found several genes possessing encoding certain GT family motifs (Figure 2-1).

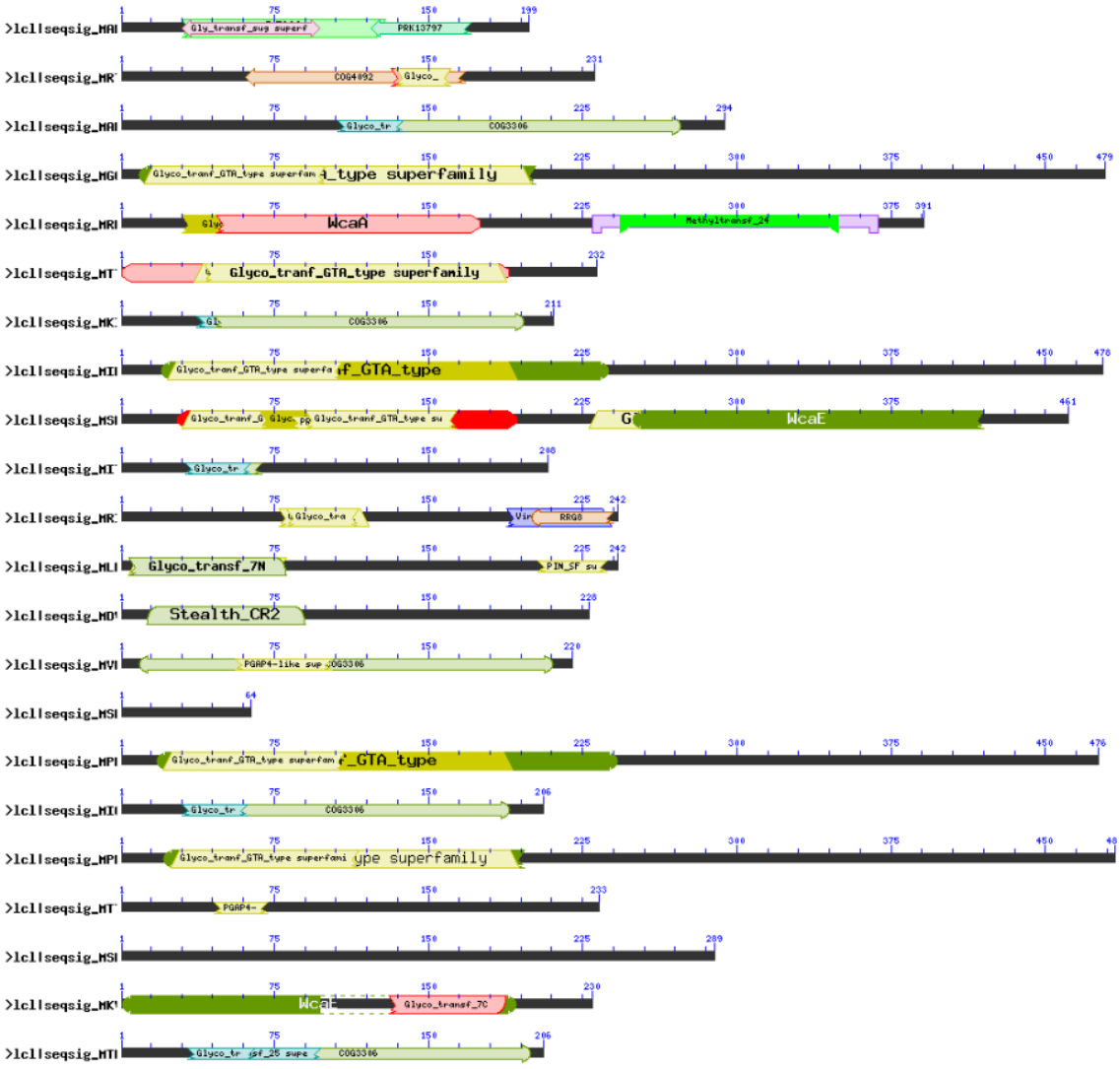


Figure 2-1 Batch conserve domain search result.

From up to down: Phrappuccino_gp142, Phrappuccino_gp143, Phrappuccino_gp154, Myrna_gp234, Che8_gp108, Che8_gp109, Che8_gp110, Fionnbharth_gp6, Sbash_gp87, Sbash_gp88, Omega_gp16, Omega_gp23, Marvin_gp81, Marvin_gp83, Nebkiss_gp116, MooMoo_gp94, MooMoo_gp96, Sparky_gp88, Sparky_gp89, Corndog_gp36, Corndog_gp37, Corndog_gp38.

doi:10.6342/NTU202402746

The MMseqs2 sequence map function in Phamerator was instrumental in mapping these genes. Some GTs were found to share moderate degrees of sequence identities despite belonging to different clusters of mycobacteriophages, Corndog gp37 and gp38 are distant homologs of Che8 gp109 and gp110 with 56% and 47% identity (Figure 2-2). Amino acid sequences of the identified genes were retrieved from PhagesDB and reverse translated into nucleotide sequences using the EMBOSS Backtranseq tool. Comparative genomics via NCBI's BLASTn and BLASTp tools provided further validation and functional annotation of these genes. Interestingly, one of the GT proteins (Che8_gp108) was identified with a N-terminal glycosyltransferase GT-A type superfamily domain and a C-terminal methyltransferase family 24 domain (Figure 2-3) which could imply this enzyme might function by the N-terminal domain facilitating the glycosylation and the C-terminal domain adding methyl groups.

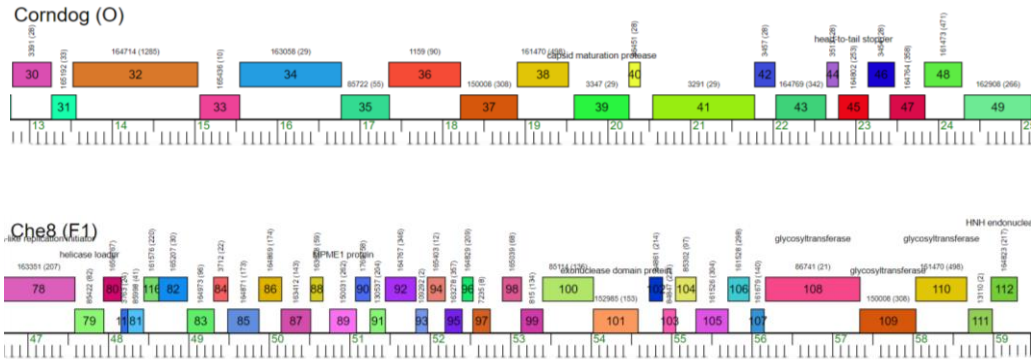


Figure 2-2 Genome maps of mycobacteriophages from different clusters, Corndog (O) and Che8 (F1). Genes are shown as colored boxes same color represents high identity. Genome ruler and putative functions are indicated.

doi:10.6342/NTU202402746

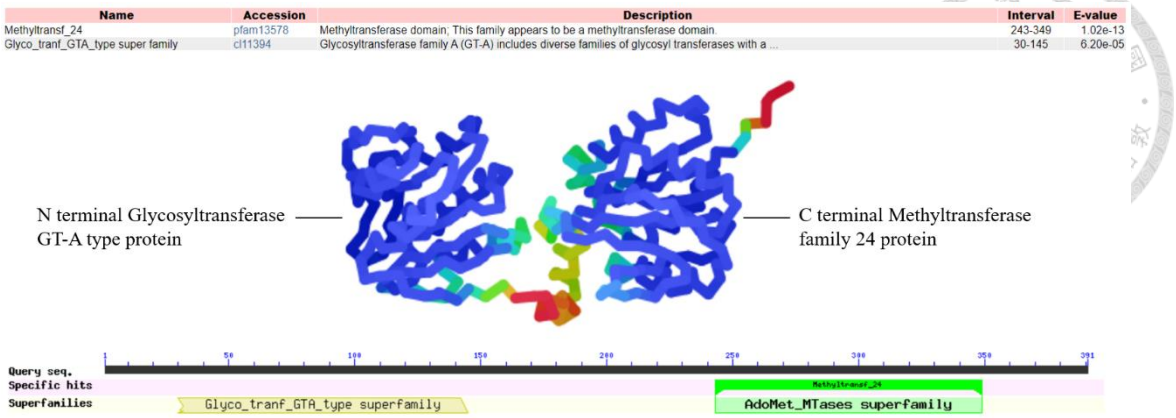


Figure 2-3 Che8_gp108 AlphaFold2 model and conserved domain functional prediction. N-terminal glycosyltransferase GT-A type superfamily domain (30–145) and C terminal methyltransferase family 24 domain (243–349).

Moreover, transmembrane domain predictions generated using the DeepTMHMM deep learning protein language model tool indicated the presence of complex topologies within these GT proteins, none of which had transmembrane domains indicating that these glycosyltransferases are likely either members of the GT-A or GT-B family. This was confirmed further with the help of the protein structure analytical tool AlphaFold2. The structural prediction via AlphaFold2 provided high-confidence three-dimensional models, facilitating a deeper understanding of the structural motifs and indicating potential interactions within these proteins. From the folding prediction models, all of the predicted proteins had one or two Rossman-like folds but none of them was found to fold with transmembrane alpha helix domain (Figure 2-4).

doi:10.6342/NTU202402746

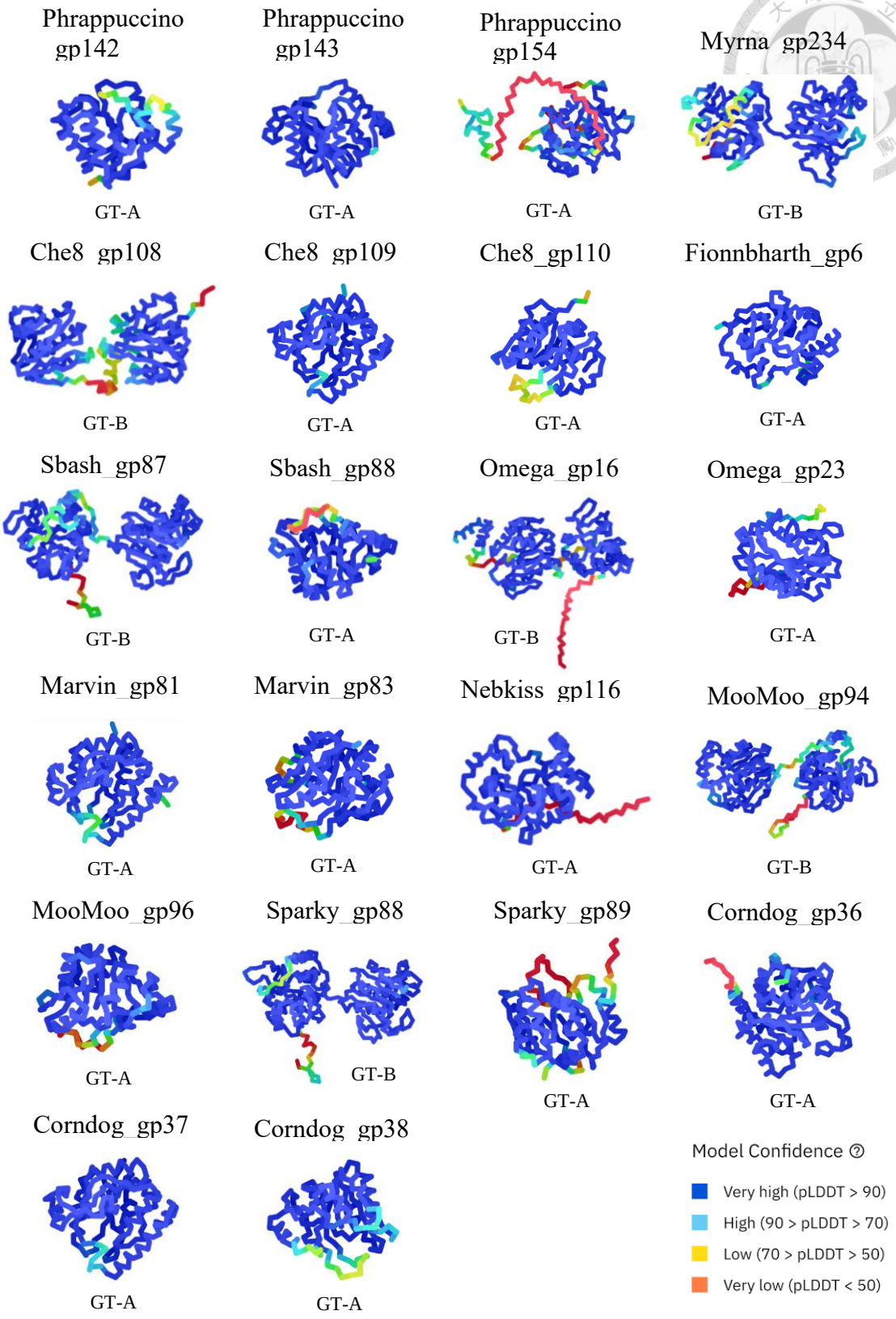


Figure 2-4 Mycobacteriophage GT structure prediction.

doi:10.6342/NTU202402746

This result is consistent with the DeepTMHMM transmembrane domain prediction results, which indicate that these proteins are most likely GT-A or GT-B type GT. The models of the proteins indicating a GT family domain were handpicked and compared to the solved protein structure models of various GTs in the CAZy database. The structure of Corndog_gp37 superimposed with the crystal structure of the catalytic domain of bovine beta1,4-galactosyltransferase (PDB entry: 1O0R) had one of the lowest Root Mean Squared Deviations (RSMD) out of the group of proteins tested with a value of 2.943 Å (Figure 2-5 (A)). Moreover, the conserved GT-A 'DxD' motif of 1O0R aligned well with the 'ESD' motif of the Corndog_gp37 (Figure 2-5 (B)), and the 'DxD' motif of 1O0R alongside the 'ESD' motif of Corndog_gp37 complexed well with a UDP-galactose and manganese that were included in the PDB structure of 1O0R (Figure 2-5 (C)). These findings suggest that certain mycobacteriophages are highly probable to encode GTs, more specifically GT-A and GT-B type enzymes, which use nucleotide sugar as sugar donors.

doi:10.6342/NTU202402746

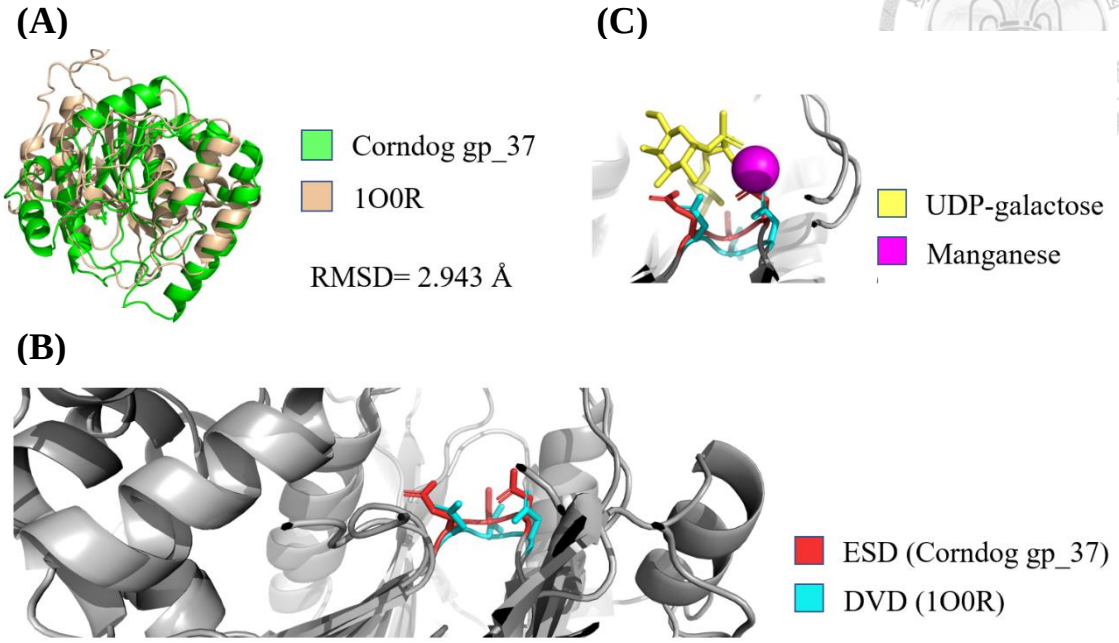


Figure 2-5 Corndog_gp37 superimposing with a crystal structure of the catalytic domain of bovine beta1,4-galactosyltransferase (PDB entry: 1O0R). (A) Protein structure superimposition of Corndog_gp37 (green) and 1O0R (beige) with an RMSD of 2.943 Å. (B) The alignment of 'DVD' motif of 1O0R (cyan) and 'ESD' motif of Corndog_gp37 (red). (C) 'DVD' motif of 1O0R (cyan) and 'ESD' motif of Corndog_gp37 (red) complexed with a UDP-galactose (yellow) and manganese (magenta).

doi:10.6342/NTU202402746

2.2 Cloning of mycobacteriophage proteins

Che8_gp109, Che8_gp110, Myrna_gp234, Myrna_gp238, Corndog_gp36, Corndog_gp37, and Corndog_gp38 were chosen to be cloned and expressed based on previous research by Hatfull and co-workers where mass spectrometry results revealed that the mycobacteriophages Che8, Myrna and Corndog are O-glycosylated [35]. This work also helped in designing the peptide acceptor for the above-mentioned potential GTs. Corndog_gp49 (major tail protein) was chosen as a potential acceptor for mycobacteriophage Corndog surface O-glycosylation. The cloning of Che8_gp109, Che8_gp110, Myrna_gp234, Myrna_gp238, Corndog_gp36, Corndog_gp37, Corndog_gp38, and Corndog_gp49 were achieved through careful experimental procedures. Custom-synthesized cDNA with a 22ACBVPP-pMA-RQ plasmid backbone were transformed into *E. coli* DH5α cells and incubated at 37 °C to amplify the plasmids. The plasmids were double digested with either restriction enzyme BamH1 and Sal1 or BamH1 and Not1 to cleave off the gene inserts. The bands of the digested DNA insert fragments were then collected according to the expected size of Che8_gp109 (708 bp), Che8_gp110 (645 bp), Myrna_gp234 (1449 bp), Myrna_gp238 (708 bp), Corndog_gp36 (879 bp), Corndog_gp37 (702 bp), Corndog_gp38 (630 bp), and Corndog_gp49 (830 bp) that corresponds to the bands in the agarose gel (Figure 2-6). The gene inserts were purified and ligated into pET28a(+) vectors using T4 DNA ligase, followed by transformation into *E. coli* BL21(DE3) cells. With this plasmid construct, the final protein products produced N-terminal His_6-tagged proteins. The DNA sequencing results confirmed that there were no frameshift or mutations observed.

doi:10.6342/NTU202402746

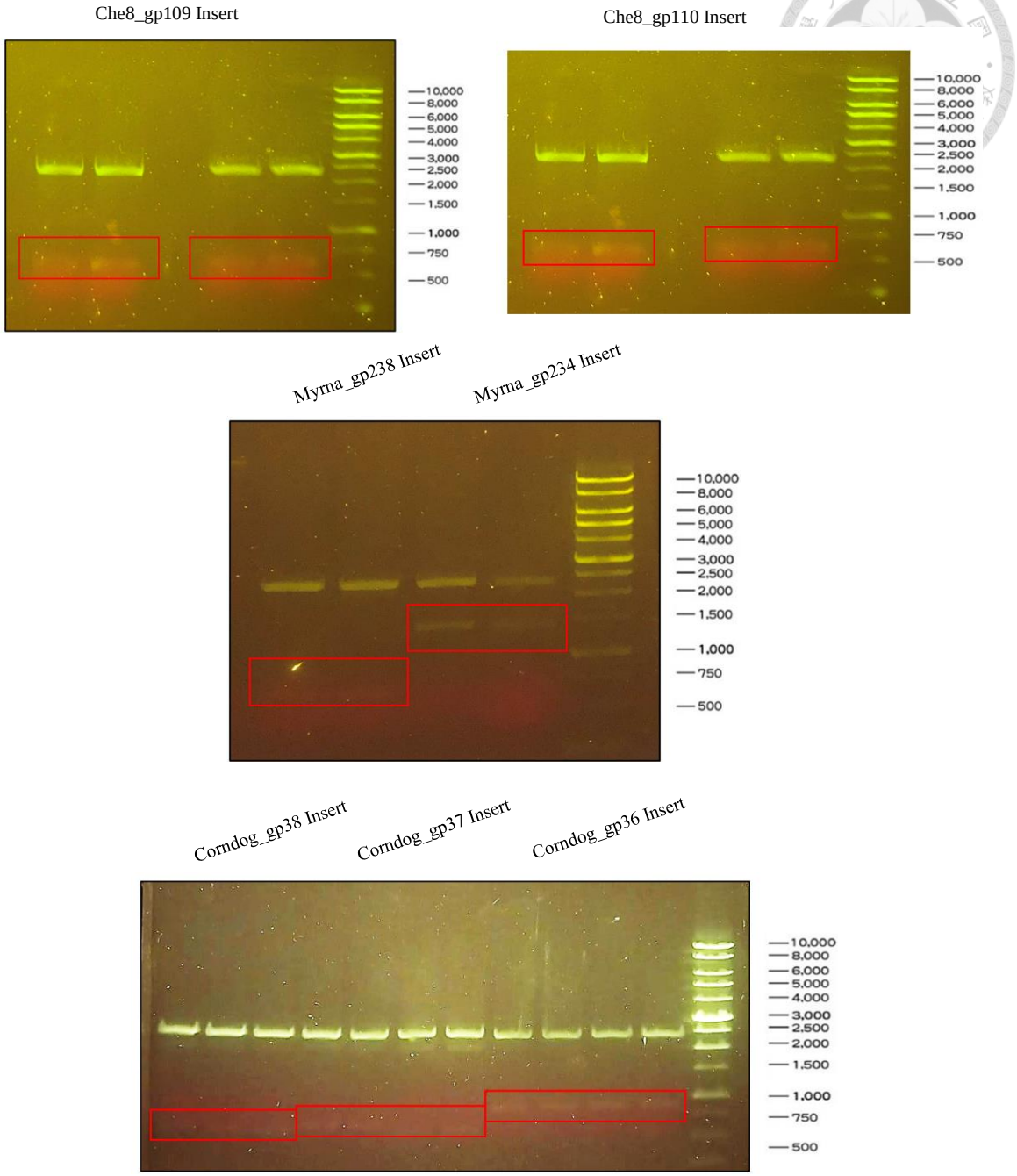


Figure 2-6 Agarose gel electrophoresis of Che8_gp109, Che8_gp110, Myrna_gp234, Myrna_gp238, Corndog_gp36, Corndog_gp37, and Corndog_gp38 inserts. Inserts are shown in red square.

doi:10.6342/NTU202402746

Another approach was used in the cloning of Corndog_gp36, Corndog_gp37, and Corndog_gp38 for more stable expression and higher yield. The synthesized 22ACBVPP-pMA-RQ plasmid was used as the template and various primers (Table 4-2) were used to perform polymerization chain reactions (PCR) amplification. The PCR products were detected by agarose gel electrophoresis (Figure 2-7).

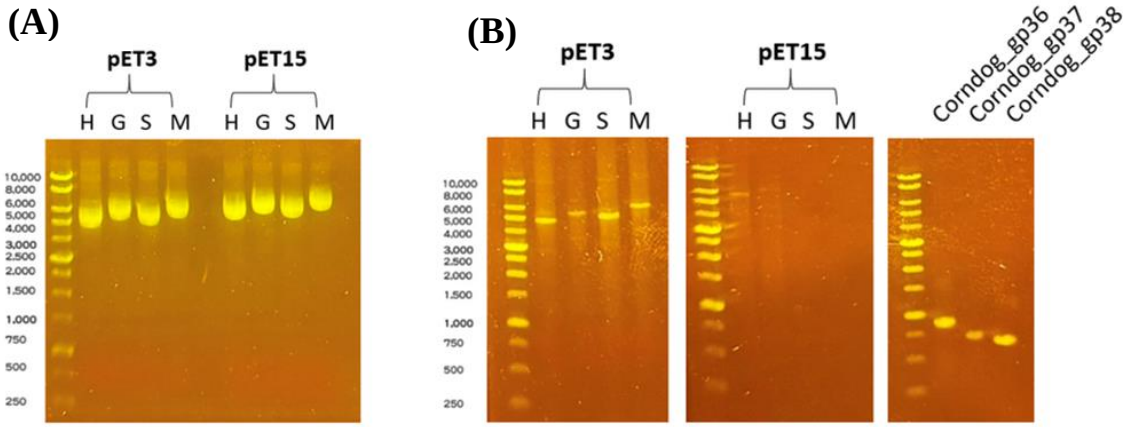


Figure 2-7 PCR amplification of pET3-His, pET3-GST, pET3-SUMO, pET3-MBP, pET15-His, pET15-GST, pET15-SUMO, pET15-MBP, Corndog_gp36, Corndog_gp37, and Corndog_gp38. (A) Positive standard for pET3 and pET15 vectors. (B) The PCR amplification products.

These inserts were then annealed with pET3-His, pET3-GST, pET3-SUMO, pET3-MBP, pET15-His, pET15-GST, pET15-SUMO, and pET15-MBP vectors. After transforming into *E. coli* BL21(DE3) cells, an ampicillin selectivity test was performed (Figure 2-8). pET3-His plasmids were selected for all three genes due to observed growth in the selectivity test, sufficient expression in the small-scale expression test (Figure 2-9), and the ease of purification with His_6-tag.

doi:10.6342/NTU202402746

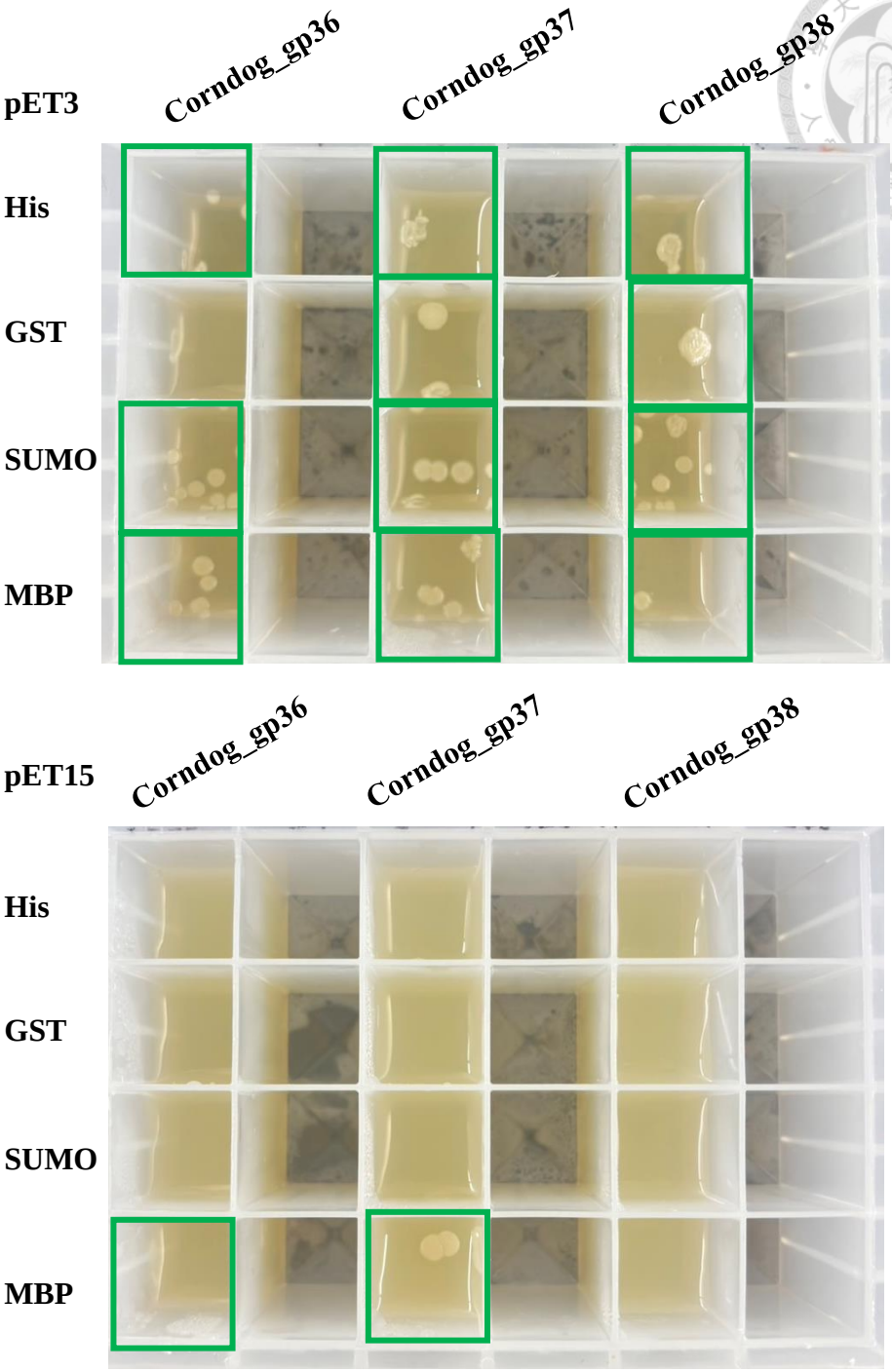


Figure 2-8 Ampicillin selective media test of pET3 and pET15 Corndog plasmids. Green square indicates colony growth.

doi:10.6342/NTU202402746

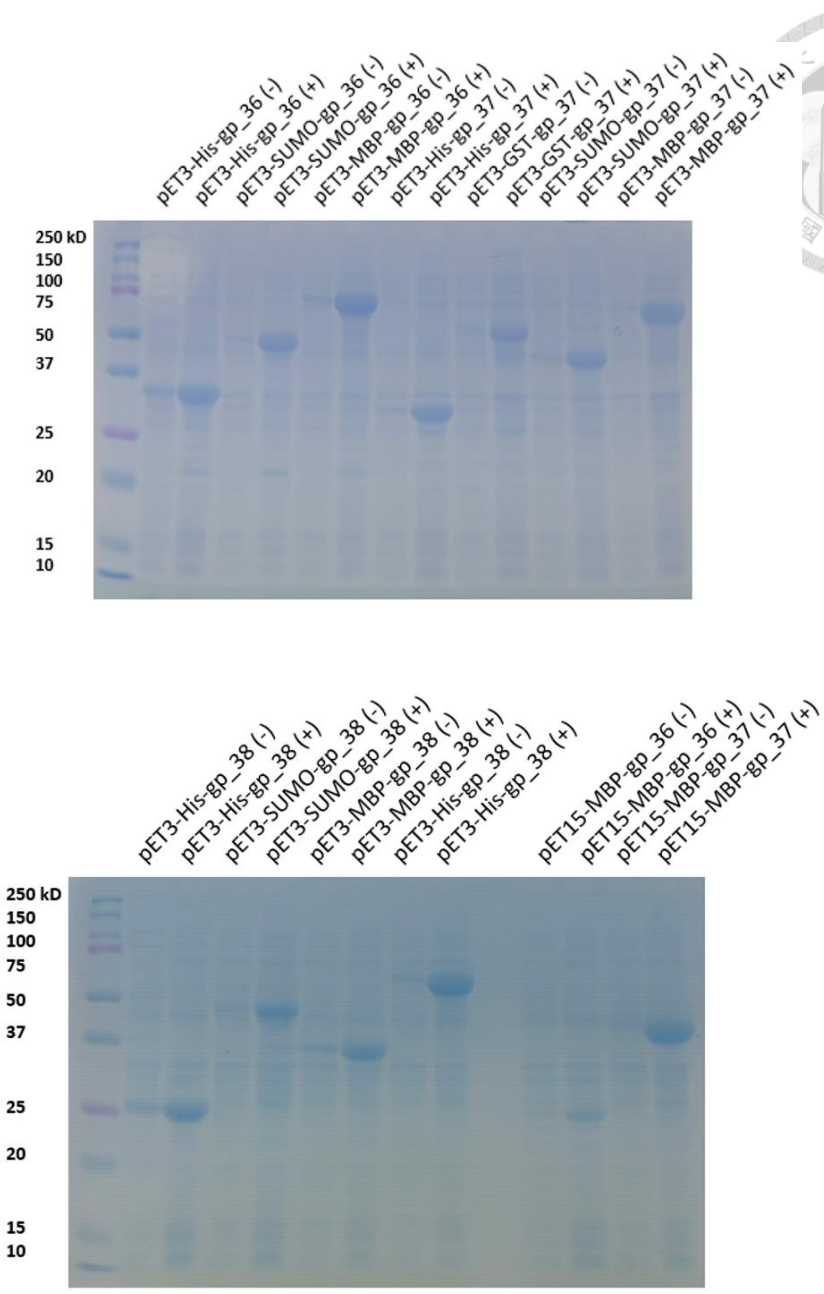


Figure 2-9 Small-scale expression tests of pET3 and pET15 fusion Corndog proteins. Lane 1 of both SDS-PAGE are protein ladder, labels of the other lanes are shown in the figure.

doi:10.6342/NTU202402746

2.3 Overexpression of mycobacteriophage proteins

The overexpression of all cloned mycobacteriophage proteins was induced using IPTG in *E. coli* BL21(DE3) cells. Recombinant strains harboring the respective plasmids were cultured in LB broth medium, and protein expression was induced at an OD_{600} of approximately 0.6. Induction done with 0.5-1.0 mM IPTG at 16 °C overnight facilitated the optimization of protein expression (Table 2-2). The harvested cells were resuspended in lysis buffer and subjected to high-pressure homogenization (18,000 psi) for cell disruption. Subsequent centrifugation and IMAC purification were then performed. SDS-PAGE results indicated that His-Che8_gp109 yielded highly pure His-tagged proteins. Relatively low yield was found for His-Corndog_gp37 and His-Corndog_gp38, and poor solubility for His-Che8_gp110, His-Myrna_gp234, and His-Myrna_gp238 proteins as they were found in the insoluble pellet fraction (Figure 2-10).

Table 2-2 Summary of the expression condition, yield, vectors used, and purification systems used for expressed proteins in this study.

Enzyme	kDa	IPTG	mg/ L culture	Vectors	Purification Systems
Che8_gp109	26.6	16 °C, 0.5 mM	35.2	pET28a	IMAC
Che8_gp110	24.6	16 °C, 0.2 mM	_	pET28a	IMAC
Corndog_gp36	33.9	16 °C, 0.2 mM	16.2	pET3	IMAC
Corndog_gp37	26.3	16 °C, 0.5 mM	11.75	pET3	IMAC
Corndog_gp38	23.3	20 °C, 0.5 mM	3.59	pET3	IMAC
Myrna_gp234	56.6	20 °C, 0.5 mM	_	pET28a	IMAC
Myrna_gp238	27.5	16 °C, 0.5 mM	_	pET28a	IMAC

doi:10.6342/NTU202402746

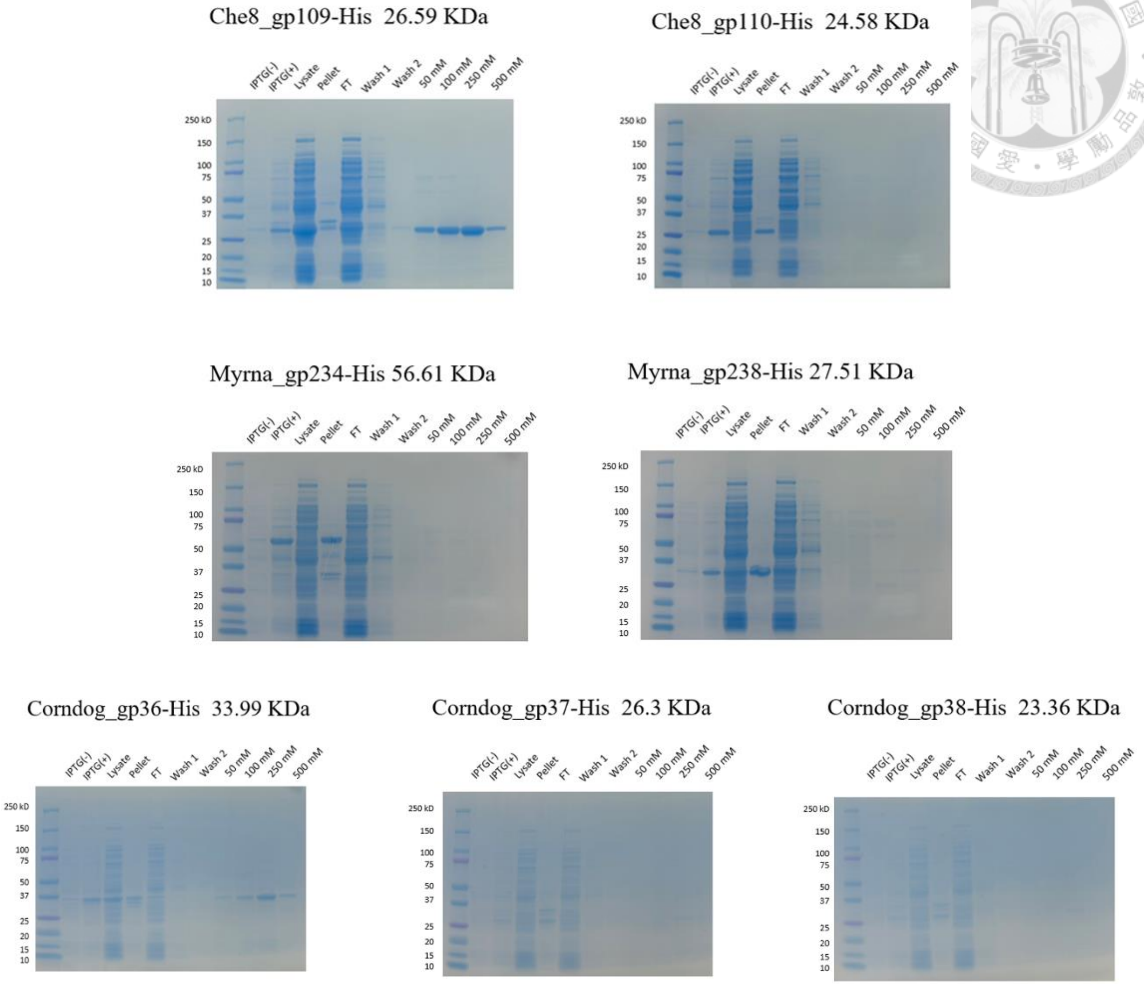


Figure 2-10 Overexpresssion and IMAC purification of Che8_gp109, Che8_gp110, Myrna_gp234, Myrna_gp238, Corndog_gp36, Corndog_gp37, and Corndog_gp38 (12% SDS-PAGE gel). Lane 1 is the protein ladder, lane 2 is the cell lysate without IPTG induction, lane 3 is the cell lysate with IPTG induction, lane 4 is the supernatant fraction of the cell lysate, lane 5 is the insoluble pellet of the cell lysate, lane 6 is the Flow Through (FT), the fraction after the initial flow through of the column, lane 7 and 8 are the wash 1 and wash 2 fractions (10 mM imidazole), lane 9 is the 50 mM imidazole elution fraction, lane 10 is the 100 mM imidazole elution fraction, lane 11 is the 250 mM imidazole elution fraction, lane 12 is the 500 mM imidazole elution fraction.

doi:10.6342/NTU202402746

A Kyte & Doolittle hydrophobicity plot suggested that these proteins have poor water solubility due to having regions being highly hydrophobic (Figure 2-11).

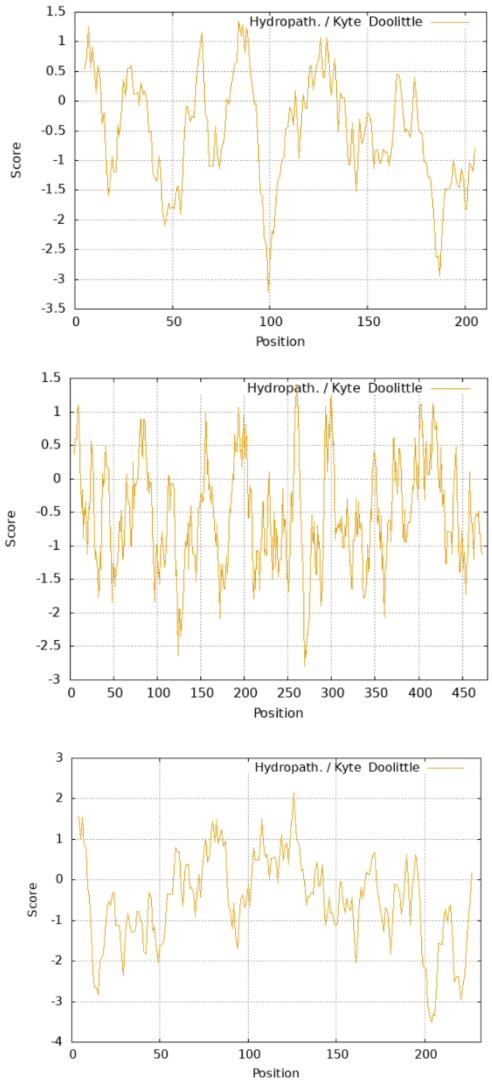


Figure 2-11 Kyte & Doolittle hydrophobicity plot of His-Che8_gp110, His-Myrna_gp234, and His-Myrna_gp238 (from top to bottom). X-axis indicates amino acid residue, Y-axis indicates hydrophobicity score: Ala: 1.800, Arg: –4.500, Asn: –3.500, Asp: –3.500, Cys: 2.500, Gln: –3.500, Glu: –3.500, Gly: –0.400, His: –3.200, Ile: 4.500, Leu: 3.800, Lys: –3.900, Met: 1.900, Phe: 2.800, Pro: –1.600, Ser: –0.800, Thr: –0.700, Trp: –0.900, Tyr: –1.300, Val: 4.200.

doi:10.6342/NTU202402746

An attempt to increase the solubility of these proteins by incubating the pellet fractions with different surfactants (Triton X 1%, dodecyl maltoside 1%, lauryldimethylamine oxide 1%, octadecyl glucoside 1%, and lauryl maltose neopentyl glycol 0.5%) were conducted but the effect was minuscule (Figure 2-12).

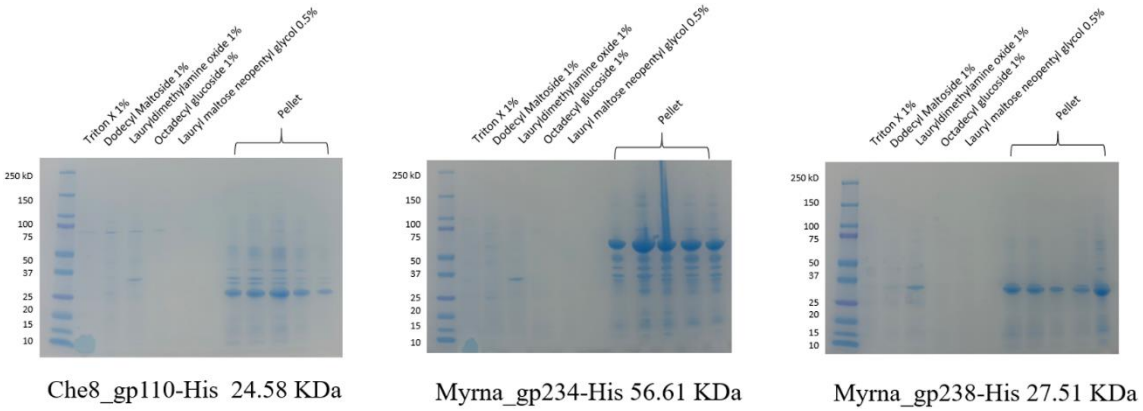


Figure 2-12 SDS-PAGE result of surfactant incubation of His-Che8_gp110, His-Myrna_gp234, and His-Myrna_gp238. Lane 1 is the protein ladder, lane 2 to lane 6 are the incubation mixture with different surfactants (Triton X 1%, dodecyl maltoside 1%, lauryldimethylamine oxide 1%, octadecyl glucoside 1%, and lauryl maltose neopentyl glycol 0.5%), lane 8 to lane 12 are the insoluble pellet fraction after surfactant incubation.

To move forward with a functional enzymatic assay, a first completed set of potential GTs from a single mycobacteriophage was tested. This made Corndog the mycobacteriophage of interest for the experiments as it was the only one with a full set of potential glycosyltransferases expressed without any significant problem. However, to address the low protein yield issue, another approach of cloning was used to generate a variety of fusion-tag proteins, namely, His-tag, GST-tag, SUMO-tag, and MBP-tag proteins using a diverse range of vectors available in our laboratory. A library of four different tagged Corndog_gp36, Corndog_gp37, and Corndog_gp38 pET3 and pET15

doi:10.6342/NTU202402746

plasmids were created and selected in a 6 × 4 selective media mentioned in the previous section (Section 2.2). SDS-PAGE of the small-scale expression revealed good expression level in all three pET3-His-Corndog proteins, pET3-GST-Corndog_gp37 and gp38, all three pET3-SUMO-Corndog proteins, all three pET3-MBP-Corndog proteins, and pET15-MBP-Corndog_gp37. However, pET15-MBP-Corndog_gp38 did not produce high yield of the fusion protein (Figure 2-10). It was decided to use pET3-His plasmids for all three Corndog proteins due to observed growth in the selectivity test, sufficient expression in the small-scale expression test (Figure 2-9), and the ease of purification with His_6-tag. With the new pET3-His plasmids, SDS-PAGE results showed high expression for all three Corndog proteins (Figure 2-13), setting the stage for the functional assays.

doi:10.6342/NTU202402746

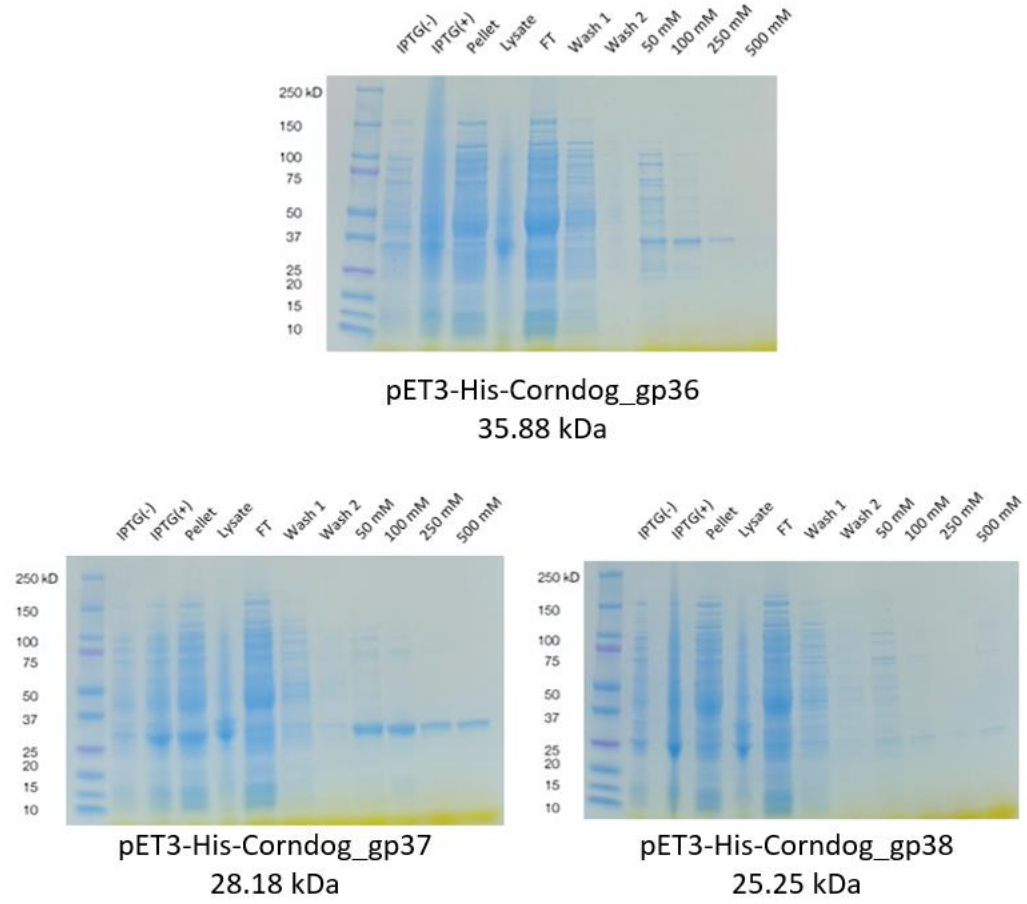
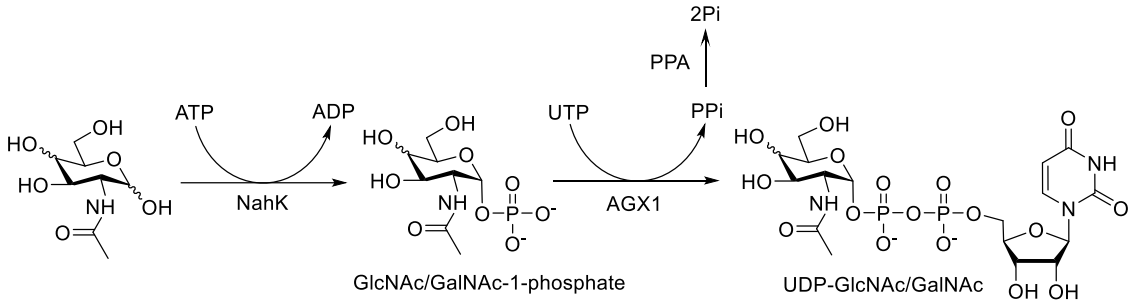


Figure 2-13 Overexpresssion and IMAC purification of Corndog_gp36, Corndog_gp37, and Corndog_gp38 (12% SDS-PAGE gel). Lane 1 is the protein ladder, lane 2 is the cell lysate without IPTG induction, lane 3 is the cell lysate with IPTG induction, lane 4 is the supernatant fraction of the cell lysate, lane 5 is the insoluble pellet of the cell lysate, lane 6 is the Flow Through (FT), the fraction after the initial flow through of the column, lane 7 and 8 are the wash 1 and wash 2 fractions (10 mM imidazole), lane 9 is the 50 mM imidazole elution fraction, lane 10 is the 100 mM imidazole elution fraction, lane 11 is the 250 mM imidazole elution fraction, lane 12 is the 500 mM imidazole elution fraction.

doi:10.6342/NTU202402746

2.4 Enzymatic synthesis of nucleotide sugars

The enzymatic syntheses of nucleotide sugars, specifically UDP-GlcNAc and UDP-GalNAc, were performed using the NahK and AGX1 enzymes (Scheme 2-1). The sugar nucleotides were chosen as the sugar donor of the glycosylation reaction due to the structural prediction results of Corndog_gp36, gp_37, and gp_38 suggesting all three of them being GT-A type glycosyltransferase which uses nucleotide sugar as sugar donors.



Scheme 2-1 Enzymatic synthesis of UDP-GlcNAc/GalNAc using NahK and AGX1

The reaction conditions, including the use of HEPES buffer, magnesium chloride, adenosine triphosphate, and the respective sugar substrates, were incubated as a reaction mixture at 37 °C overnight. The syntheses involved sequential addition of GlcNAc/GalNAc, adenosine triphosphate, magnesium chloride, uridine triphosphate, and the enzymes to ensure the formation of the UDP-GlcNAc/GalNAc. Thin layer chromatography (TLC) monitored the progression and completion of the reactions (Figure 2-14).

doi:10.6342/NTU202402746

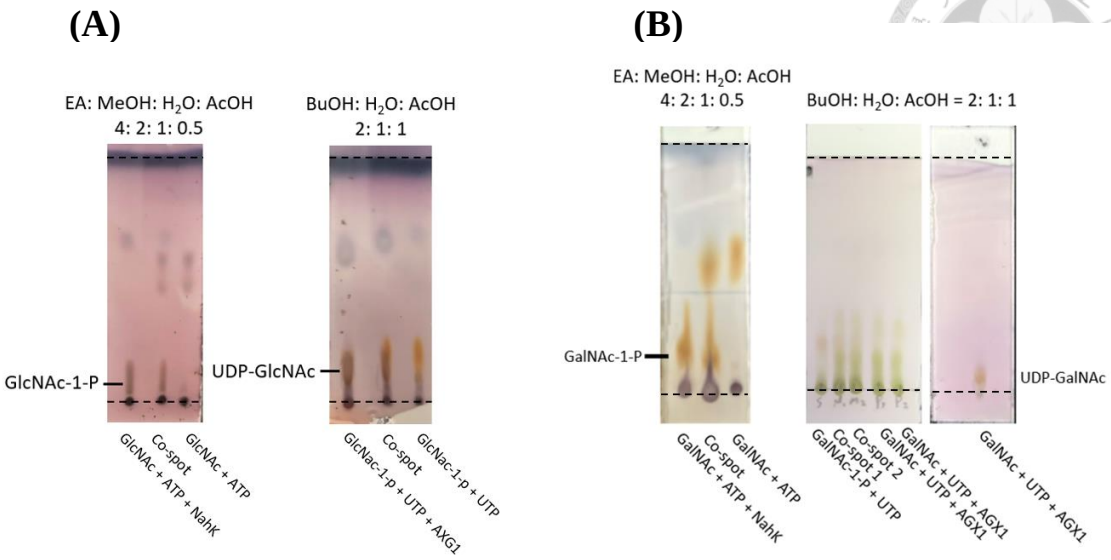


Figure 2-14 TLC results of UDP-GlcNAc and UDP-GalNAc enzymatic syntheses. (A) UDP-GlcNAc (B) UDP-GalNAc.

The synthesized products were purified by barium chloride precipitation and size-exclusion chromatography. The purity and identity of the products were confirmed by nuclear magnetic resonance (NMR) spectroscopy. According to the ^{1}H NMR spectra of the synthesized UDP-GlcNAc and UDP-GalNAc the chemical shifts of the various protons agreed with the values from previous reports (Figure 2-15, 2-16) [49]. Overall, the syntheses of UDP-GlcNAc and UDP-GalNAc resulted in nucleotide sugars with high purity suitable for the enzymatic assays.

doi:10.6342/NTU202402746

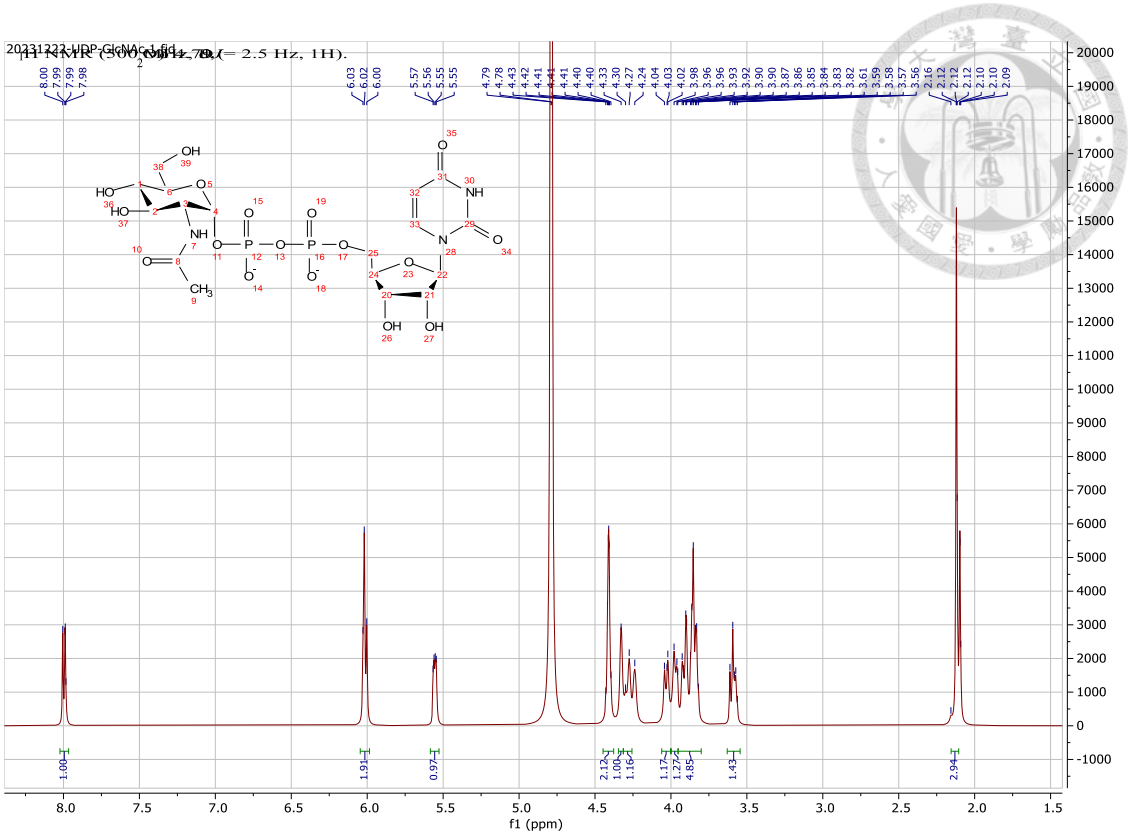


Figure 2-15 500 MHz ^{1}H NMR spectrum of UDP-GlcNAc, measured in D_2O

doi:10.6342/NTU202402746

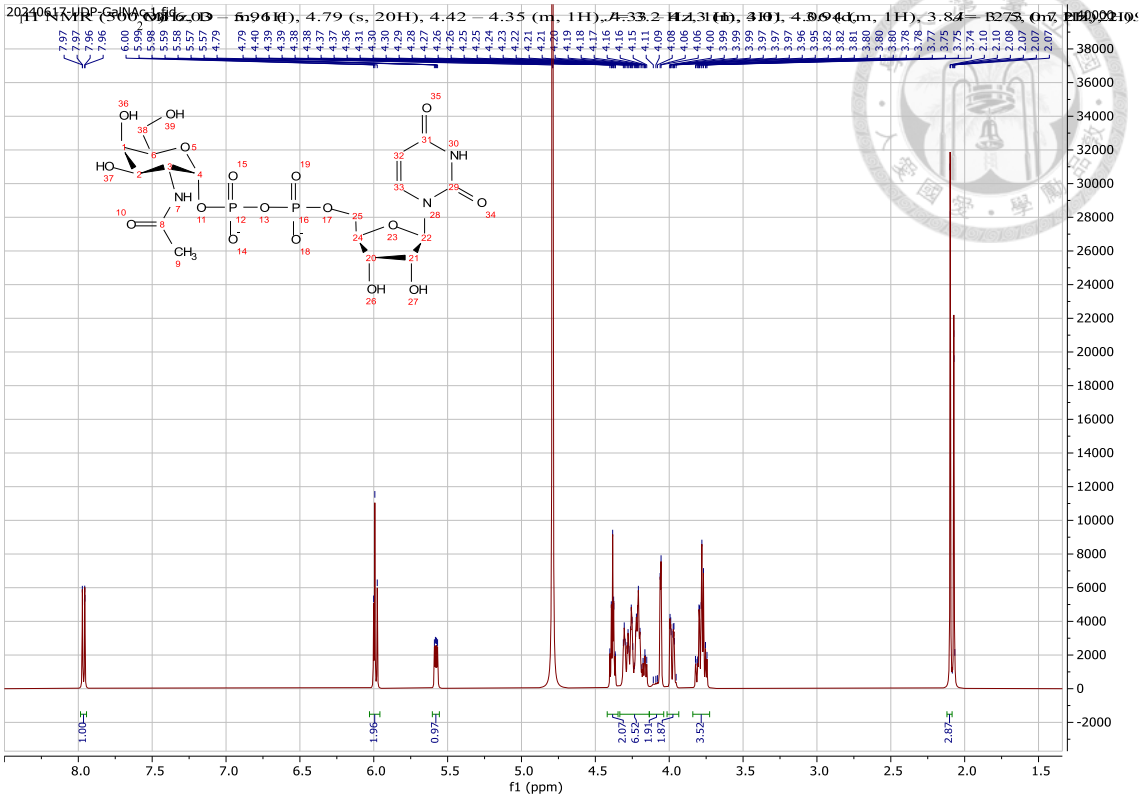


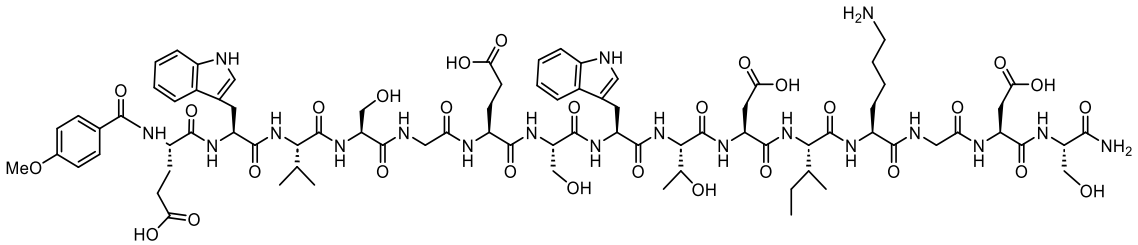
Figure 2-16 500 MHz ^{1}H NMR spectrum of UDP-GalNAc, measured in D_2O

doi:10.6342/NTU202402746

2.5 Enzymatic assay test of Corndog_gp36, Corndog_gp37, and Corndog_gp38

The activities of the putative mycobacteriophage GTs Corndog_gp36, Corndog_gp37, and Corndog_gp38 were evaluated using the synthesized nucleotide sugars (UDP-GlcNAc and UDP-GalNAc), commercially purchased nucleotide sugars (UDP-glucose, UDP-galactose, and GDP-mannose) and peptide acceptors (pMBz-EWVSGESWTDIKGDS and GLAGGS(GlcNAc)G, where pMBz is a *p*-methoxybenzamide tag added to facilitate detection by UV light. The choice of these acceptors is described below.

The assays were conducted in 100 mM Tris-HCl buffer with divalent cations (20 mM $MgCl_2$ or $MnCl_2$) at 37 °C with agitation. The reaction products were analyzed by thin layer chromatography (TLC), and the presence of expected glycosylated products was confirmed by electrospray ionization mass spectrometry (ESI-MS) and Matrix-Assisted Laser Desorption Ionization mass spectrometry (MALDI-MS). Acceptor **1** (pMBz-EWVSGESWTDIKGDS) was the first peptide used as the acceptor in the enzymatic assay. This peptide was a truncated mycobacteriophage Corndog capsid protein and a C-terminal serine for O-glycosylation. This truncated peptide was used as the acceptor because the full capsid proteins of phages are highly unstable to produce, as capsid proteins have the tendency to self-assemble and self-aggregate.

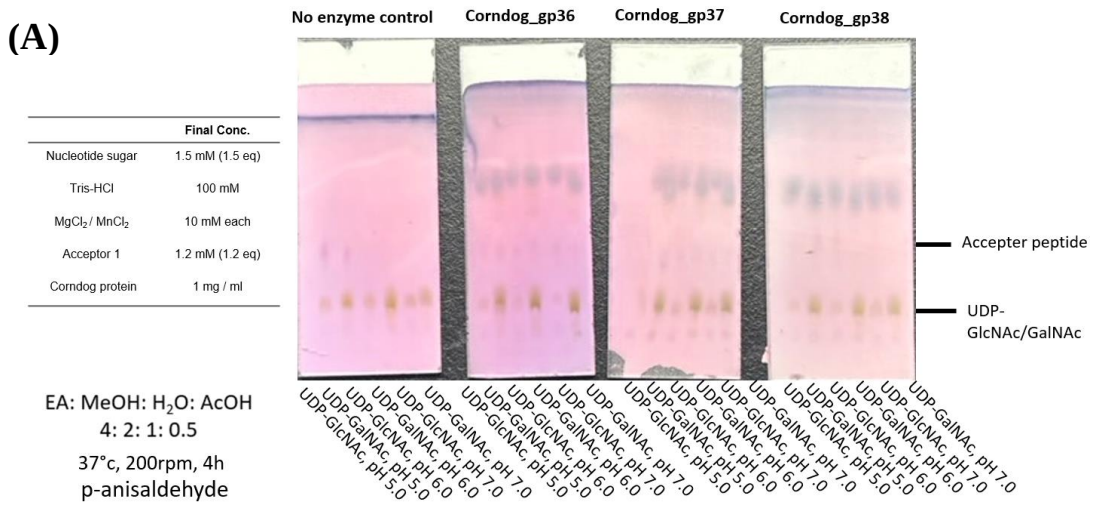


pMBz- EWVSGESWTDIKGDS Mw: 1829.91 g/ mol

Scheme 2-2 Structure of acceptor **1**

doi:10.6342/NTU202402746

Several attempts to develop a functional assay with Corndog_gp36, Corndog_gp37, and Corndog_gp38 were done with acceptor **1** (Figure 2-17). First, the acceptor was tested under different pH ranging from 5.0 to 7.0, with a total reaction volume of 2 μL containing acceptor **1** (1 eq.), and UDP-GlcNAc/GalNAc (1.5 eq.) (Figure 2-17 (A)). Under these conditions, all three enzymes showed no new spots on the TLC when compared to the no enzyme negative control, suggesting an unsuccessful glycosylation reaction. Next, a reaction with 5 eq. of nucleotide sugar donor was conducted (total volume of 10 μL, pH 7.0), providing the reactions with access to more nucleotide sugar donor (Figure 2-17 (B)). Similar to the first attempt, no reaction was observed. Finally, an attempt using the crude expression *E. coli* cell lysate to promote glycosylation reaction instead of the purified and expressed Corndog_gp36, gp_37, and gp_38 was conducted (Figure 2-17 (C)). Unfortunately, the crude lysate under the tested conditions was also unable to demonstrate the transfer of sugar residues to acceptor **1**. ESI-MS and MALDI experiments performed on all the reaction mixtures above also failed to indicate glycosylation reaction, that is, no new peaks corresponding to the molecular weights of the desired products were observed on the MS spectra (Corresponding MS spectra not shown).



doi:10.6342/NTU202402746

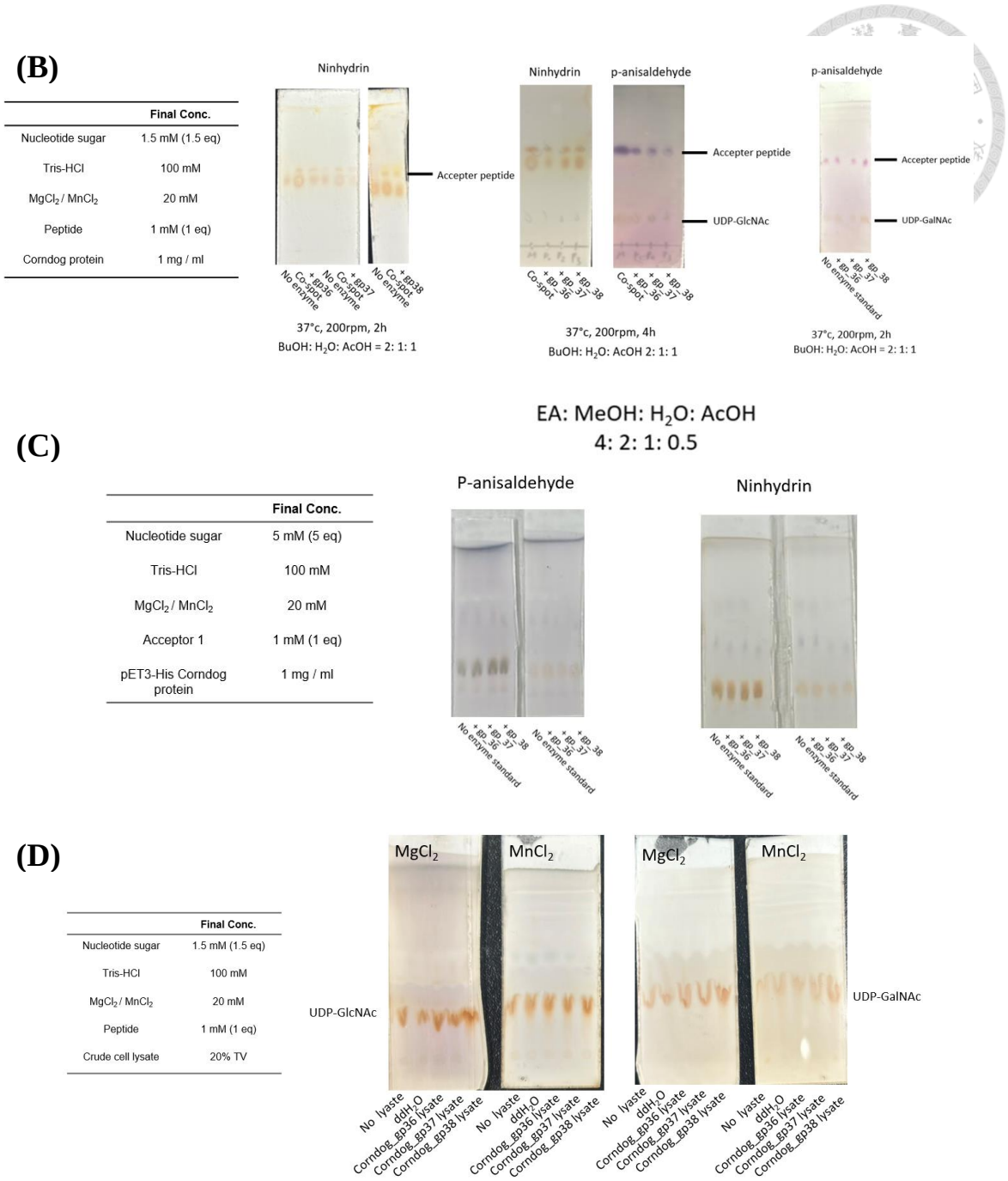
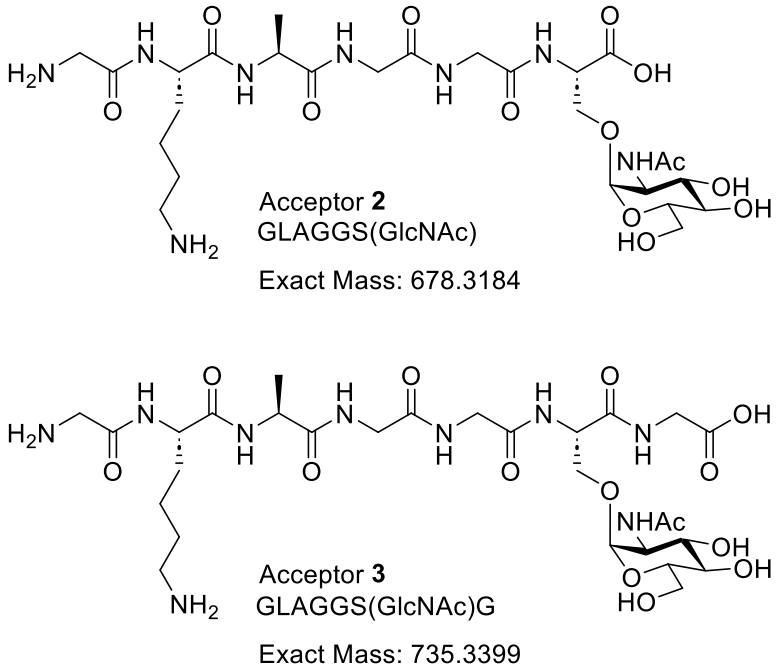


Figure 2-17 TLC results of functional assay with Corndog_gp36, Corndog_gp37, and Corndog_gp38 done with acceptor **1**. (A) 1.5 eq. nucleotide sugar donor, 100 mM Tris-HCl, 10 mM MgCl2, 10 mM MnCl2, 1.2 eq. UDP-GlcNAc/GalNAc, and 1 mg/mL Corndog proteins with pH ranging from 5.0 to 7.0. (B) 1.5 eq. nucleotide sugar donor, 100 mM Tris-HCl, 10 mM MgCl2, 10 mM MnCl2, 1 eq. acceptor **1**, and 1 mg/mL

doi:10.6342/NTU202402746

Corndog proteins (pH 7.0). (C) 5 eq. nucleotide sugar donor, 100 mM Tris-HCl, 10 mM MgCl2, 10 mM MnCl2, 1 eq. acceptor **1**, and 1 mg/mL Corndog proteins (pH 7.0). (D) 1.5 eq. nucleotide sugar donor, 100 mM Tris-HCl, 10 mM MgCl2, 10 mM MnCl2, 1 eq. acceptor **1**, and 20% total volume of crude expression *E. coli* cell lysate (pH 7.0).

The second attempt was with the synthetic acceptor **2** and acceptor **3** (GLAGGS(GlcNAc) and GLAGGS(GlcNAc)G, respectively), which were short flexible general glycosyltransferase substrate design with a GlcNAc residue already attached to the C-terminal serine to demonstrate the successive transfer of hexose sugars to a HexNAc already attached to proteins. An additional glycine was added to the C-terminus of one of the acceptors as this facilitated solid-phase synthesis.



Scheme 2-3 Structures of acceptor **2** and acceptor **3**

doi:10.6342/NTU202402746

The activity tests with acceptor **3** using Corndog_gp36, gp37 and gp38 were conducted in 10 μL volume containing acceptor **3** (1 eq.) and UDP-glucose, UDP-galactose, and GDP-mannose (1.5 eq.). The TLC showed no new spot with all three enzymes tested compared to the no enzyme negative control. Unfortunately, after testing acceptor **3** under different reaction conditions, it failed to demonstrate the function of Corndog_gp36, Corndog_gp37, and Corndog_gp38 as glycosyltransferases (Figure 2-18 (A)).

Lastly, an attempt using only GlcNAc and GalNAc (1 eq.) as the acceptor was conducted with UDP-glucose, UDP-galactose, and GDP-mannose (1.5 eq.) (Figure 2-18 (B)). This attempt was to test the possibility of UDP-HexNAc alone as the substrate of Corndog_gp36, gp37 and gp38. Similar to the previous reactions, the results of ESI-MS and MALDI experiments showed no new peaks corresponding to the molecular weights of the desired products confirming the unsuccessful transfer of sugars in the above reactions (Corresponding MS spectra not shown).

doi:10.6342/NTU202402746

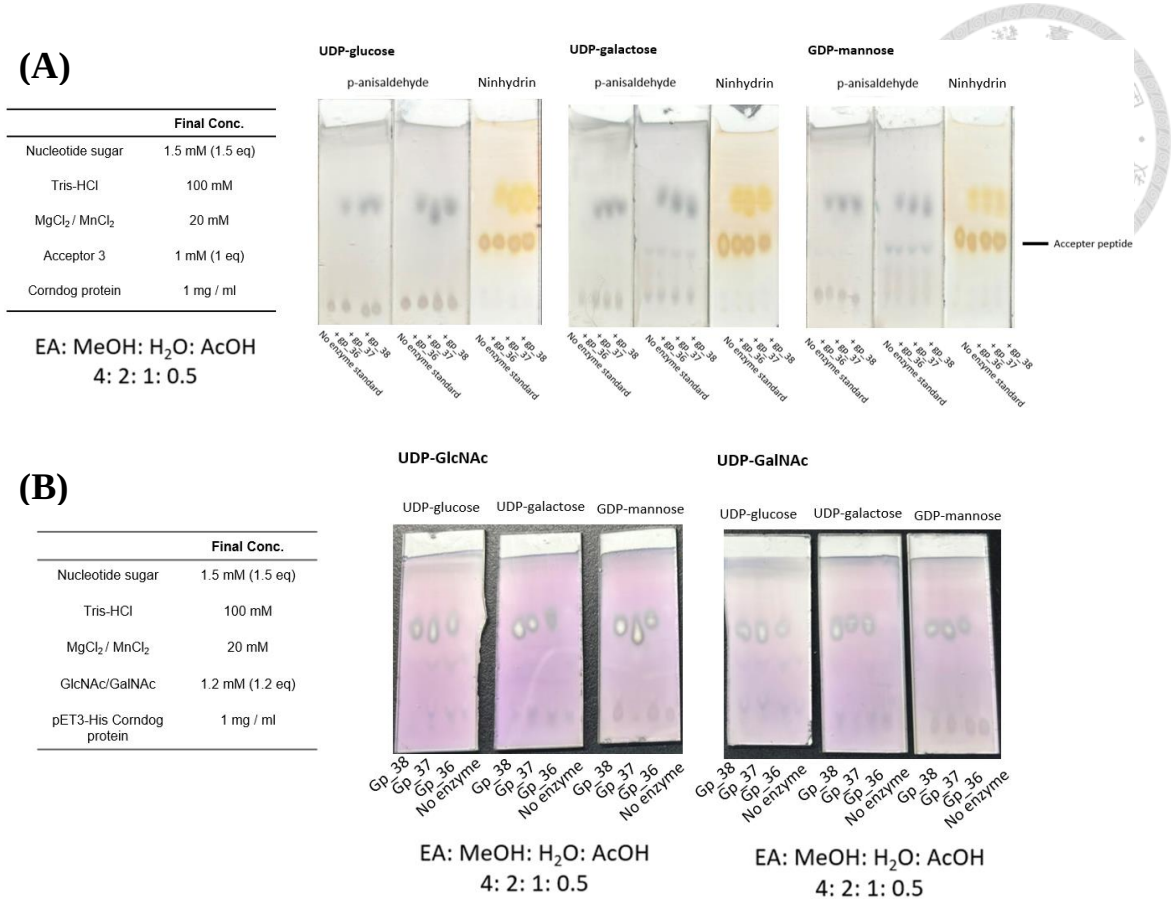


Figure 2-18 TLC results of functional assay with Corndog_gp36, Corndog_gp37, and Corndog_gp38 done with acceptor **3**. (A) 1.5 eq. nucleotide sugar donor, 100 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 1 eq. acceptor **3**, and 1 mg/mL Corndog proteins (pH 7.0). (B) 1.5 eq. nucleotide sugar donor, 100 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM $MnCl_2$, 1.2 eq. UDP-GlcNAc/GalNAc, and 1 mg/mL Corndog proteins (pH 7.0).

doi:10.6342/NTU202402746

The third attempt at developing a functional assay was with the major tail protein Corndog_gp49 as the acceptor. The idea behind this attempt was based on the reported occurrence of glycosylation appearing on the major and minor tail proteins of other mycobacteriophages [35], where mycobacteriophage Che8 was found to perform surface O-glycosylation on the capsid and the tail proteins. Corndog_gp49 having a C-terminal serine residue that is shown to protrude outside according to AlphaFold2 modeling (Figure 2-19) was then tested as a potential acceptor.

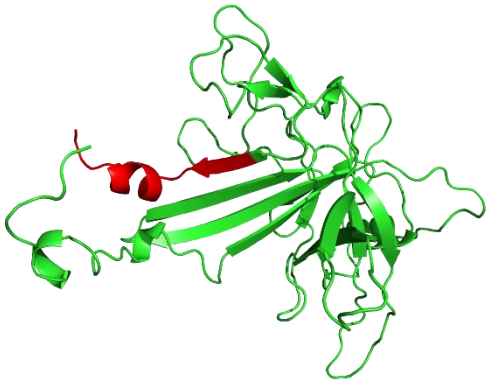


Figure 2-19 Corndog_gp49 AlphaFold2 model. C-terminal tail shown in red.

The enzymatic assay of Corndog_gp36, Corndog_gp37, and Corndog_gp38 using acceptor **2** and Corndog_gp49 are still under investigation and results of these experiments are expected in the near future.

doi:10.6342/NTU202402746

3. Conclusion

In this study, I performed an analysis of mycobacteriophage GT candidates, that has provided valuable insights into the genetic and structural characteristics of these enzymes. Utilizing advanced bioinformatics tools and experimental methodologies, I was able to identify and characterize potential GT genes from various mycobacteriophages, particularly focusing on mycobacteriophage Corndog.

Performing genome analysis revealed the presence of several GT-encoding genes within the mycobacteriophage genomes. By leveraging the PhagesDB database and the Phamerator tool, I identified multiple GT candidates and mapped their sequences using the MMseqs2 sequence map function. Notably, the identified genes exhibited conserved GT family motifs, which were confirmed through comparative genomics and structural predictions. Structural analysis using AlphaFold2 provided high-confidence three-dimensional models of these GTs, elucidating their structural motifs and potential functional sites. The absence of transmembrane domains in these proteins, as predicted by DeepTMHMM and confirmed by structural predictions, suggests that these GTs are likely to be soluble enzymes belonging to the GT-A or GT-B families. This finding aligns with the presence of Rossman-like folds, which are characteristic of these GT families. One notable discovery was the GT protein Che8_gp108, which contained both an N-terminal glycosyltransferase GT-A type superfamily domain and a C-terminal methyltransferase family 24 domain. This dual-domain structure suggests a potential complex function, possibly involving both glycosylation and methylation processes. Bioinformatics and structural prediction results provided new insights into the putative GTs involved in mycobacteriophage O-glycosylation.

doi:10.6342/NTU202402746

Despite some challenges with protein solubility, particularly for the Che8 and Myrna enzymes, the expression of Corndog GTs (gp36, gp37, and gp38) was achieved with relatively high efficiency. The low yield issues were mitigated by employing various fusion tags, with the pET3-His proteins showing high expression and purification profiles.

The successful synthesis of nucleotide sugars, specifically UDP-GlcNAc and UDP-GalNAc, was a critical step in preparing for the functional assays of the GTs. The optimized reaction conditions and purification processes ensured the production of high-purity nucleotide sugars, which were confirmed by ^{1}H NMR spectroscopy.

The functional assays of the Corndog GTs using synthesized nucleotide sugars and peptide acceptors, although initially promising, did not result in detectable glycosylation. Despite multiple attempts with different acceptor designs, the GTs were unable to transfer sugar residues to the acceptors. This outcome could result from several possibilities: the donor and acceptor specificities of these enzymes may be more complex than initially anticipated, or the *in vitro* conditions used in the assays failed to simulate or fully replicate the native environment required for enzymatic activity. The identification of specific donor–acceptor pairs remain a significant challenge and a potential bottleneck in demonstrating the precise functionalities of these novel enzymes.

To summarize, this study has provided an analysis of mycobacteriophage GTs, highlighting their genetic diversity, structural features, and potential functional roles. While challenges remain in demonstrating their enzymatic activities, the foundational knowledge gained from this research sets the stage for future investigations and

doi:10.6342/NTU202402746

applications, such as the possibility of using mycobacteriophage GTs to deploy surface glycan shield to mitigate innate immune response in phage therapy. The continued exploration of mycobacteriophage GTs holds promise for advancing our understanding of mycobacteriophage surface glycosylation processes and leveraging these enzymes for innovative solutions for phage therapy.

doi:10.6342/NTU202402746

4. Material and Methods

4.1 Mycobacteriophage genome analysis

Mycobacteriophage genome analysis was done by using several bioinformatics tools including the PhagesDB database (available at https://phagesdb.org) in conjunction with Phamerator (available at http://phamerator.org) and Many-against-Many sequence searching (MMseqs2). First, Phamerator was used to identify all the possible glycosyltransferase encoding genes in the complete Actinobacteriophage genome database (Actino_Draft version 561) using the MMseqs2 sequence map function. Then the amino acid sequence of potential glycosyltransferase genes was retrieved from the PhagesDB database. The amino acid sequence was reverse translated into nucleotide sequence using EMBOSS Backtranseq online tool (available at https://www.ebi.ac.uk/jdispatcher/st/emboss_backtranseq). Comparative genomics were conducted with National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLASTn and BLASTp) (available at https://blast.ncbi.nlm.nih.gov/Blast.cgi). The conserved domains of the glycosyltransferase families were extracted from NCBI's Conserved Domain Database (CDD) (available at https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) using Reverse Position-Specific BLAST (RPS-BLAST). Transmembrane domain predictions were generated using DeepTMHMM deep learning protein language model tool provided by the Phamerator webpage to identify and predict the topology of alpha helical and beta barrel proteins across all mycobacteriophage glycosyltransferase amino acid sequences. The sequences were then keyed in to perform protein structure and complex prediction using AlphaFold2.

doi:10.6342/NTU202402746

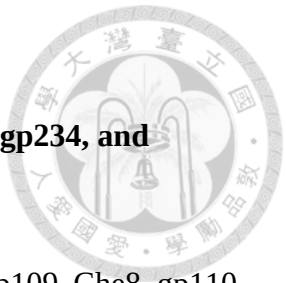
4.2 Mycobacteriophage protein structure folding prediction

The protein folding prediction was performed using AlphaFold2, accessed through the Google Colab servers provided by ChimeraX (May 2024 version). All prediction work were done with multiple runs and the model with the highest predicted Local Distance Difference Test (pLDDT) score was picked as the representative three-dimensional protein structure. The confidence measure for the predicted structures is scaled from 0 to 100 using the pLDDT, which quantifies the reliability of the model's predictions. This measure is made possible by performing B-factor calculations. Predicted structures were analyzed using PyMOL to visualize the three-dimensional conformations and identify structural motifs. PyMOL facilitated the detailed examination of secondary and tertiary structures, enabling the identification of critical interactions and functional sites. All three-dimensional protein structure images were generated at a resolution of 300 dpi, using the ray setting in PyMOL to ensure high-quality visual representation.

4.3 Mycobacteriophage Corndog glycosyltransferase function prediction

The glycosyltransferase functions and functional sites of Corndog phage proteins were predicted by NCBI's CDD using RPS-BLAST. Additionally, sequence homology searches were performed using BLASTp against the Protein Data Bank (PDB) and cross referenced with CAZy database (available at http://afmb.cnrs-mrs.fr/CAZY) to find similar structures with known functions. Functional prediction was further reinforced by the comparison of AlphaFold2 structural modeling with glycosyltransferase structures with known glycosyltransferase functions in the UniProt protein database (available at https://www.uniprot.org/).

doi:10.6342/NTU202402746



4.4 Cloning and overexpression

4.4.1 Cloning of Che8_gp108, Che8_gp109, Che8_gp110, Myrna_gp234, and Myrna_gp238

The cloning of mycobacteriophage gene Che8_gp108, Che8_gp109, Che8_gp110, Myrna_gp234, and Myrna_gp238 were performed by using custom-synthesized plasmid (Mission Biotech Ltd, Taiwan) (Figure 4-1) and by *E. coli* DH5α cell transformation via the heat shock method. 1 μL of the plasmid was added to 20 μL of the competent cells with efficiency of > 5 x 10^7 cfu/μg. The samples were incubated on ice for 45 min, heat shocked in a 42℃ water bath for one min and then transferred back on ice for additional 15 min incubation. For each transformation, 100 μL of *E. coli* DH5α cells were spread-plated at room temperature onto Luria-Bertani (LB) agar plates (containing 10.0 g/L tryptone, 5.0 g/L yeast extract, 5.0 g/L NaCl, and 15.0 g/L agar) supplemented with either 50 μg/mL kanamycin or 100 μg/mL ampicillin antibiotics depending on the antibiotic resistance of the plasmid. The antibiotics used were pre-filtered using sterile, non-pyrogenic Millex® Syringe Filters with a pore size of 0.20 μm (Merck, Darmstadt, Germany). The cultures were then incubated overnight at 37 ℃. The amplified plasmid samples were harvested by using the Presto™ Mini Plasmid Kit (Generaid Biotech Ltd, Taiwan) after overnight culture at 37 °C. Restriction enzyme BamH1 and Sal1 were used to double digest the gene inserts Che8_gp108, Che8_gp109, Che8_gp110, Myrna_gp234, and Myrna_gp238 from the custom plasmids by following the enzyme manufacturer's instruction. Gene inserts were then detected by agarose gel electrophoresis (1%), excised then further purified by GenepHlow™ Gel/PCR Kit (Generaid Biotech Ltd, Taiwan). The purified gene inserts were then ligated onto pET28a(+) vectors with T4 DNA Ligase (NEB, USA) and transformed into *E.coli* BL21(DE3) competent cells by the heat shock

doi:10.6342/NTU202402746

method. This was followed by overnight incubation at 37 °C. The whole LB agar plates were sent to Mission Biotech Ltd for DNA sequencing using T7 forward and T7t reverse sequencing primers (Mission Biotech Ltd, Taiwan) for sequence conformation of the plasmid constructs.

4.4.2 Cloning of Corndog_gp36, Corndog_gp37, and Corndog_gp38

The cloning of mycobacteriophage gene Corndog_gp36, Corndog_gp37 and Corndog_gp38 were done by performing PCR using a Clubio Technology polymerase chain reactor. The PCR reactions with total volumes of 50 µL (Table 4-1) were performed with specifically designed primers generated using NEBaseChanger online application (https://nebasechanger.neb.com/) (Table 4-2). Each reaction underwent 30 annealing cycles at temperatures recommended by ThermoFisher based on their Tm calculator (available at www.thermofisher.com/tmcalculator). The annealing temperatures from the ThermoFisher Tm calculator were used because the annealing rules for Phusion DNA Polymerases are different from many common DNA polymerases. The PCR products were treated with DnpI (New England Biolabs, USA) then incubated at room temperature for 5 min after PCR. Annealing inserts with vectors were done by incubating 3 µL of ddH_2O with 2 µL of insert and 3 µL vector at room temperature (25 °C) for 10 min. This was done by adding 2 µl 100 mM EDTA to the mixture, incubation at 75 °C for 10 min then additional 10 min incubation at 25 °C.

doi:10.6342/NTU202402746

doi:10.6342/NTU202402746

Table 4-1 PCR mix for Corndog_gp36, Corndog_gp37, Corndog_gp38 and LIC_Vector.

	Corndog_gp36	**Corndog_gp37**	**Corndog_gp38**	**Vector**
ddH_2O	20 μL	20 μL	20 μL	7 μL
2X Phusion Flash MM	25 μL	25 μL	25 μL	10 μL
FWD primer	2 μL (0.5 μM)	2 μL (0.5 μM)	2 μL (0.5 μM)	1 μL (0.5 μM)
RVRS primer	2 μL (0.5 μM)	2 μL (0.5 μM)	2 μL (0.5 μM)	1 μL (0.5 μM)
DNA template	1 μL (20 ng)	1 μL (20 ng)	1 μL (20 ng)	1 μL (10 ng)
	50 μL	50 μL	50 μL	20 μL

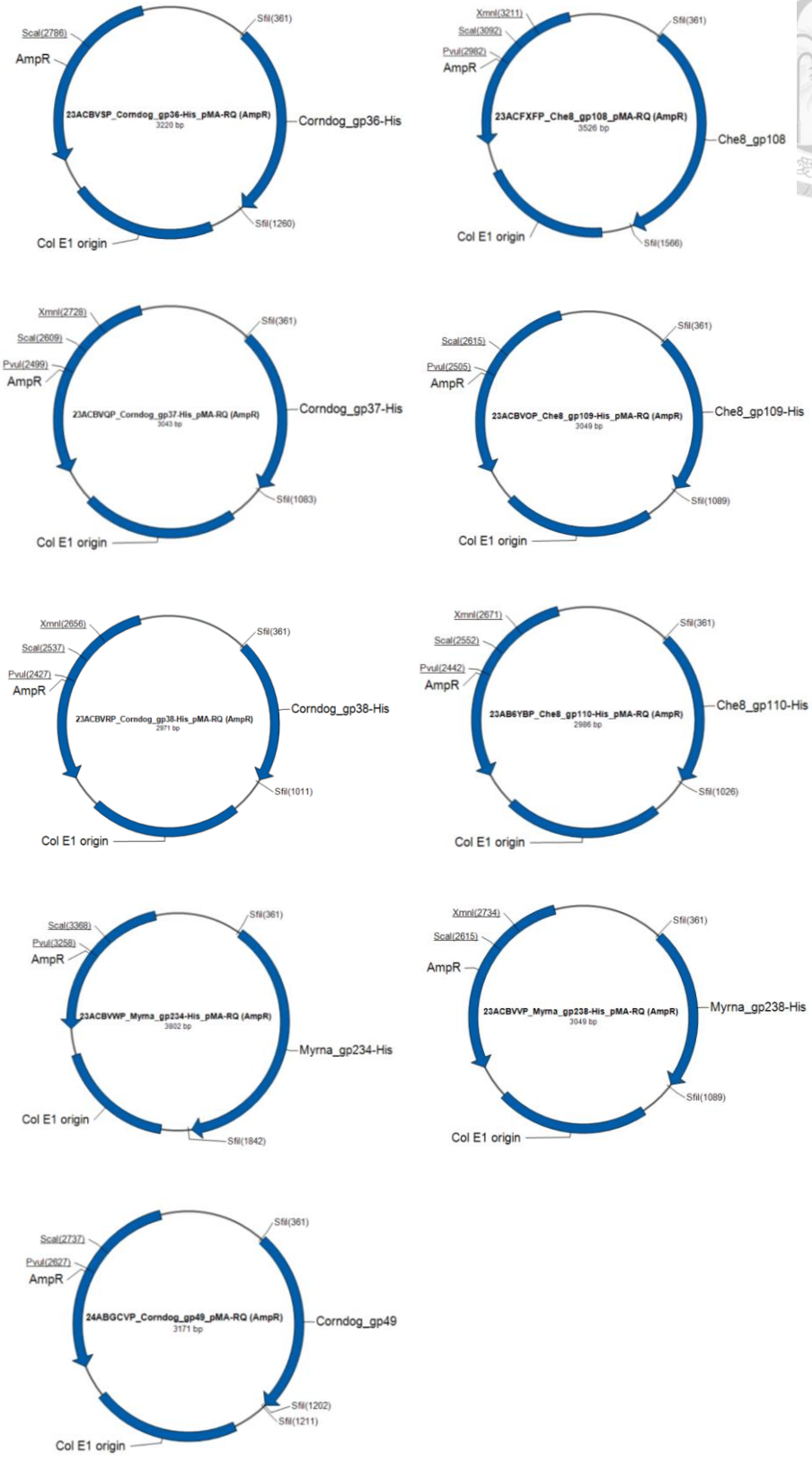


Figure 4-1 Custom-synthesized plasmid used in this study.

doi:10.6342/NTU202402746

Table 4-2 Forward and reverse primers for Corndog_gp36, Corndog_gp37, Corndog_gp38 and LIC_Vector used in this study.

PRIMER (NAME F/R)	OLIGO (5'-3')	LENGTH (BP)	%GC	TM
Corndog_36_FW	TAC TTC CAA TCC AAT GGG ATG AGC GAT CGT CCG GTT ACA CTG	42	50.0%	85.9 °C
Corndog_36_RV	TTA TCC ACT TCC AAT GGC TAT TAC AGA CCA CCA CCA GGT GTA TTC G	46	45.7%	83.4 °C
Corndog_37_FW	TAC TTC CAA TCC AAT GGG ATG AAA GTT GCA GTT GTT ATT CCG	42	40.5%	82.0 °C
Corndog_37_RV	TTA TCC ACT TCC AAT GGC TAT TAA CGT GCG GTC AGT TTA CGA ATC	45	42.2%	83.8 °C
Corndog_38_FW	TAC TTC CAA TCC AAT GGG ATG ACA CTG ATT GGT ATT GTT GC	41	41.5%	82.0 °C
Corndog_38_RV	TTA TCC ACT TCC AAT GGC TAA ATT GCA ATG GTT TTA TTA CGC CAC	45	37.8%	82.0 °C
LIC_Vector_FW	ATT GGA AGT GGA TAA CGG ATC CG	23	47.8%	71.8 °C
LIC_Vector_RV	ATT GGA TTG GAA GTA CAG GTT CTC GGT ACC	30	46.7%	78.1 °C

4.4.3 Cloning of Corndog_gp49

The cloning of mycobacteriophage gene Corndog_gp49 involved the use of custom-synthesized plasmid (Figure 4-1), which was transformed into *E. coli* DH5α cells via the heat shock method. After overnight culture at 37 °C, the plasmid was amplified using the Presto™ Mini Plasmid Kit (Generaid Biotech Ltd, Taiwan). The Corndog_gp49, gene insert was double-digested from the custom plasmid with BamH1 and Not1 restriction enzymes at 37 °C for 2 hours. The gene insert was identified and extracted using 1% agarose gel electrophoresis then purified with the GenepHlow™ Gel/PCR Kit (Generaid Biotech Ltd, Taiwan). The purified gene insert was ligated into pET28a(+) vector using T4 DNA Ligase (New England Biolabs, USA), followed by transformation into *E. coli* BL21(DE3) competent cells through heat shock method and overnight incubation at 37 °C. The entire LB agar plate was sent to Mission Biotech Ltd for DNA sequencing using T7 forward and T7t reverse sequencing primers (Mission Biotech Ltd, Taiwan) for the final plasmid confirmation.

4.4.4 Isopropyl β-D-1-thiogalactopyranoside (IPTG) induction

Final plasmid constructs with correct sequences were transformed into *E. coli* BL21(DE3). Recombinant strains harboring the respective plasmids were grown in LB broth medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L sodium chloride) with either 50 µg/mL kanamycin or 100 µg/mL ampicillin antibiotics. The cultivation of the *E. coli* BL21(DE3) was done at 37 °C, with agitation at 200 rpm in a S300R Orbital Shaking Incubator (Firstek, New Taipei City, Taiwan) until to an OD_{600} of approximately 0.6. Protein expression was induced with 0.5 to 1 mM of IPTG at 16 °C overnight (16–20 h).

doi:10.6342/NTU202402746

4.5 Cell resuspension and disruption

A Heraeus Megafuge 8R Centrifuge (ThermoFisher, Germany) was used in the harvesting of the cells (4 °C, 6,000 × g, 30 min). The recovered pellets were resuspended in appropriate volumes (10^9 cells/mL or 40 mL per pellet of 1L culture) of lysis buffer (20 mM Tris HCl, 200 mM NaCl, pH 7.5). One tablet of cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche, Basel, Switzerland), 40μL Triton X-100 and 2μL of Benzonase were then added for every 40 mL of resuspended samples. The mixtures were then injected though a 25-gauge x 1" (0.50 x 25 mm) needlepoint to reduce major clumping in the samples. The expressed proteins were released into the supernatant solution by using a Nanolyzer N2 Desktop High-Pressure Homogenizer (Gogene Ltd, Taiwan). During the process, the cooler was first powered on 15 min prior to the experiment to ensure the sample reservoir and loops were equilibrated at 4 °C. Then, 200 mL of cold ddH_2O was slowly poured into the reservoir to wash the entire system loops under 5,000 psi. This process was followed by another washing step with 50 mL of cold lysis buffer. Cells were then disrupted, and the lysates were collected under a single 18,000 psi cycle at 4 °C. After cell disruption, the sample reservoir and loops were washed with 50 mL of cold lysis buffer, followed immediately by an additional washing step with 500 mL of ddH_2O to avoid contamination. Finally, the bottom of the reservoir cup was submerged and stored with 50 mL of 20% EtOH to halt bacterial and fungal growth. The collected lysates were centrifuged at 4 °C, 20,000 × g for 30 min to pellet cell debris. The supernatant containing the proteins were transferred to new 50 mL Falcon tubes. Protein samples were then purified by immobilized metal affinity chromatography (IMAC) purification and the expression levels were verified by SDS-PAGE analysis.

doi:10.6342/NTU202402746

4.6 Immobilized metal affinity chromatography (IMAC) purification

The isolated proteins were purified using Ni-NTA agarose resin. The bacterial lysates were prepared, and the His-tagged proteins were purified under native conditions. The Ni-NTA resin was equilibrated first by washing with 5 column volumes of double distilled water followed by 5 column volumes of wash buffer (200 mM Tris HCl, 500 mM NaCl, 10 mM imidazole, pH 7.5). The resin was allowed to settle and air bubbles removed by gently tapping the column. For every lysate runs, the sample was added slowly onto the Ni-NTA column and allowed to sit for 30 min at 4 ℃ to provide adequate time for His-tagged protein to bind to the resin. The samples were then eluted at a flow rate of 1 mL/min. The flow though fractions were collected for later analysis. The washing step was done by washing the column with 10–20 column volumes of wash buffer (200 mM Tris HCl, 500 mM NaCl, 10 mM imidazole, pH 7.5). This step was performed two times to ensure the removal of all non-specifically bound proteins. Both wash fractions were collected for later analysis. The elution of the bound His-tagged proteins was done with 5 column volumes of elution buffer containing 50 mM, 100 mM, 250 mM and 500 mM of imidazole. Washing and elution were carried out at 4 ℃ and all fractions were kept at –80 ℃ until further use. The eluted proteins were analyzed by SDS-PAGE and concentrated by using Amicon Ultra centrifugal filters (Merck, Darmstadt, Germany). The final protein concentrations was determined by measuring the absorbance at 280 nm using a Nanodrop spectrophotometer (Thermo Fisher Scientific) or using the Bradford assay.

doi:10.6342/NTU202402746

4.7 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

The protein samples were mixed with SDS sample buffer and heated at 95 °C for 5 min. The proteins were separated on a 12% SDS-PAGE resolving gel (3.3 mL ddH_2O, 4.0 mL 30% acrylamide mix, 2.5 mL 1.5 M Tris (pH 8.8), 0.1 mL 10% SDS, 0.1 mL 10% APS, and 4.0 μL tetramethylethylenediamine), with 5% acrylamide stacking gel (2.1 mL ddH_2O, 0.5 mL 30% acrylamide mix, 0.38 mL 1.5 M Tris (pH 8.8), 0.03 mL 10% SDS, 0.03 mL 10% APS, and 2.0 μL tetramethylethylenediamine). Tris-Glycine-SDS was used as the running buffer for all SDS-PAGE. The electrophoresis runs were conducted at a voltage of 60 volts until the dye front reached the top of the resolving gel and at 120 volts for additional 90 min until the dye front reached the bottom of the resolving gel. The gels were stained with Coomassie Brilliant Blue R-250 and destained twice using ddH_2O for 20 min with 200 rpm shaking until bands were clearly visible.

4.8 Protein concentration assay

4.8.1 Bradford assay

Protein concentrations were determined using the Bradford protein assay. Samples were mixed with the Bradford reagent and the absorbance measured at 595 nm using a BioTek Synergy HTX multi-mode reader (Cold Spring Biotech, New Taipei City, Taiwan). BSA was used as the standard for the calibration curve. First, a standard curve was created with bovine serum albumin (BSA). Protein dye was prepared by mixing one part InstantBlue® Coomassie Protein Stain (Expedeon Ltd, United Kingdom) with four parts sterile ddH_2O. To generate a standard curve, 20 μL of BSA samples of known concentrations (0.2–1 mg/mL) were mixed with 980 μL of the protein dye. The mixture was homogenized by vortexing and left to stand at room temperature for 5 min.

doi:10.6342/NTU202402746

Absorbances were then measured at 595 nm using a BioTek Synergy HTX multi-mode reader. For each sample, 200 μL of the dye-protein mix was placed in a well of a 96-well plate. To determine the concentration of proteins, 20 μL of each sample was mixed with 980 μL of the protein dye. The samples were homogenized, and the absorbance measured at 595 nm after 5 min. Protein concentrations were calculated using the BSA standard curve.

4.8.2 Nanodrop protein concentration detection

Protein concentrations were also measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific). For this method, 2 μL of each sample was used to measure the absorbance at 280 nm. Before measurement, the spectrophotometer was calibrated with a blank sample of the buffer or ddH_2O used to dissolve the proteins. Measurements were taken in triplicate to ensure accuracy and consistency of the results. The obtained protein concentrations were recorded and compared with those measured by the Bradford assay to validate the consistency between the two methods.

4.9 Nucleotide sugar donor enzymatic synthesis

4.9.1 UDP-GlcNAc enzymatic synthesis

UDP-GlcNAc was synthesized using NahK and AGX1 enzymes[48]. The synthesis of UDP-GlcNAc was carried out by mixing 250 μL of 1M HEPES buffer (100 mM, pH8), 50 μL of 1M magnesium chloride (20 mM), adenosine triphosphate (66 mg 0.12 mM, 1.2 eq.), GlcNAc (22 mg, 0.1 mM, 1 eq.), and 68 μL of 7.3 mg/mL NahK. The reaction mixture was incubated at 37 °C with agitation at 200 rpm overnight (16–18 h). The progression of the reaction (formation of GlcNAc-1-phosphate) was monitored by thin

doi:10.6342/NTU202402746

layered chromatography (TLC) using ethyl acetate–methanol–water–acetic acid = 4:2:1:0.5 (*v/v/v*) as the solvent system. Once GlcNAc was no longer observed on the TLC, uridine triphosphate (58 mg, 0.12 mM, 1.2 eq.) and 96 µL of 1.3 mg/mL AGX1 were added to the mixture. The reaction mixture was then incubated at 37 °C, with agitation at 200 rpm overnight (16–18 h). Formation of the final product (UDP-GlcNAc) was monitored by TLC using *n*-butanol–water–acetic acid = 2:1:1 (*v/v/v*) as the solvent system. Ethanol was added once GlcNAc-1-phosphate was no longer observed on the TLC.

4.9.2 UDP-GalNAc enzymatic synthesis

UDP-GalNAc was synthesized using a similar protocol and enzymes as mentioned in the previous section (Section 4.9.1). The synthesis of UDP-GalNAc was done by mixing 250 µL of 1M HEPES buffer (100 mM, pH8), 50 µL of 1 M magnesium chloride (20 mM), adenosine triphosphate (297 mg 0.12 mM, 1.2 eq.), GalNAc (100 mg, 0.1 mM, 1 eq.), and 68 µL of 7.3 mg/mL NahK. The reaction was incubated at 37 °C with agitation at 200 rpm overnight (16–18 h). The progression of the reaction (GalNAc-1-phosphate) was monitored by TLC using ethyl acetate–methanol–water–acetic acid = 4:2:1:0.5 (*v/v/v*) as the solvent system. Once GalNAc was no longer observed on the TLC, uridine triphosphate (261 mg, 0.12 mM, 1.2 eq.) and 96 µL of 1.3 mg/mL AGX1 were added to the mixture and incubated at 37 °C with agitation at 200 rpm overnight (16–18 h). The final product (UDP-GalNAc) was monitored by TLC using *n*-butanol–water–acetic acid = 2:1:1 (*v/v/v*) as the solvent system. Ethanol was added once GalNAc-1-phosphate was no longer observed on the TLC.

doi:10.6342/NTU202402746

4.10 Nucleotide sugar donor purification

The nucleotide sugars synthesized in Sections 4.9.1 and 4.9.2 were purified by slowly adding barium chloride into the solution until no more precipitation was observed. The mixture was centrifuged at 6000 × g for 15 min and the supernatants were collected and transferred to new 50 mL Falcon tube. The pellets were washed with ddH_2O and centrifuged again at 6000 × g for 15 min. The supernatants were then collected to the same 50 mL Falcon tube. Size-exclusion chromatography gel (HW-40F) was used to isolate and purify the nucleotide sugar donors from the supernatant. The desired products were detected by TLC using *n*-butanol–water–acetic acid = 2:1:1 (*v/v/v*) as the solvent system. Every eluted fractions containing pure final product were pooled together and then lyophilized using a UNISS Manifold Freeze Dryer (UNISS, Taiwan). The final dried powder products were then stored in -80 ℃ freezer until further use.

4.11 Peptide acceptor synthesis

Peptide acceptors pMBz-EWVSGESWTDIKGDS and GLAGGS(GlcNAc)G were synthesized by Institute of Biological Chemistry Synthesis Core Facility (IBCSCF). Peptide acceptors were synthesized using standard solid-phase peptide synthesis (SPPS) techniques. Peptides were cleaved from the resin, purified by High-Performance Liquid Chromatography (HPLC), and characterized by Matrix-Assisted Laser Desorption Ionization–time-of-flight mass spectrometry (MALDI-TOF) mass spectrometry.

doi:10.6342/NTU202402746

4.12 Mycobacteriophage glycosyltransferase enzymatic assay

The glycosyltransferase activity of Corndog_gp36, Corndog_gp37, and Corndog_gp38 were assessed by using these expressed mycobacteriophage glycosyltransferase proteins (1.0 mg/mL) and the synthesized nucleotide sugar donors (1.0 eq., or 5.0 eq.) and peptide acceptors (1 mM, 1 eq.). Reactions were conducted in Tris-HCl buffer (100 mM, pH 5–8) containing divalent cations such as magnesium chloride or manganese chloride (20 mM) and incubated at 37 °C with agitation at 200 rpm overnight (16–18 h). The products were analyzed by TLC using ethyl acetate–-methanol–water–acetic acid = 4:2:1:0.5 (*v/v/v*) as the solvent system. Electrospray ionization mass spectrometry (ESI-MS) and Matrix-Assisted Laser Desorption Ionization mass spectrometry (MALDI-MS) were performed on the sample mixtures to confirm the outcome of the reactions.

doi:10.6342/NTU202402746

References

1. Shinnick, T. M.; Good, R. C. Mycobacterial taxonomy. *European Journal of Clinical Microbiology & Infectious Diseases*. **1994**, 13 (11), 884-901.
2. Parte, A. C.; Sardà Carbasse, J.; Meier-Kolthoff, J. P.; Reimer, L. C.; Göker, M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*. **2020**, 70 (11), 5607-5612.
3. Koch, R. Die Äetiologie der Tuberkulose. *Berliner klinische Wochenschrift*. **1882**, 19 (15), 221–230.
4. World Health Organization, *On the Road to Ending TB: Highlights from the 30 Highest TB Burden Countries*; World Health Organization, 2016.
5. World Health Organization, *Global tuberculosis report 2023*; World Health Organization, 2023.
6. McQuaid, C. F.; Vassall, A.; Cohen, T.; Fiekert. K.; White, R. G. The impact of COVID-19 on TB: a review of the data. *The International Journal of Tuberculosis and Lung Disease*. **2021**, 25 (6), 436-446.
7. Joel, C. L.; Marilda Aparecida, M. M. A. Leprosy: review of the epidemiological, clinical, and etiopathogenic aspects - Part 1. *Anais Brasileiros de Dermatologia*. **2014**, 89 (2), 05–218.
8. Summers,W. C., *Bacteriophage discovered. In Felix d'Herelle and the Origins of Molecular Biology*; Yale University Press, 1999.
9. Abedon, S. T., *Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses*; Cambridge University Press, 2008.

doi:10.6342/NTU202402746

10. Fauquet, C. M.; Pringle, C. R. Abbreviations for bacterial and fungal virus species names. *Archives of Virology*. **2000**, 145, 197–203.
11. Van Regenmortel, M. H. Virus species and virus identification: past and current controversies. *Infection Genetics and Evolution*. **2006**, 7, 133–144.
12. Catalao, M. J.; Gil, F.; Moniz-Pereira, J.; Sao-Jose, C.; Pimentel, M. Diversity in bacterial lysis systems: bacteriophages show the way. *FEMS Microbiology Reviews*. **2013**, 37 (4), 554-571.
13. Little, J. W., *Lysogeny, Prophage Induction, and Lysogenic Conversion*; Wiley Online library, 2005.
14. Campbell, A. The future of bacteriophage biology. *Nature Reviews Genetics*. **2003**, 4, 471–477.
15. Hatfull, G. F.; Dedrick, R. M.; Schooley, R. T. Phage Therapy for Antibiotic-Resistant Bacterial Infections. *Annual Review of Medicine*. **2022**, 73, 197-211.
16. Harper, D.R. Introduction to Bacteriophages. In: Harper, D.R.; Abedon, S.T.; Burrowes, B.H.; McConville, M.L. (eds) Bacteriophages. *Springer*. 2021; pp 3-16.
17. Hatfull, G. F. Actinobacteriophages: Genomics, Dynamics, and Applications. *Annual Review of Virology*. **2020**, 7 (1), 37-61.
18. Russell, D. A.; Hatfull, G. F. PhagesDB: the actinobacteriophage database. *Bioinformatics*. **2017**, 33 (5), 784-786.
19. Taher, A.; Mehrdad, M.; Mohammad, J. N.; Sahar, S.; Samira, K.; Ahmad, N. Phage therapy as a renewed therapeutic approach to mycobacterial infections: a comprehensive review. *Infection and Drug Resistance*. **2019**, 12, 2943–2959.

doi:10.6342/NTU202402746

20. Smith, H. W.; Huggins, M. B. Successful treatment of experimental Escherichia coli infections in mice using phage: its general superiority over antibiotics. *Journal of General Microbiology*. **1982**, 128 (2), 307-318.
21. Chanishvili N. Phage therapy—history from Twort and d'Herelle through Soviet experience to current approaches. *Advances in Virus Research*. **2012**, 83 (3), 40.
22. Kortright, K. E.; Chan, B. K.; Koff, J. L.; Turner, P. E. Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell Host & Microbe.* **2019**, 25, 219–32.
23. Jiaxing, Y.; Laiying, Z.; Wenliang, Q.; Youfu, Luo. Mycobacterium tuberculosis: Pathogenesis and therapeutic targets. *MedComm*. **2020**, 4 (5), 353.
24. Singh, V.; Chibale, K. Strategies to Combat Multi-Drug Resistance in Tuberculosis. *Accounts of Chemical Research*. **2021**, 54 (10), 2361-2376.
25. Hatfull, G. F. Mycobacteriophages: windows into tuberculosis. *PLOS Pathogens*. **2014**, 10 (3).
26. Azimi, T.; Mosadegh, M.; Nasiri, M.J.; Sabour, S.; Karimaei, S.; Nasser, A. Phage therapy as a renewed therapeutic approach to mycobacterial infections: A comprehensive review. *Infection and Drug Resistance*. **2019**, 12, 2943–2959.
27. Sula, L.; Sulova, J.; Stolcpartova, M. Therapy of experimental tuberculosis in guinea pigs with mycobacterial phages DS-6A, GR-21 T, My-327. *CzechMed*. **1981**, 4, 209–214.
28. Peng, L.; Luo, Y.; Chen, B.; Li, Y.; Shen, X.; Su, C.; Wang, G. Therapeutic effect of bacteriophage D29 in the treatment for guinea pigs infected with sensitive strain of Mycobacterium tuberculosis. *Chinese Journal of Zoonoses.* **2009**, 25, 733–736.

doi:10.6342/NTU202402746

29. Hatfull, G. F. Phage Therapy for Nontuberculous Mycobacteria: Challenges and Opportunities. *Pulmonary Therapy.* **2023**, 9 (1), 91–107.
30. Dedrick, R. M.; Freeman, K. G.; Nguyen, J. A. Potent antibody-mediated neutralization limits bacteriophage treatment of a pulmonary Mycobacterium abscessus infection. *Nature Medicine.* **2021**, 27, 1357–1361.
31. Moremen, K.W.; Tiemeyer, M.; Nairn, A.V. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nature Reviews Molecular Cell Biology.* **2012**, 13, 448–462.
32. Reily, C.; Stewart, T.J.; Renfrow, M.B.; Novak, J. Glycosylation in health and disease. *Nature Reviews Nephrology*. **2019**, 15, 346–366.
33. Schjoldager, K.T.; Narimatsu, Y.; Joshi, H.J.; Clausen, H. Global view of human protein glycosylation pathways and functions. *Nature Reviews Molecular Cell Biology*. **2020**, 21, 729–749.
34. Varki, A. Biological roles of glycans. *Glycobiology*. **2017**, 27, 3–49.
35. Freeman, K. G.; Robotham, A. C.; Parks, O. B.; Abad, L.; Jacobs-Sera, D.; Lauer, M. J.; Podgorski, J. M.; Zhang, Y.; Williams, J. V.; White, S. J.; Kelly, J. F.; Hatfull, G. F.; Pope, W. H. Virion glycosylation influences mycobacteriophage immune recognition. *Cell Host & Microbe*. **2023**, 31, 1216.
36. Christelle, B.; Lenka Šnajdrová; Charlotte, J.; Jaroslav, K.; Anne, I. Structures and mechanisms of glycosyltransferases. *Glycobiology*. **2006**, 16 (2), 29-37.
37. Rini, J. M.; Moremen, K. W.; Davis, B. G.; Esko, J. D. Glycosyltransferases and Glycan-Processing Enzymes. In Essentials of Glycobiology, 4th ed.; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Mohnen, D.,

doi:10.6342/NTU202402746

Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2022; pp. 67–78.

38. Pasala, C.; Sharma, S.; Roychowdhury, T.; Moroni, E.; Colombo, G.; Chiosis, G. N-Glycosylation as a Modulator of Protein Conformation and Assembly in Disease. *Biomolecules.* **2024**, 14, 282.
39. Lairson, L. L.; Henrissat, B.; Davies, G. J.; Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annual Review of Biochemistry*. **2008**, 77, 521-55.
40. Moremen, K. W.; Haltiwanger, R. S. Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nature Chemical Biology*. **2019**, 15 (9), 853-864.
41. Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H. Essentials of Glycobiology [Internet]. 3rd ed.; Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 2015.
42. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G (1998). "Concepts and principles of O-linked glycosylation". Critical Reviews in Biochemistry and Molecular Biology. 33 (3): 151–208.
43. Vik, A.; Aas, F. E.; Anonsen, J. H.; Bilsborough, S.; Schneider, A.; Egge-Jacobsen, W.; Koomey, M. Broad spectrum O-linked protein glycosylation in the human pathogen Neisseria gonorrhoeae. *Proceedings of the National Academy of Sciences of the United States of America*. **2009**, 106 (11), 4447–52.
44. Lithgow, K. V.; Scott, N. E.; Iwashkiw, J. A.; Thomson, E. L.; Foster, L. J.; Feldman, M. F.; Dennis, J. J. A general protein O-glycosylation system within the Burkholderia

doi:10.6342/NTU202402746

cepacia complex is involved in motility and virulence. *Molecular Microbiology*. **2014**, 92 (1), 116–37.

45. Iwashkiw, J. A.; Seper, A.; Weber, B. S.; Scott, N. E.; Vinogradov, E.; Stratilo, C. Identification of a general O-linked protein glycosylation system in Acinetobacter baumannii and its role in virulence and biofilm formation. *PLoS Pathogens*. **2012**, 8 (6).
46. Hounsell, E. F.; Davies, M. J.; Renouf, D. V. O-linked protein glycosylation structure and function. *Glycoconjugate Journal*. **1996**, 13 (1), 19–26.
47. Lin, B.; Qing, X.; Liao, J.; Zhuo, K. Role of Protein Glycosylation in Host-Pathogen Interaction. *Cells*. **2020**, 9 (4), 1022.
48. Guan, W.; Cai, L.; Wang, PG. Highly efficient synthesis of UDP-GalNAc/GlcNAc analogues with promiscuous recombinant human UDP-GalNAc pyrophosphorylase AGX1. *Chemistry*. **2010**, 16 (45), 13343–13345.
49. Shuang, Li.; Shuaishuai, Wang.; Yaqian. Wang.; Jingyao, Qu.; Xian-wei, Liu.; Peng George, Wang.; Junqiang, Fang. Gram-scale production of sugar nucleotides and their derivatives. *Green Chemistry*. **2021**, 23, 2628–2633.

doi:10.6342/NTU202402746

Appendix

Protein View: gi|2024040301

HIS-CORNDOG_36

Database: IBC-114
Score: 120
Expect: 3.8e-011
Monoisotopic mass (M_r): 35976
Calculated pI: 6.20

Sequence similarity is available as an NCBI BLAST search of gi|2024040301 against nr.

Search parameters

Enzyme: Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Acetyl (Protein N-term), Oxidation (M)
Mass values searched: 30
Mass values matched: 11

Protein sequence coverage: 38%

Matched peptides shown in ***bold red***.

```
  1 MHHHHHHSSG VDLGTENLYF QSNGMSDRPV TLFMFAGREG NMRCNLPLIR
 51 QILDENPNVQ FDIWNLARKP ADAEYLRTIQ SGSGLRVINN FAGPRAYRFL
101 NKVWSYYAQD KFKDQLFVKM DDDVVFIETE KFDAFVDEVE AHPEHVLSAE
151 VINNGACTEF VPGLNTKFRK MGIDLLDVHE SNAYALMAHQ HMERNWEKLV
201 GRKTSVTDIE TWLSINFIGM NWDVLRRLNG HIGNRAPEWI AGRQWLPHHR
251 IGDEGAANML PRAVMQGFTV GHLGFGPQKL TDAQEDEWRE VYTKIGQKYL
301 LDVRKPNTPG GGL
```

Unformatted sequence string: 313 residues (for pasting into other applications).

Sort by (•) residue number ( ) increasing mass ( ) decreasing mass
Show (•) matched peptides only ( ) predicted peptides also

Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
39 – 43	622.2460	621.2387	621.2540	-24.7	0	R.EGNMR.C + Oxidation (M)
44 – 50	885.5028	884.4955	884.4902	6.02	0	R.CNLPLIR.Q
69 – 77	1062.5528	1061.5456	1061.5505	-4.67	0	R.KPADAEYLR.T
78 – 86	918.4951	917.4878	917.4930	-5.71	0	R.TIQSGSGLR.V
87 – 95	987.5345	986.5272	986.5298	-2.55	0	R.VINNFAGPR.A
171 – 194	2829.2323	2828.2250	2828.2680	-15.2	0	K.MGIDLLDVHESNAYALMAHQHMER.N + 3 Oxidation (M)
236 – 243	899.4727	898.4654	898.4661	-0.77	0	R.APEWIAGR.Q
251 – 262	1259.5962	1258.5889	1258.5976	-6.88	0	R.IGDEGAANMLPR.A + Oxidation (M)
263 – 279	1789.9013	1788.8941	1788.8982	-2.30	0	R.AVMQGFTVGHLGFGPQK.L + Oxidation (M)
280 – 294	1882.8782	1881.8709	1881.8745	-1.88	1	K.LTDAQEDEWREVYTK.I
299 – 304	778.4456	777.4383	777.4385	-0.19	0	K.YLLDVR.K

Index

	Accession	Mass	Score	Description
1.	gi|2024040301	35976	120	HIS-CORNDOG_36
2.	gi|2023011102	28379	22	Wzt
3.	gi|2022053102	35200	6	Optimus_52
4.	gi|2021101504	12794	0	Sumo
5.	gi|2022011702	51436	0	Nanosmite_27 with MBP tag (Sample 2)
6.	gi|2022012503	43738	0	His-SUMO-Bxb1_19 (Sample 3)
7.	gi|2022012701	55238	0	Sample 1
8.	gi|2022012702	53211	0	Sample 2
9.	gi|2022021502	51227	0	His-MBP-Nanosmite_27
10.	gi|2022021601	53211	0	WYC_01

Results List

1. gi|2024040301 **Mass:** 35976 **Score:** 120 **Expect:** 3.8e-011 **Matches:** 11
HIS-CORNDOG_36

Observed	Mr(expt)	Mr(calc)	ppm	Start	End	Miss	Peptide
622.2460	621.2387	621.2540	-24.65	39	43	0	R.EGNMR.C + Oxidation (M)
778.4456	777.4383	777.4385	-0.19	299	304	0	K.YLLDVR.K
885.5028	884.4955	884.4902	6.02	44	50	0	R.CNLPLIR.Q
899.4727	898.4654	898.4661	-0.77	236	243	0	R.APEWIAGR.Q
918.4951	917.4878	917.4930	-5.71	78	86	0	R.TIQSGSGLR.V
987.5345	986.5272	986.5298	-2.55	87	95	0	R.VINNFAGPR.A
1062.5528	1061.5456	1061.5505	-4.67	69	77	0	R.KPADAEYLR.T
1259.5962	1258.5889	1258.5976	-6.88	251	262	0	R.IGDEGAANMLPR.A + Oxidation (M)
1789.9013	1788.8941	1788.8982	-2.30	263	279	0	R.AVMQGFTVGHLGFGPQK.L + Oxidation (M)
1882.8782	1881.8709	1881.8745	-1.88	280	294	1	K.LTDAQEDEWREVYTK.I
2829.2323	2828.2250	2828.2680	-15.19	171	194	0	K.MGIDLLDVHESNAYALMAHQHMER.N + 3 Oxidation (M)

Figure S1 Protein ID results of His-Corndog_gp36

doi:10.6342/NTU202402746

Protein View: gi|2024040302

His-CORNDOG_37

Database:	IBC-114
Score:	255
Expect:	1.2e-024
Monoisotopic mass (Mr):	28184
Calculated pI:	6.05

Sequence similarity is available as an NCBI BLAST search of gi|2024040302 against nr.

Search parameters

Enzyme:	Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Acetyl (Protein N-term), Oxidation (M)
Mass values searched:	69
Mass values matched:	22

Protein sequence coverage: 72%

Matched peptides shown in ***bold red***.

```
  1 MHHHHHHSSG VDLGTENLYF QSNGMKVAVV IPFRDRGLDP LRLENLARVA
 51 DHWAGYSGEV IVSGDGRSGD AHFNRSAAYN RGVTDTDADM LIFAESDLLV
101 SYAQIDRAIE MASDSLGMVV PFSWFMALSP EDSVRVRNRE VEPEGLSATP
151 IKGHRGSIGA INVLSRDTYE AVGGYDEQFE GAWYDDDAMK IAFDVAAGPT
201 RWVEGSAYHL YHLSGGRGAH LTAEDRAATA RNRRRLRLYR QARTAEQIRK
251 LTAR
```

Unformatted sequence string: 254 residues (for pasting into other applications).

Sort by ◉ residue number ○ increasing mass ○ decreasing mass
Show ◉ matched peptides only ○ predicted peptides also

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
1 - 26	3016.2383	3015.2310	3015.3253	-31.3	0	-.MHHHHHHSSGVDLGTENLYFQSNGMK.V + Oxidation (M)
27 - 34	900.5617	899.5544	899.5593	-5.39	0	K.VAVVIPFR.D
27 - 36	1171.6849	1170.6776	1170.6873	-8.30	1	K.VAVVIPFRDR.G
35 - 48	1637.8885	1636.8812	1636.9009	-12.0	2	R.DRGLDPLRLENLAR.V
37 - 42	670.3728	669.3655	669.3810	-23.0	0	R.GLDPLR.L
37 - 48	1366.7649	1365.7576	1365.7728	-11.1	1	R.GLDPLRLENLAR.V
43 - 48	715.3946	714.3873	714.4024	-21.1	0	R.LENLAR.V
49 - 67	1974.9190	1973.9117	1973.9232	-5.80	0	R.VADHWAGYSGEVIVSGDGR.S
68 - 75	903.3990	902.3917	902.3995	-8.62	0	R.SGDAHFNR.S
76 - 81	681.3205	680.3132	680.3242	-16.1	0	R.SAAYNR.G
82 - 107	2858.3159	2857.3086	2857.3692	-21.2	0	R.GVTDTDADMLIFAESDLLVSYAQIDR.A
82 - 107	2874.3134	2873.3061	2873.3641	-20.2	0	R.GVTDTDADMLIFAESDLLVSYAQIDR.A + Oxidation (M)
138 - 152	1639.8525	1638.8453	1638.8577	-7.57	1	R.NREVEPEGLSATPIK.G
138 - 155	1990.0220	1989.0147	1989.0392	-12.3	2	R.NREVEPEGLSATPIKGHR.G
140 - 155	1719.8960	1718.8888	1718.8951	-3.70	1	R.EVEPEGLSATPIKGHR.G
156 - 166	1086.6137	1085.6064	1085.6193	-11.8	0	R.GSIGAINVLSR.D
191 - 201	1117.5886	1116.5813	1116.5928	-10.2	0	K.IAFDVAAGPTR.W
202 - 217	1831.8757	1830.8685	1830.8802	-6.39	0	R.WVEGSAYHLYHLSGGR.G
218 - 226	969.4585	968.4512	968.4675	-16.9	0	R.GAHLTAEDR.A
236 - 240	720.4491	719.4418	719.4442	-3.38	1	R.LRLYR.Q
244 - 249	717.3789	716.3716	716.3817	-14.0	0	R.TAEQIR.K
244 - 250	845.4974	844.4901	844.4766	16.0	1	R.TAEQIRK.L

Index

	Accession	Mass	Score	Description
1.	gi|2024040302	28184	255	His-CORNDOG_37
2.	gi|2024040303	25354	5	His-CORNDOG_38
3.	gi|2024040301	35976	4	HIS-CORNDOG_36
4.	gi|2022012703	35236	4	Sample 3
5.	gi|2022021602	35236	4	WYC_02
6.	gi|20220301	35236	4	D29
7.	gi|2022053101	35236	4	D29_10
8.	gi|20220705	35236	4	Endolysin D29
9.	gi|2021101504	12794	0	Sumo
10.	gi|2023011101	29671	0	Wzm

Results List

1. gi|2024040302 **Mass:** 28184 **Score:** 255 **Expect:** 1.2e-024 **Matches:** 22
His-CORNDOG_37

Observed	Mr(expt)	Mr(calc)	ppm	Start	End	Miss	Peptide
670.3728	669.3655	669.3810	-23.05	37	42	0	R.GLDPLR.L
681.3205	680.3132	680.3242	-16.05	76	81	0	R.SAAYNR.G
715.3946	714.3873	714.4024	-21.15	43	48	0	R.LENLAR.V
717.3789	716.3716	716.3817	-14.03	244	249	0	R.TAEQIR.K
720.4491	719.4418	719.4442	-3.38	236	240	1	R.LRLYR.Q
845.4974	844.4901	844.4766	16.0	244	250	1	R.TAEQIRK.L
900.5617	899.5544	899.5593	-5.39	27	34	0	K.VAVVIPFR.D
903.3990	902.3917	902.3995	-8.62	68	75	0	R.SGDAHFNR.S
969.4585	968.4512	968.4675	-16.85	218	226	0	R.GAHLTAEDR.A
1086.6137	1085.6064	1085.6193	-11.85	156	166	0	R.GSIGAINVLSR.D
1117.5886	1116.5813	1116.5928	-10.24	191	201	0	K.IAFDVAAGPTR.W
1171.6849	1170.6776	1170.6873	-8.30	27	36	1	K.VAVVIPFRDR.G
1366.7649	1365.7576	1365.7728	-11.12	37	48	1	R.GLDPLRLENLAR.V
1637.8885	1636.8812	1636.9009	-12.02	35	48	2	R.DRGLDPLRLENLAR.V
1639.8525	1638.8453	1638.8577	-7.57	138	152	1	R.NREVEPEGLSATPIK.G
1719.8960	1718.8888	1718.8951	-3.70	140	155	1	R.EVEPEGLSATPIKGHR.G
1831.8757	1830.8685	1830.8802	-6.39	202	217	0	R.WVEGSAYHLYHLSGGR.G
1974.9190	1973.9117	1973.9232	-5.80	49	67	0	R.VADHWAGYSGEVIVSGDGR.S
1990.0220	1989.0147	1989.0392	-12.29	138	155	2	R.NREVEPEGLSATPIKGHR.G
2858.3159	2857.3086	2857.3692	-21.19	82	107	0	R.GVTDTDADMLIFAESDLLVSYAQIDR.A
2874.3134	2873.3061	2873.3641	-20.17	82	107	0	R.GVTDTDADMLIFAESDLLVSYAQIDR.A + Oxidation (M)
3016.2383	3015.2310	3015.3253	-31.26	1	26	0	-.MHHHHHHSSGVDLGTENLYFQSNGMK.V + Oxidation (M)

Figure S2 Protein ID results of His-Corndog_gp37

doi:10.6342/NTU202402746

Protein View: gi|2024040303

His-CORNDOG_38

Database:	IBC-114
Score:	108
Expect:	6e-010
Monoisotopic mass (M_r):	25354
Calculated pI:	6.56

Sequence similarity is available as an NCBI BLAST search of gi|2024040303 against nr.

Search parameters

Enzyme:	Trypsin: cuts C-term side of KR unless next residue is P.
Fixed modifications:	Carbamidomethyl (C)
Variable modifications:	Acetyl (Protein N-term), Oxidation (M)
Mass values searched:	20
Mass values matched:	9

Protein sequence coverage: 41%

Matched peptides shown in ***bold red***.

```
  1 MHHHHHHSSG VDLGTENLYF QSNGMTLIGI VADARRDGAG EMLAAKVQAD
 51 YLSIDQGELG CAQNHAKVWR TLAALSEGHS HCVVLEDDAV PVDGFRDQLD
101 AALEAAPAPI VSLYLGRGYI GDRYMSGHIA RADQLDAHWL TSPAIMHAVA
151 LAVRTDLLPG LVTALPSKDQ AIDRTLSLWA RRQGHRVAYT HPSLVDHDDG
201 PSLVSRYKRA ERRAWRVGGR DVWRNKTIAI
```

Unformatted sequence string: 230 residues (for pasting into other applications).

Sort by (•) residue number ( ) increasing mass ( ) decreasing mass
Show (•) matched peptides only ( ) predicted peptides also

Start – End	Observed	Mr(expt)	Mr(calc)	ppm	M	Peptide
47 – 67	2317.1049	2316.0977	2316.0804	7.43	0	K.VQADYLSIDQGELGCAQNHAK.V
118 – 131	1595.7755	1594.7682	1594.7674	0.49	1	R.GYIGDRYMSGHIAR.A
118 – 131	1611.7811	1610.7738	1610.7623	7.10	1	R.GYIGDRYMSGHIAR.A + Oxidation (M)
155 – 174	2123.1969	2122.1896	2122.1634	12.4	1	R.TDLLPGLVTALPSKDQAIDR.T
175 – 181	846.4905	845.4832	845.4759	8.65	0	R.TLSLWAR.R
182 – 186	653.3547	652.3474	652.3517	-6.61	1	R.RQGHR.V
187 – 206	2165.0863	2164.0790	2164.0549	11.1	0	R.VAYTHPSLVDHDDGPSLVSR.Y
209 – 213	687.4163	686.4090	686.3936	22.5	2	K.RAERR.A
210 – 216	944.5228	943.5156	943.5100	5.90	2	R.AERRAWR.V

Index

	Accession	Mass	Score	Description
1.	gi|2024040303	25354	108	His-CORNDOG_38
2.	gi|2024040301	35976	6	HIS-CORNDOG_36
3.	gi|2022012703	35236	6	Sample 3
4.	gi|2022021602	35236	6	WYC_02
5.	gi|20220301	35236	6	D29
6.	gi|2022053101	35236	6	D29_10
7.	gi|20220705	35236	6	Endolysin D29
8.	gi|20220607	73511	5	His-MBP-BXB1_14
9.	gi|2022042102	72945	5	AftB
10.	gi|2022053103	72945	5	AftB mutants

Results List

1. gi|2024040303 **Mass:** 25354 **Score:** 108 **Expect:** 6e-010 **Matches:** 9
His-CORNDOG_38

Observed	Mr(expt)	Mr(calc)	ppm	Start	End	Miss	Peptide
653.3547	652.3474	652.3517	-6.61	182	186	1	R.RQGHR.V
687.4163	686.4090	686.3936	22.5	209	213	2	K.RAERR.A
846.4905	845.4832	845.4759	8.65	175	181	0	R.TLSLWAR.R
944.5228	943.5156	943.5100	5.90	210	216	2	R.AERRAWR.V
1595.7755	1594.7682	1594.7674	0.49	118	131	1	R.GYIGDRYMSGHIAR.A
1611.7811	1610.7738	1610.7623	7.10	118	131	1	R.GYIGDRYMSGHIAR.A + Oxidation (M)
2123.1969	2122.1896	2122.1634	12.4	155	174	1	R.TDLLPGLVTALPSKDQAIDR.T
2165.0863	2164.0790	2164.0549	11.1	187	206	0	R.VAYTHPSLVDHDDGPSLVSR.Y
2317.1049	2316.0977	2316.0804	7.43	47	67	0	K.VQADYLSIDQGELGCAQNHAK.V

No match to: 607.2959, 620.3206, 862.4846, 878.4851, 976.5119, 1046.6097, 1082.5975, 1627.767

Figure S3 Protein ID results of His-Corndog_gp38

doi:10.6342/NTU202402746