Spring 4-8-2021

Characterization of Novel Chlorovirus Glycosyltransferases That Synthesize Atypical Glycans

Eric Noel
University of Nebraska-Lincoln, enoel@huskers.unl.edu

CHARACTERIZATION OF NOVEL CHLOROVIRUS GLYCOSYLTRANSFERASES THAT SYNTHESIZE ATYPICAL GLYCANS

by

Eric A. Noel

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Biological Sciences

(Genetics, Cellular & Molecular Biology)

Under the Supervision of Professor James L. Van Etten

Lincoln, Nebraska

April 2021

CHARACTERIZATION OF NOVEL CHLOROVIRUS GLYCOSYLTRANSFERASES THAT SYNTHESIZE ATYPICAL GLYCANS

Eric A. Noel, Ph.D.

University of Nebraska, 2021

Advisor: James L. Van Etten

Giant dsDNA chloroviruses encode a diverse repertoire of glycosyltransferases (GTases) and methyltransferases (MTases) that biosynthesize unusual, methylated sugars independent of their host *chlorella*-like green algae prompting a reexamination of glycobiology systems. Unlike most other viruses, the prototype chlorovirus PBCV-1 encodes most, if not all, of the machinery required to glycosylate its major capsid protein (MCP). The structures of the four N-linked glycans do not resemble any other glycans in the three domains of life. Here, we investigated the potential involvement of chlorovirus-encoded putative GTases and MTases in glycosylation of the viral MCP. First, we aimed to generate site-directed virus mutants by targeting associated viral genes. We tested transformation methods using cell wall-degrading enzymes, electroporation, SiC whiskers, cell-penetrating peptides, and *Agrobacterium* to generate GT-gene mutations in the chlorovirus CA-4B. We successfully delivered preassembled Cas9 protein-sgRNA ribonucleoproteins (RNPs) to macerozyme-treated NC64A cells that resulted in a frameshift mutation in the CA-4B-encoded gene *034r*, a homolog of PBCV-1 GT gene *a064r*.

Unable to duplicate these results, we shifted our focus to characterize PBCV-1-encoded proteins involved in glycan synthesis. Here, we demonstrated that protein A064R has three functional domains: domain 1 is a β-L-rhamnosyltransferase, domain 2

is an α-L-rhamnosyltransferase, and domain 3 is a MT that methylates the C-2 hydroxyl group of the terminal α-L-rhamnose unit (α-L-Rha). We also established that methylation of the C-3 hydroxyl group of the terminal α-L-Rha is achieved by A061L. Moreover, genetic and structural analyses indicate the protein coded by PBCV-1 gene *a111/114r*, conserved in all chloroviruses, is a GT with three putative domains: galactosyltransferase (domain 1), xylosyltransferase (domain 2), and fucosyltransferase (domain 3). Hydrolytic assays supported these predictions suggesting that A111/114R is likely responsible for the attachment of three of the five conserved residues of the core region of this complex glycan. These findings provide additional support that the chloroviruses do not use the canonical host ER-Golgi glycosylation pathway to glycosylate their glycoproteins; instead, they perform glycosylation independent of cellular organelles using virus-encoded enzymes.

To my dad, Nick Noel, for the past.

To my mom, Cindy Noel, for the past and present.

To my wife, Mary Noel, for the present and future.

ACKNOWLEGEMENTS

First, I must express my sincere gratitude to my advisor, mentor, and friend, Dr. James L. Van Etten. The magnitude of opportunity you provided me as a graduate student cannot be measured or repaid in full value. Your enthusiasm for science is contagious and is only magnified by your warm personality and constant encouragement. It has been my absolute privilege to work beside a world class research scientist, and the Godfather of *Chlorella* viruses. Thank you for personally investing in me and introducing me to the fascinating world of giant viruses. You are an extraordinary person and an admirable professional I aspire to be.

Special thanks to my committee members, Dr. James Van Etten, Dr. Donald Weeks, Dr. Heriberto Cerutti, Dr. Peter Angeletti, and Dr. Eric Weaver for their participation, support and suggestions that greatly improved my research. Thank you, Dr. Van Etten and Dr. Weeks, for serving as readers as well. Thank you, Dr. Charles Wood, the former director of the Nebraska Center for Virology, for your dedication to training young virologists and endorsing NCV as a premier research institution.

My colleagues in the Van Etten lab offered invaluable input that contributed significantly to my dissertation. I cannot thank enough Dr. Irina Agarkova, Dr. David Dunigan, Dr. Gary Duncan, and Dr. Jayadri Sekhar for their influential advice, amusement and friendship. To Dr. Maitham Al-Sammak and Dr. Zeina Al-Ameeli, thank you for treating me as family and your warm company. To Fatima Al-Sammak, thank you for all your help at the benchtop and your patient reception to my barking orders. To Dr. William McClung, it was a sincere joy being office mates on Thursdays. To Rebecca Carlson, the Van Etten lab would crumble without your tenured wizardry. To Roger

Carlson, thank you for being the swiss-army knife of the lab. And Ray Zdanowicz, I'd have a beer with you any day.

Thank you to our "Italian Angels" collaborators across Italy. I strongly value our productive partnership and friendship. Special thanks to Dr. Cristina De Castro and Dr. Immacolata Speciale for their glycochemistry expertise and leadership. Thank you Dr. Michela Tonetti for your consistent input and Dr. Todd Lowary and Sicheng Lin for sharing your beautiful sugar oligomers. To our other collaborators, Dr. Luca Unione, Dr. Domenico Garozzo, Dr. Jesus Jimenez-Barbero, Dr. Elena Laugieri, and Dr. Antonio Molinaro, thank you.

I thank my family, a continuous source of love and encouragement. Mom, thank you for your unwavering support and engraving in me early the values of relentless perseverance. You are the most selfless person I know. As a toddler, you and Dad provided me with everything I needed (and more) to flourish as a student, athlete, and now as an adult. And to my older brother Nick, thank you for your love, support, and friendship. Our closeness is a testament to your grace and patience you've showed with your bothersome younger brother.

I want to express my deepest appreciation to my new family, especially my father and mother-in-law, Mike and Sally Hillis. Thank you for making me feel so loved and welcome in your family. I am so grateful for your advice, your help, and particularly for being so positive about my future. I also thank my new brother-in-law, Charlie, for showing me how to be cool.

Lastly, I want to thank the most important person during my years as a graduate student (again), my wife and best friend, Mary. She has been my endless source of love,

support, and encouragement. With her amazing ability of making me laugh every single day, she has made my days easier and more enjoyable. Thank you for letting me attend AVW9 and marrying me in the same week. And thank you for moving back to Nebraska for me. I know two years became four years, but here we are.

GRANT INFORMATION

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2505060195001. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

This research was also supported by the National Institutes of Health under Ruth L. Kirschstein National Research Service Award 1 T32 AI125207 from the National Institute of Allergy and Infectious Diseases.

PREFACE

Chapter 3 has been published in *Proceedings of the National Academy of Sciences* (Speciale, I.*, Laugieri, M. E.*, Noel, E. A.*, Lin, S., Lowary, T., Molinaro, A., Duncan, G. A., Agarkova, I. V., Garozzo, M. G., Tonetti, A., Van Etten, J. L., & De Castro, C. (2020). Chlorovirus PBCV-1 protein A064R has three of the transferase activities necessary to synthesize its capsid protein N-linked glycans. *Proceedings of the National Academy of Sciences*, 117(46), 28735-28742. doi: 10.1073/pnas.2016626117) *co-first authors

Chapter 4 has been published in *Viruses* (Noel, E. A.*, Notaro, A.*, Speciale, I.*, Duncan, G. A., De Castro, C., Van Etten, J. L. (2020). Chlorovirus PBCV-1 multidomain protein A111/114R encodes three glycosyltransferases involved in the synthesis of atypical N-glycans. *Viruses* 13(1), 87. doi: 10.3390/v13010087) *co-first authors

TABLE OF CONTENTS

ABBREVIATIONS LIST

5mC	5-methylcytosine
6mA	N^6-methyladenine
aa	Amino acids
ADP-RGH	ADP-ribosyl glycohydrolase
APMV	*Acanthamoeba* polyphaga mimivirus
AT	Acetyltransferase (AT)
ATP	Adenosine triphosphate
Ara	Arabinose
Asn	Asparagine
BBM	Bold's basal medium
BLASTp	Basic local alignment search tool protein
CA-4B	*Paramecium bursaria* chlorella virus CA-4B
CAZy	Carbohydrate-active enzymes
CDS	Coding sequence
chs	chitin synthase gene
CRISPR-Cas9	Clustered regularly interspaced short palindromic repeats/CRISPR associated gene 9
Cryo-EM	Cryogenic electron microscopy
DdRp	DNA-dependent RNA polymerase
DSB	Double-strand break
dsDNA	Double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EM	Electron-microscopy
FACS	Fluorescence-activated cell sorting
FRD	Fumarate reductase
Fuc	Fucose
Gal	Galactose
GalNAc	N-acetylgalactosamine
GDP	Guanine diphosphate
GFAT	Glutamine:fructose-6-phosphate amino transferase
GFP	Green fluorescent protein
Glc	Glucose
GlcA	Glucuronic acid
GlcDH	Glucose dehydrogenase
GlcNAc	*β*-1,3-*N*-acetylglucosamine
GMD	GDP-D-Man 4,6 dehydratase
GMER	GDP-4-keto-6-deoxy-D-Man epimerase reductase
GT	Glycosyltransferase
GUS	β-glucuronidase
has	Hyaluronan synthase gene
HDR	Homology directed repair
HGT	Horizontal gene transfer
HMBC	Heteronuclear multiple bond correlation
HPLC	High-pressure liquid chromatography

HSQC	Heteronuclear single quantum correlation
InDel	Insertion-deletion mutation
Kb	Kilobase
kDa	Kilodaltons
LB	Luria-Bertani media
LbCpf1	Lachnospiraceae bacterium Cpf1
LD	Lactate dehydrogenase
Man	Mannose
MBBM	Modified Bold's basal medium
MCP	Major capsid protein
min	Minute
ms	millisecond
MOI	Multiplicity of infection
MRPN	Median reads per-nucleotide
MT	Methyltransferase
NADPH	Nicotinamide adenine dinucleotide phosphate
nm	Nanometers
NC64A	*Chlorella variabilis* NC64A
NCLDV	Nucleocytoplasmic large DNA viruses
NHEJ	Non-homologous end joining
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser effect
nt	Nucleotide
OD	Optical density
PBCV-1	Paramecium bursaria chlorella virus-1
Pbi	*Micractinium conductrix* Pbi
PFU	Plaque-forming units
p.i.	Post infection
Rha	Rhamnose
RNP	Ribonucleoprotein
SAG	*Chlorella heliozoae* SAG 3.83
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
sgRNA	Single guide RNA
STEM	Scanning transmission electron microscopy
SDS/PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SiC	Silicon carbide
SpyCas9	*Streptococcus pyogenes* strain Cas9
ssDNA	Single-stranded DNA
ssODNs	single-stranded oligodeoxynucleotides
Syngen	*Chlorella variabilis* Syngen 2–3
TOCSY	Total correlation spectroscopy
UDP	Uridine diphosphate
VLP	Virus-like particle
Xyl	Xylose

FIGURES LIST

CHAPTER I

INTRODUCTION AND BACKGROUND

1.1. OVERVIEW

This Ph.D. dissertation thesis will introduce chlorovirus-encoded enzymes involved in the manipulation of carbohydrates that has provided many interesting and unexpected findings and concepts to the scientific community over the past 40 years. The aims addressed are twofold. First, to develop a transformation system that supports the genetic modification of the large dsDNA chloroviruses that infect unicellular *Chlorella*-like green microalgae. Specifically, to investigate the progression of chlorovirus glycosylation that takes place independent of its host cell's biosynthetic machinery using CRISPR-Cas9 gene editing. And second, to identify chlorovirus-encoded putative glycosyltransferases (GTases) involved in the unique glycosylation of its major capsid protein (MCP). Investigations of these sugar-manipulating enzymes will provide the field of glycobiology with new discoveries and may ultimately answer the central question about what benefits these viruses have by maintaining their own glycosylation machinery through evolution.

1.2. *PHYCODNAVIRIDAE* FAMILY

Conservative estimates suggest there is somewhere between 100,000 to several million species of algae and that only approximately 40,000 have been identified [1]. Microalgae form the base of the marine food web and their photosynthetic activities provide an important carbon sink that influences the global carbon cycle and even climate. The *Phycodnaviridae* consists of a family of viruses that infect these globally

important players and comprise a genetically diverse [2], yet morphologically similar, family of large icosahedral viruses that infect marine or freshwater eukaryotic algae. These viruses have dsDNA genomes ranging from 160 to 560 kb containing up to ~550 predicted protein-encoding genes (CDSs) [3]. Members of the *Phycodnaviridae* are currently grouped into six genera (named after the hosts they infect): Chlorovirus, Coccolithovirus, Prasinovirus, Prymnesiovirus, Phaeovirus and Raphidovirus. In the broader evolution picture, phylogenetic analyses of numerous chlorovirus genes indicate that the viruses most closely related to the chloroviruses are members of the prasinoviruses. This genus includes viruses that infect the smallest eukaryotic cell *Ostreococcus*, and related species in the class *Mamiellophyceae* [4]. However, the chloroviruses evolved mainly by gene duplications and losses of genes belonging to large paralogous families, whereas *Micromonas* and *Ostreococcus* phycodnaviruses derive most of their genetic novelties though horizontal gene transfers (HGT) [5, 6].

The discovery of phycodnaviruses expanded viral diversity among higher plants given that plants typically are infected by small, plus-stranded RNA containing particles that only encode a few genes [7]. Their genomes exhibit diverse structures, some with large regions of noncoding sequence and others with regions of ssDNA. To date, there are more than 30 species in the *Phycodnaviridae*, divided among 6 genera, and several members have been sequenced [2]. The genome analyses have revealed more than 1000 unique genes, with only 14 homologous genes in common among the three genera of phycodnaviruses sequenced to date [8]. Thus, their gene diversity far exceeds the number of so-called core genes. Not much is known about the replication of these viruses, but the

consequences of these infections on phytoplankton have global affects, including influencing geochemical cycling and weather patterns [9-11].

1.2.A. History of Giant Viruses

The second half of the 20th century was the golden age of virus discovery and most of the 2,000 recognized species of animal, plant, and bacterial viruses were discovered during these years [12]. Nevertheless, it wasn't until the onset of a new century that introduced new infectious agents, enormous by comparison, that would challenge our previous guidelines that defined a virus. Their hallmark characteristics, namely their small size, tiny genomes and parasitic dependence on cellular hosts for replication, set them apart from all other living things despite their animation. However, with the onset of giant virus discovery, new descriptions of viruses have straddled life and nonlife, a divide that usually isn't difficult to navigate. The discovery of giant viruses infecting protists pioneered by the isolation of *Acanthamoeba* polyphaga mimivirus (APMV), is one of the most unexpected and spectacular breakthroughs in virology in decades [13-17]. The giant viruses shatter the textbook definition of viruses as "filterable" infectious agents because their virions do not pass bacterial filters and obliterate all boundaries between viruses and cellular life forms in terms of size and number of genes. Indeed, not only are the particles of giant viruses larger than the cells of numerous bacteria and archaea but also the genomes of pandoraviruses, the current record holders at approximately 2.5 Mb [18], are larger and more diverse in gene content than many bacterial and archaeal genomes, from both parasites and free-living microbes [19]. The recently identified pandoraviruses and pithoviruses [20] are not only observable under light microscope, but also possess a previously unseen, asymmetrical virion

structure. These observations highlight the realization that discovery of the extensive diversity of giant viruses is in its infancy.

Viruses in the family *Phycodnaviridae*, including the chloroviruses, are proposed to have an ancient, common evolutionary ancestry with some other large dsDNA viruses in the *Poxviridae*, *Iridoviridae*, *Ascoviridae*, *Asfarviridae*, *Mimiviridae, Marseillivirdae,* and *Pithoviridae* virus families. However, many more large DNA viruses are rapidly being discovered including pandoraviruses, faustoviruses, mollivirus, kaumoebavirus, cetratvirus, pacmanvirus, and orpheovirus [21] and the evolutionary relationships among these viruses is just starting to be analyzed [22-24]. Collectively, these giant viruses are referred to as nucleocytoplasmic large dsDNA viruses (NCLDVs) [22, 25-27] and it has been proposed that all these viruses should be included in a new order named *Megavirales* [28]. Diverse eukaryotes, such as protozoa, vertebrate animals, and insects, host these viruses. As their name suggests, NCLDVs replicate in both the nucleus and cytoplasm (phycodnaviruses, asfarvavirus, and iridovirus), or only within the cytoplasm (poxviruses). They possess large genomes (100 kb to up to 2.77 Mb) and a diverse gene repertoire that encodes up to 2556 predicted proteins [18]. To put the size of these large viral genomes into perspective, the smallest free-living bacterium, *Mycoplasma genitalium*, encodes 470 proteins [29]. Indeed, their genomes appear somehow closer in size to a typical cell than any other described virus, encoding proteins that have never been previously identified in viruses but have closely related eukaryotic homologs.

Although the low levels of genetic similarity among NCLDVs complicate the precise phylogenetic placement of giant viruses, the relationships between NCLDV families have been reconstructed using multiple conserved genes [24]. Using five genes

that are nearly universal in the NCLDVs, a phylogenetic tree indicates that the NCLDVs form three major branches, with the *Phycodnaviridae* forming one branch along with the *Mimiviridae* and the pandoraviruses [22, 23]. Despite 93% of pandoravirus genes being previously unknown to biologists [18], the phylogenetic relationships of DNA polymerase genes from NCLDVs confidently places pandoraviruses with the family *Phycodnaviridae*, with an apparent specific affinity with coccolithoviruses like *Emiliana huxleyi* virus (Figure 1.1). Therefore, pandoraviruses and phycodnaviruses appear to be related and phylogenetically distinct from the rest of the giant viruses. Also residing here in the same ancestral branch is another genus of phycodnaviruses, chloroviruses, that are the prime focus of this thesis. Although a common evolutionary ancestry of NCLDVs is generally accepted for at least some of the viruses considered to be NCLDVs, disagreements exist regarding the role these viruses played in the evolution of eukaryotes.

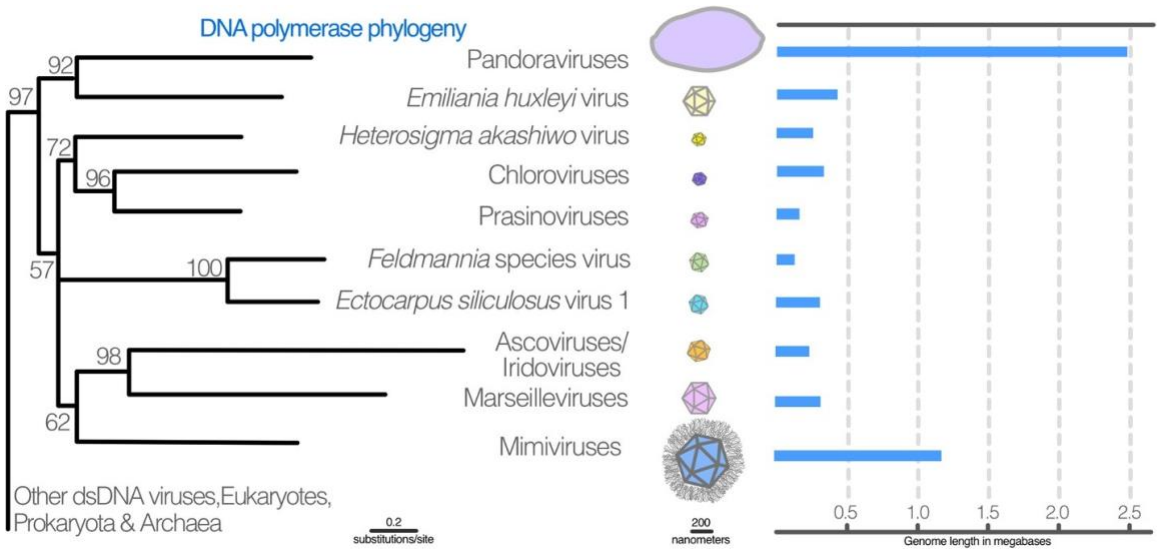

Figure 1.1. Maximum likelihood tree of DNA polymerase with genome length and virion size of various NCLDVs. The maximum likelihood tree shown is a simplified version of the NCLDV subtree for DNA polymerase adapted from [30], where collapsed clades in the original tree are represented by single branches. Numbers at each node are expected likelihood weights from 1,000 rearrangements. The scale throughout the figure is approximate. Figure from [31].

1.2.B. Chlorovirus Discovery

Reports of virus-like particles (VLPs) in at least 44 taxa of eukaryotic algae have appeared between the early 1970s and 1991 [32]. These include incidental observations of VLPs in electron micrographs, for example the first description of VLPs in *E. huxleyi* was reported in 1974 [33]. Although these investigators did not publish an electron micrograph showing VLPs in *E. huxleyi*, they mentioned that VLPs in *Chrysochromulina mantoniae* resembled those of VLPs commonly found in moribund or dead *Coccolithus huxleyi* cells (now referred to as *E. huxleyi*). It was not until 1979

that a phycodnavirus was even isolated; the virus infected the marine unicellular alga *Micromonas pusilla* [34]. However, this report was largely ignored until the early 1990s [35] when high concentrations of viruses in aquatic environments were being described [36]. These early observations were not further pursued by virologists for several reasons that include: (i) evidence for the presence of viruses in algae was solely microscopic and many times it corresponded to a single observation; (ii) in some cases, algae were field-collected and they were no longer available for study; (iii) cultures were not axenic; (iv) technical problems regarding virus isolation (e.g., low concentration); (v) the inability to culture the host [37].

In 1978, Kawakami and Kawakami [38] described VLPs in *Chlorella* isolated from the protozoan *Paramecium bursaria*. However, the virus particles were neither isolated nor characterized. In the early 1980s, a series of landmark discoveries were made by Van Etten and Meints [39-42] when they characterized a large virus with polyhedron structure present in the isolated symbiotic alga *Chlorella* isolated from *Hydra viridis.* They reported that the virus lysed the whole algal population in 24 h and virus particles were absent in endosymbiotic *Chlorella*. Further characterization determined that the virus possessed a 250 kb dsDNA with at least 19 structural proteins, including a 46 kDa major capsid protein (MCP). Six other hydra *Chlorella* viruses were isolated within a short time. Initial restriction fragment analysis of all the viruses that infected endosymbiotic *Chlorella* (now referred to as chloroviruses) revealed four different types of isolates and showed no correlation between genomic composition and geographical origin of the viruses.

One of the six genera in the *Phycodnaviridae*, chloroviruses are icosahedral-shaped dsDNA-containing viruses that infect and replicate in certain strains of unicellular, symbiotic, *chlorella*-like green algae. Chloroviruses are widespread in freshwater environments in all parts of the globe and have been isolated from freshwater sources in Europe, Asia, Australia, as well as North and South America [43]. Chlorovirus titers are variable by season and location, but typically fluctuate between 1 and 100 PFU/mL, although high abundances of up to 100,000 PFU/mL may occur in some environments. Due to the rich genetic diversity and high specialization of individual species with respect to infectious range, variations in their ecology are not unusual, resulting in unique spatio-temporal patterns, which ultimately depend on lifestyle and nature of the host. As such, titers fluctuate during the year with the highest titers typically occurring in the spring, followed by another increase in the late fall while in mid-summer there can be few detectable viruses in the water column [44]. Chlorovirus genome sequences have also been detected in metagenomes from marine environments [45] but no chloroviruses have been isolated from these environments.

Known chlorovirus hosts, which are normally mutualistic endosymbionts and are often referred to as zoochlorellae [46], are associated with the ciliate protozoan *Paramecium bursaria* (Figure 1.2a) [38], the coelenterate *Hydra viridis* [39, 40], or the heliozoon *Acanthocystis turfacea* [47]. Four such zoochlorellae and their viruses are *Chlorella variabilis* NC64A (NC64A) and its viruses (referred to as NC64A viruses), *Chlorella variabilis* Syngen 2–3 and its viruses (referred to as Osy viruses), *Chlorella heliozoae* SAG 3.83 and its viruses (referred to as SAG viruses) and *Micractinium conductrix* Pbi and its viruses (referred to as Pbi viruses). The

zoochlorellae are resistant to virus infection when they are in their symbiotic relationship, because the viruses have no way of reaching their hosts. For example, the zoochlorellae NC64A lives inside the ciliate *P. bursaria* in a special perialgal vacuole (Figure 1.2b) that is surrounded by a host-derived membrane, which protects them from attack by host lytic enzymes. While an individual protist can harbor up to several hundred algal cells at any given time, free-floating algae are highly susceptible to chloroviruses, indicating that such endosymbiosis serves to provide resistance from infection. In this relationship, *P. bursaria* can produce glutamine and Mg^{2+} for the symbiotic algae, and the algae can take glutamine as a nitrogen source and utilize Mg^{2+} for chlorophyll-based photosynthesis [48]. In return, symbiotic algae may provide photosynthetic products, such as fructose, maltose, and oxygen, to host cells. Previous studies also demonstrate existing nutrient trading systems between hosts and symbionts (e.g., O_2, CO_2, maltose and amino acids) [49].

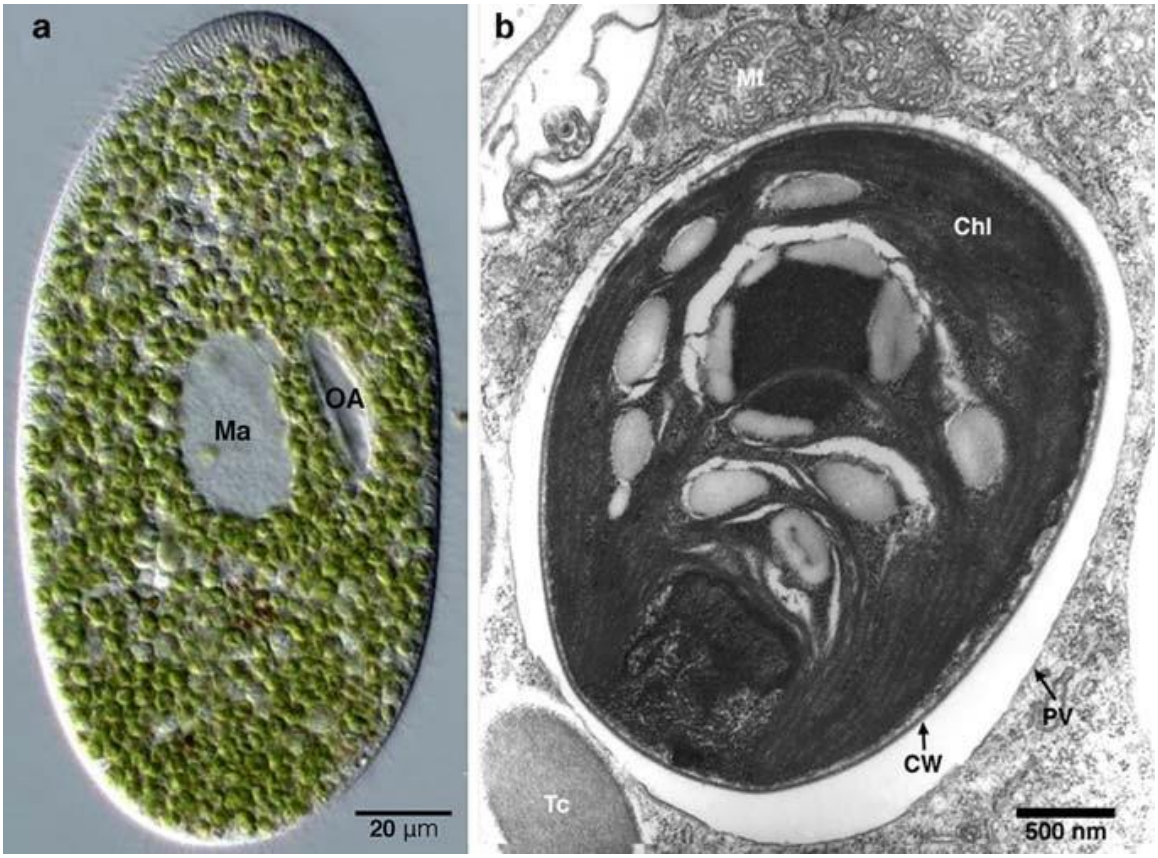

Figure 1.2. Light and transmission electron micrographs of *P. bursaria with endosymbiotic chlorella*. (a) Differential interference contrast (DIC); (b) Transmission electron microscopy (TEM); *Ma,* macronucleus; *OA*, oral apparatus; *Chl, chlorophyll*; *PV,* perialgal vacuole membrane; *CW,* cell wall; *Mt,* mitochondrion; *Tc*, trichocyst. Figure from [50].

A phylogenetic analysis of 32 concatenated conserved genes from 41 chloroviruses that infect three zoochlorellae hosts, i.e., 14 NC64A viruses, 14 Pbi viruses, and 13 SAG viruses, revealed three important features about chlorovirus evolution (Figure 1.3a): (i) viruses infecting the same zoochlorella host clustered in monophyletic clades despite being isolated in diverse locations across five continents; (ii) The

branching pattern of the three main virus clades does not match the phylogeny of the 18S RNA alignment of their zoochlorellae hosts (Figure 1.3b), which eliminates the simplest co-evolution scenario whereby the algae and virus lineages separated at the same time; (iii) The phylogenetic position of virus NE-JV-1 within the Pbi virus clade is the first representative of a previously unknown subgroup of chloroviruses. The most recently discovered chlorovirus group, the Osy viruses, are interesting because phylogenetic analysis of 29 proteins encoded by the 32 conserved concatenated core genes of chlorovirus Osy-NE5 indicates that this virus resides between two separate phylogenetic subclades of NC64A viruses (not included in Figure 1.3) [51].

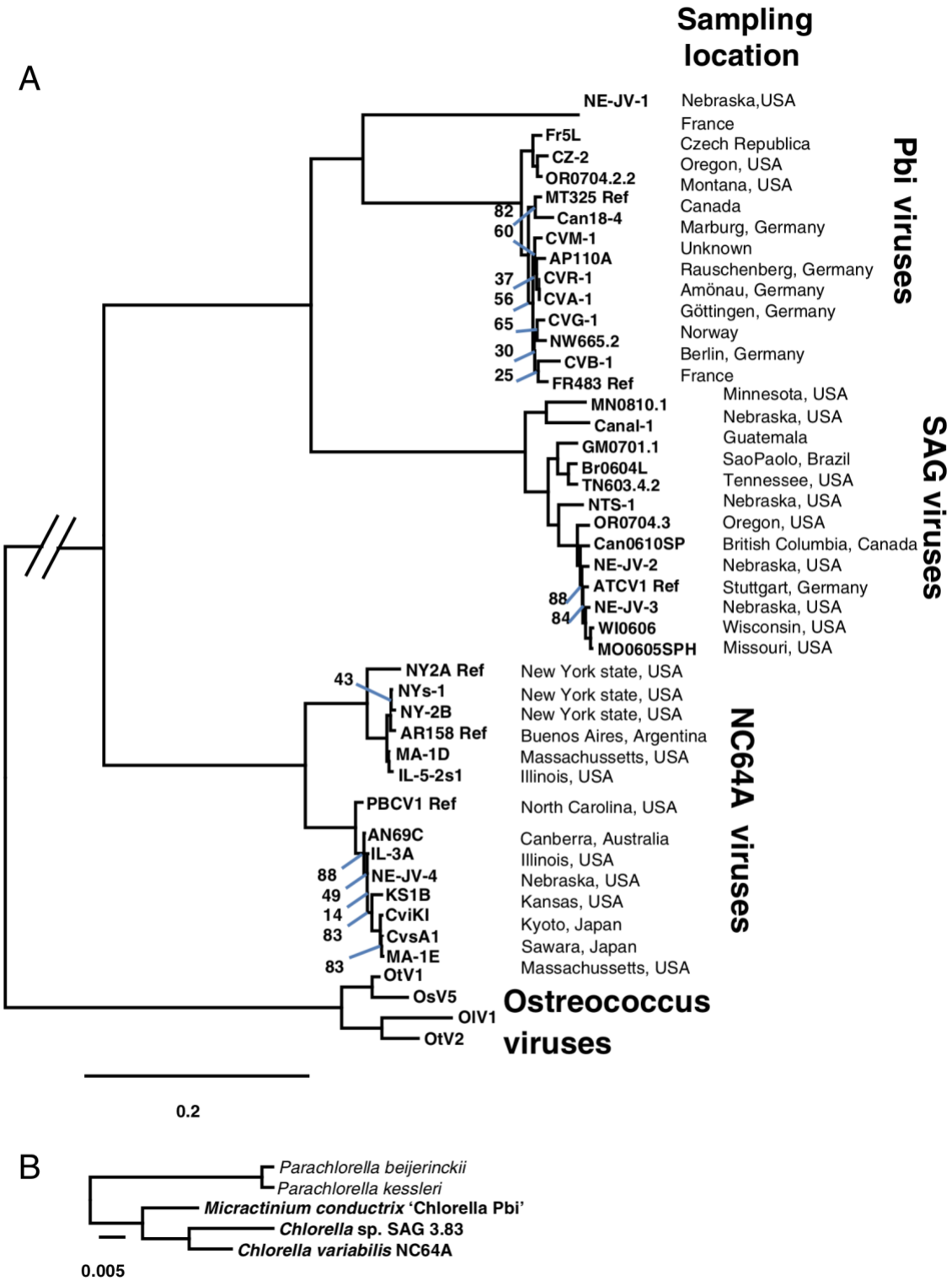

Figure 1.3. Phylogenies of chloroviruses and algal hosts. (a) Phylogenetic tree shows the evolutionary relationships between 47 chloroviruses based on amino acid sequences

encoded by 32 conserved concatenated genes. The Maximum Likelihood tree was constructed using the MEGA 6.0 software with the Maximum Likelihood algorithm and default setting. The bar length represents 0.2 substitutions per amino acid site. A recently characterized Osy chlorovirus species resides between separate phylogenetic subclades of NC64A viruses (not included); Four Ostreococcus virus sequences served as outgroups to root the tree. (b) Maximum Likelihood tree of three algal hosts based on 18S RNA alignment. Figure from [52].

Large amounts of chloroviruses can be produced and assayed by plaque formation on lawns of algae using standard bacteriophage techniques. This provides a sensitive bioassay that is still used today. Attempts to culture the zoochlorellae from hydra to serve as a host for the viruses that infect hydra zoochlorellae were unsuccessful [37]. Given *Chlorella* can be grown independently of their symbiotic partners in the laboratory, permitting plaque assay of the viruses and synchronous infection of their hosts, NC64A virus Paramecium bursaria chlorella virus-1 (PBCV-1) naturally become the prototype virus in the laboratory allowing one to study the life cycle of chloroviruses in detail. Extending behind virological orthodoxy, the chloroviruses, especially PBCV-1 and its host NC64A, are serving as a flexible model system to address concepts of co-evolution at the phenotypic and genotypic levels, ecological dynamics, and predator-prey modeling [53-55].

1.2.C. Prototype chlorovirus Paramecium bursaria chlorella virus-1

Several hundred plaque-forming chloroviruses have been characterized to various degrees. The genomes of 43 chloroviruses infecting four different alga hosts have been

sequenced, assembled and annotated [51, 52, 56-59]. Collectively, the viruses encode genes from 643 predicted protein families; however, any given chlorovirus only has 330 to 416 protein-encoding genes. Thus, the genetic diversity among these viruses is large, and many of the proteins are unexpected for a virus. With the exception of homologs solely in other chlorovirus members, about 50% of their protein-encoding genes do not match anything in the databases. Given the coding capacity of the chloroviruses, it is not surprising that they encode many unusual proteins. Some of these viruses also have introns and inteins, which are rare in viruses. The most studied chlorovirus and the first phycodnavirus to be sequenced was PBCV-1 which infects NC64A [60]. The *C. variabilis* 46.2 Mb genome was sequenced [61] and the availability of both host and virus sequences makes chloroviruses an attractive model system.

Cryo-electron microscopy (cryo-EM) and six-fold symmetry averaging three-dimensional reconstruction of PBCV-1 virions at ~26 Å resolution indicated that the capsid was an icosahedron that is 1900 Å in diameter from point-to-point with a triangulation number of 169d [62]. The capsid covers a single lipid bi-layered envelope, which is required for infection [63, 64]. The trimeric capsomers are arranged into 20 triangular facets (trisymmetrons, each containing 66 capsomers) and 12 pentagonal facets (pentasymmetrons, each containing 30 trimers and one pentamer at each of the icosahedral vertices) (Figure 1.4a). However, as cryo-EM procedures have improved, more details have emerged about the structure of the PBCV-1 virion; the virus is much more complex than the original icosahedral structure suggested.

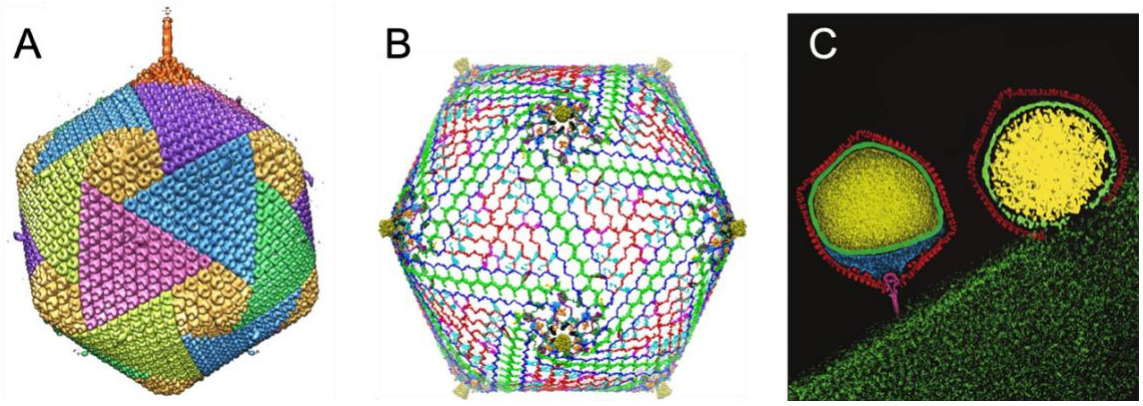

Figure 1.4. Cryo-EM structure of PBCV-1. (a) Hexagonal arrays of major capsomers form trisymmetrons and pentasymmetrons. The unique vertex with its spike structure is at the top. Capsomers in neighboring trisymmetrons are related by a 60° rotation, giving rise to the boundary between trisymmetrons. A fiber extends from one of the capsomers in each of the trisymmetrons. (b) The cryo-EM density (3.5 Å resolution) of PBCV-1 after removing the MCP so that 14 minor capsid proteins are visible. Each protein is shown in a different color. (c) Cross section of a five-fold averaged cryo-EM image of PBCV-1 reveals a long narrow cylindrical spike structure at one vertex and the viral internal membrane (green) surrounding the viral genome asymmetrically (left). Cross section of a five-fold averaged cryo-EM image of PBCV-1 as the virus is ready to release its DNA into the host cell (right). (a) is from [63], (b) is from [65], and (c) is from the cover of J. Virology issue 17, 2012.

Improvement to 8.5 Å resolution and five-fold symmetry 3D reconstruction of cryo-EM images revealed that one PBCV-1 vertex has a 560 Å long spike-structure (Figure 1.4a); 340 Å extends from the surface of the virus [63, 64]. The part of the spike

structure outside the capsid has an external diameter of 35 Å at the tip, expanding to 70 Å at the base. The spike structure widens to 160 Å inside the capsid and forms a closed cavity inside a large pocket between the capsid and the membrane enclosing the virus DNA (Figure 1.4c). Therefore, the internal virus membrane departs from icosahedral symmetry adjacent to the unique vertex (Figure 1.4c). Consequently, the virus DNA is packaged non-uniformly in the particle. (This asymmetric packaging of the genome was reported previously in an ultra-structural study of the Pbi virus CVG-1 [66] and is apparent in other earlier electron microscopy studies). External "fibers" extend from some of the trisymmetron capsomers (one per trisymmetron) and are rigid enough to be resolved with cryo-EM methods; the fibers presumably aid in virus attachment to the host (Figure 1.4a). The fiber-containing capsomer is always located in the middle of the second row of capsomers in the trisymmetron.

Recently, a 3.5 Å near-atomic resolution structure of PBCV-1 by cryo-EM was reported [65]. This resolution was accomplished by correcting for the Ewald sphere effect in single cryo-EM reconstructions [67]. Otherwise, the technology was stalled at ~4.5 Å resolution. The 3.5 Å resolution led to the identification of 14 virus-encoded, minor capsid proteins that form a hexagonal network below the outer capsid shell, stabilizing the capsid by binding neighboring capsomers together (Figure 1.4b). This minor capsid protein shell cements most of the gaps between neighboring capsomers and uses transmembrane helices to mediate inner viral membrane association. The size of the viral capsid is determined by a "tape-measure" minor capsid protein of which there are 60 copies in the virion. Homologs of the tape-measure protein and some of the other minor

capsid proteins exist in other NCLDVs. Thus, a similar capsid assembly pathway might be used by other NCLDVs.

The PBCV-1 genome is a linear ~331-kb, dsDNA molecule with 35 nucleotide-long, covalently closed hairpin termini that exist in one of two forms. The two forms are complementary when the 35-nucleotide sequences are inverted [68]. Identical ~2.2-kb inverted repeats flank each hairpin end [69]. The remainder of the PBCV-1 genome contains primarily single-copy DNA. The GC content of the PBCV-1 genome is ~40%; by contrast, its host nuclear genome is ~67% GC. The PBCV-1 genome was re-sequenced in 2012 to correct mistakes in the original sequence that was reported in the mid-1990s [60]. Using 40 codons as the minimum CDS size and avoiding large overlaps, PBCV-1 contains 416 CDSs (three outermost rings in Figure 1.5). Of the predicted CDSs, ~50% resemble proteins of known function, including many that are novel for a virus. The CDSs are evenly distributed on both DNA strands with minimal intergenic spaces. One exception is a 1,788-nucleotide sequence in the middle of the PBCV-1 genome encoding 11 tRNAs, which is co-transcribed as a large precursor and then processed to mature tRNAs [70].

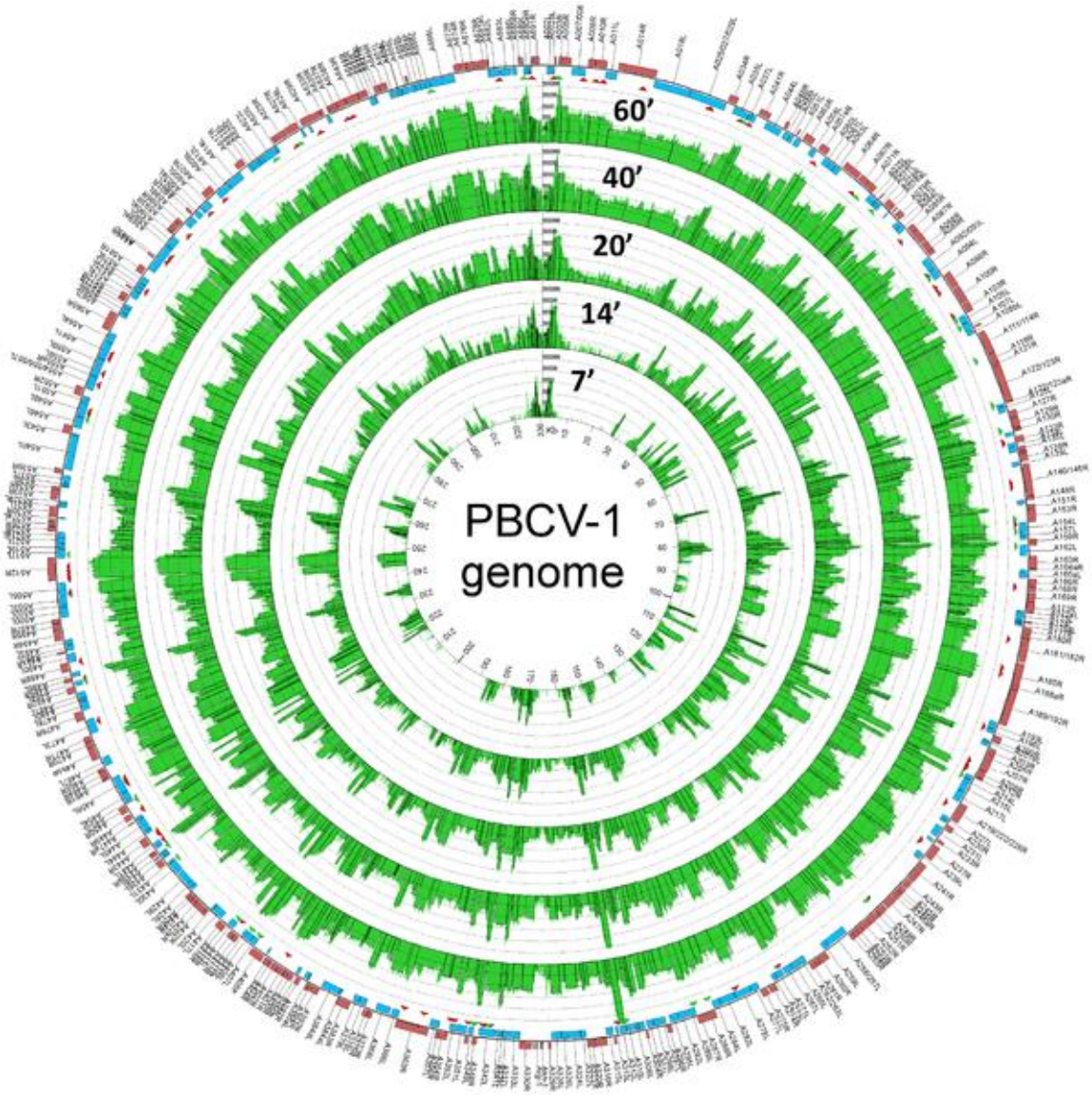

Figure 1.5. Mapping of the PBCV-1 transcriptome. PBCV-1 genes on the forward and reverse strands are depicted by red and blue boxes respectively. The green curves in the interior concentric circles represent the normalized read coverage for each time point (7, 14, 20, 40, and 60 min p.i.) of the experiment in logarithmic scale (base 10). Open boxes are superimposed onto the read coverage curves to show calculated MRPN values for each gene. Below the outer circle, red and green triangles indicate the location of

predicted splice events and polyadenylation cleavage site, respectively. Note: the PBCV-1 genome is a linear molecule with inverted repeats and closed hairpin ends that is depicted as a circle. The two ends are at the 12 o'clock position. Figure from [71].

PBCV-1 and the other chlorovirus genomes contain methylated bases, which occur in specific DNA sequences [72]. Genomes from 37 chloroviruses have 5-methylcytosine (5mC) in amounts ranging from 0.12 to 47.5% of the total cytosines. In addition, 24 of the 37 viral DNAs contain N^6-methyladenine (6mA) in amounts ranging from 1.5 to 37% of the total adenines [32]. The methylated bases occur in specific DNA sequences, which led to the discovery that chloroviruses encode multiple 5mC and 6mA DNA methyltransferases (MTases). Interestingly, the methylated bases are part of the virus-encoded DNA restriction and DNA modification systems [73]. About 25% of the virus-encoded DNA MTases have companion DNA site-specific (restriction) endonucleases, including some with unique cleavage specificities (2 to 4 bp) that are sold commercially [74, 75]. Prior to the discovery that the chloroviruses encoded enzymes with DNA restriction endonuclease activity, all known DNA restriction endonucleases were from bacteria. The purpose of the other virus-encoded DNA MTases is unknown. One possibility is that they were originally associated with companion DNA restriction endonucleases that have subsequently been lost in evolution or serve some other unknown physiological function.

The sum total of chlorovirus-encoded CDSs from 45 analyzed viruses (NC64A, Pbi, SAG viruses) include ~650 protein families. Any one virus encodes no more than ~400 CDSs, thus the chloroviruses have a lot of genetic diversity. Furthermore, some

chlorovirus-encoded CDSs have as many as three distinct functional domains, and each domain encodes an independent enzyme activity. Therefore, the genetic diversity of the chloroviruses is much larger than the total number of genes. Why chloroviruses need this large amount of gene diversity and the origins of many of these genes is unknown.

Three-dimensional reconstruction of PBCV-1 in the presence of NC64A cell walls supports the hypothesis that the virus spike structure first contacts the cell wall (Figure 1.4c) [76] and that fibers appear to aid in holding the virus to the wall. The PBCV-1 protein A140/145R, referred to as Vp130, was reported to be involved in recognizing the host receptor [77] and to be located at a unique vertex [78]. The virus spike structure is too narrow to deliver DNA into the host and likely serves to puncture the wall using a virus-associated enzyme(s) (Figure 1.6a) before it is pushed aside. PBCV-1 attachment to its host receptor is rapid and specific with an adsorption rate of 5×10^{-9}/mL/min [79]. The early stages of PBCV-1 attachment are reversible [80]. The inability to attach to non-host cells is a major factor in limiting chloroviruses' host range. The identity of the host receptor, which is uniformly present over the entire surface of the alga, is unknown but circumstantial evidence suggests that it is probably a carbohydrate [81].

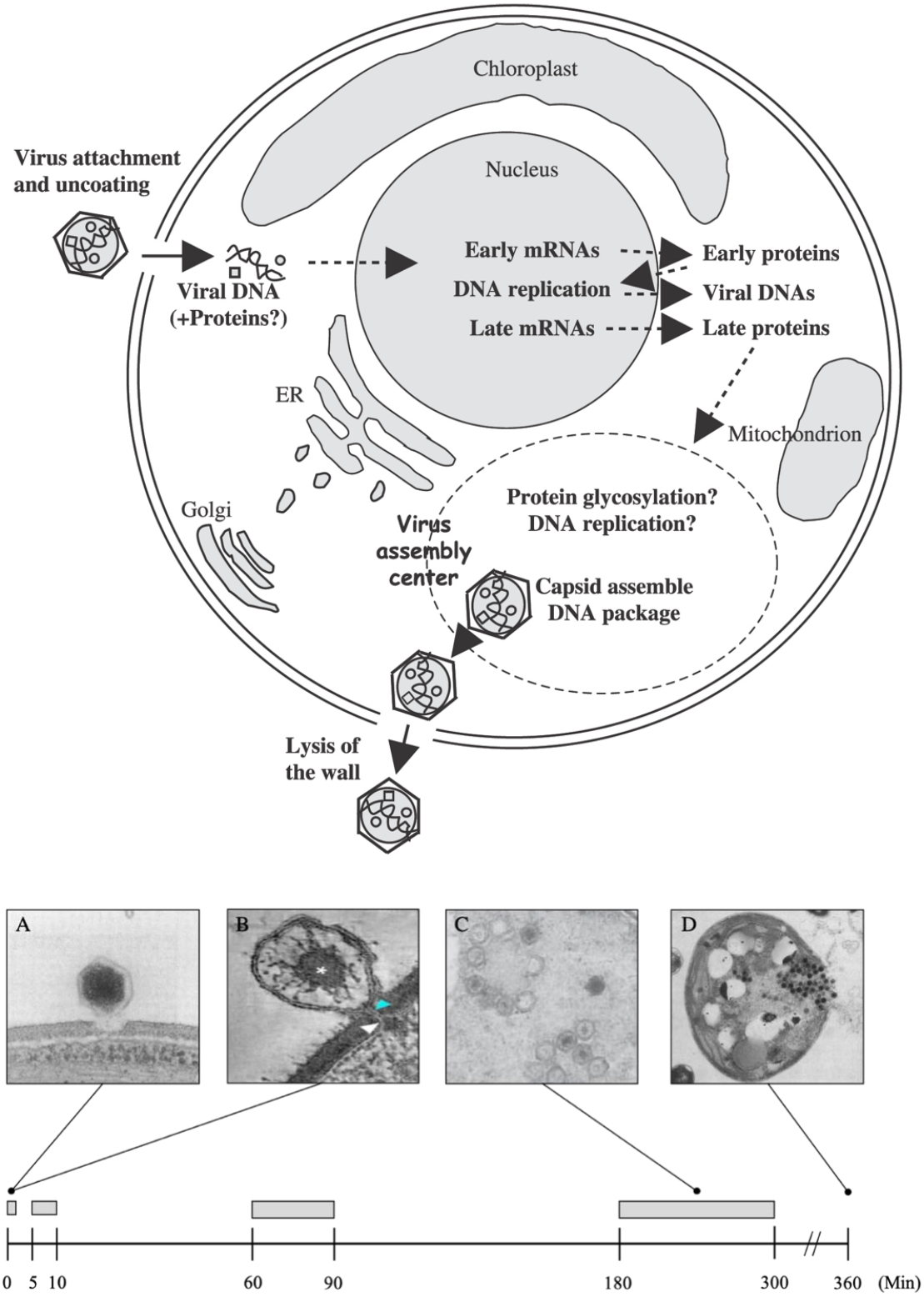
Chloroplast
Nucleus
Virus attachment
and uncoating
Viral DNA
(+Proteins?)
Early mRNAs
DNA replication
Late mRNAs
Early proteins
Viral DNAs
Late proteins
ER
Mitochondrion
Golgi
Protein glycosylation?
DNA replication?
Virus
assembly
center
Capsid assemble
DNA package
Lysis of
the wall
A
B
C
D
0 5 10 60 90 180 300 360 (Min)

Figure 1.6. Proposed replication cycle of PBCV-1 (*top*) and infection timeline (*bottom*). Attachment of PBCV-1 to the algal wall and digestion of the wall at the point of attachment (a). STEM tomograph showing the close proximity between the viral and host internal membrane forming a membrane-lined tunnel between the virus and its host (b). The viral DNA moves to the nucleus where early gene transcription begins at 5 to 10 min p.i. Early mRNAs move to the cytoplasm for translation, and at least some early proteins presumably return to the viral genome located at or near to the nucleus to initiate viral DNA replication, which begins ~60 p.i., followed by late gene transcription. Late mRNAs move to the cytoplasm for translation and many of these late proteins are targeted to the virus assembly centers (c), where virus capsids are formed, and DNA is packaged. Mature infectious virus particles virus particles appear in the cytoplasm of the cell ~45 min prior to virus release. The *Chlorella* cell membrane and wall lyses (d), and infectious PBCV-1 progeny viruses are released at 6 to 8 h p.i. (→) Known events; (- ->) hypothesized events. EM micrographs by Kit Lee. Figure adapted from [82, 83].

PBCV-1 encodes five proteins that degrade potential cell wall polysaccharides including two chitinases [84], a chitosanase [84], a β-1,3 glucanase [85], and an alginate lyase-like enzyme [86-88]. Following host cell wall degradation, the virus internal membrane fuses with the host membrane (Figure 1.6b), forming a membrane-lined tunnel between the virus and its host [83], leaving an empty capsid attached to the surface. The virus-host membrane fusion process triggers rapid depolarization of the host plasma membrane resulting in the release of K^+ from the cell, presumably initiated by a virus encoded K^+ channel that is located in the virus internal membrane [89]. The

depolarization of the host's cell membrane is also thought to prevent secondary infection from another virus or secondary transporters. The rapid loss of K^+ from the host and associated water fluxes significantly reduce the host turgor pressure, which aids ejection of viral DNA and virion-associated proteins into the host (Figure 1.6b) [90].

Circumstantial evidence plus microscopic images [83] indicates that the viral DNA and suspected DNA-associated proteins quickly move to the nucleus taking swift command of the host transcription machinery. Given none of the chloroviruses encode a recognizable DNA-dependent RNA polymerase (DdRp) gene, its rapid transit to the host nucleus supports these observations. RNAseq experiments revealed that ~50 PBCV-1 genes are expressed within the first 7 min p.i. [71]. By 60 min p.i., essentially all of the PBCV-1 genes are expressed at some level and ~40% of the poly (A+) containing RNAs in the infected cell are PBCV-1 transcripts. This rapid increase in viral mRNAs probably involves increased viral transcription together with the selective degradation of host mRNAs within 5 min p.i., presumably by the PBCV-1 encoded and packaged DNA restriction endonucleases. Further, host chromatin remodeling is probably involved because PBCV-1 also encodes and packages an enzyme (referred to as vSET) that tri-methylates Lys-27 in histone 3 [91] which likely suppresses host transcription. Not surprisingly, transcription of host genes is altered within the first 7 min after PBCV-1 infection [92]. This reversal in transcription control from host to virus implies that some component(s) must facilitate active transport of the virus genome to the nucleus.

Presumably virus DNA synthesis begins in the nucleus before moving to the cytoplasm. By 4 h p.i., the amount of virus DNA in the cell is at least four times higher than the DNA in the cells at the time of virus infection. At 3 to 4 h p.i., assembly of

PBCV-1 capsomers begins in localized regions of the cytoplasm, which become prominent 4 to 5 h p.i. (Figure 1.6c). These localized regions, called virus assembly centers or virus factories, consist of host cisternae that are derived from the ER next to the nuclear membrane [93]. By 5 to 6 h p.i. infectious PBCV-1 progeny accumulates in the cytoplasm and localized lysis of the host cell membrane and cell wall releases progeny at 6 to 8 h p.i. (Figure 1.6d). Each infected algal cell releases up to ~1000 particles, of which ~25% form plaques. Mechanical disruption of cells releases infectious virus particles 30 to 50 min prior to cell lysis, indicating the virus is mature in the cell and that it does not acquire its glycoprotein capsid by budding through the host plasma membrane as it is released from the cell [94]. Currently nothing is known about the clock mechanism that must operate to time the release of the nascent virions; however, the viruses encode several cell wall-degrading enzymes that presumably function for the virus to escape the cell. A schematic diagram of the predicted PBCV-1 replication cycle is reported in Figure 6.

1.3. CARBOHYDRATE MANIPULATING GENES IN CHLOROVIRUSES

Chloroviruses are unusual because they encode enzymes involved in manipulating carbohydrates [95-97]. These include enzymes involved in making extracellular polysaccharides, such as hyaluronan and chitin, enzymes that make nucleotide sugars, such as GDP-L-fucose (GDP-L-Fuc) and GDP-D-rhamnose (GDP-D-Rha) and enzymes involved in the synthesis of glycans attached to the virus MCPs. This latter process differs from that of all other glycoprotein containing viruses that traditionally use the host endoplasmic reticulum (ER) and Golgi machinery to synthesize and transfer the glycans. Van Etten *et al.* (2017) [97] listed putative chlorovirus genes

involved in carbohydrate metabolism, which are encoded by the 43 chloroviruses whose genomes have been sequenced, in Tables 1.1-1.3. Recombinant proteins have been produced from some of these genes, and the proteins have been characterized (indicated in bold in the tables). When some of the genes were initially cloned and the recombinant proteins characterized, the genes were hybridized to many other chlorovirus genomes by dot blots to determine the distribution of the genes.

Table 1.1.

Chlorovirus encoded enzymes involved in the synthesis of polysaccharides

Host	Viruses	HAS [1]	CHS [2]	CBP [3]
NC64A	AN69C		390R	395L, 438L
	AR158		C418R	C423L, C475L
	CviK1	102R	365R	370L, 414L
	CvsA1		375R	380L, 427L
	IL-3A			386L, 432L
	IL-5-2s1	134L		503L, 562L
	KS-1B			314L, 360L
	MA-1D	485L	362R	367L, 472L
	MA-1E		113L	407L, 451L
	NE-JV-4			390L
	NY-2B			542L, 484L
	NY2A		B139R, B472R	B480L
	NYs-1	137R		360R, 495L, 555L
	PBCV-1	[a] **A098R**		A333L, A348Dl, A383R
SYN	OSY-NE-5	038R		167L, 184L
Pbi	AP110A		152L, 175R	828R
	CVA-1		150L, 169R	834R
	CVB-1		177L	791R
	CVG-1	146R		792R
	CVM-1		165L, 186R	832R
	CVR-1			838R
	CZ-2			798R
	Can18-4	163R		839R
	FR483	N124R		N690R
	FR5L		151L	797R
	MT325	M128R		M701R
	NE-JV-1		278R, 282R	734L
	NW665	133R		821R
	OR0704.2.2	116R		804R
SAG	ATCV-1			Z734R
	Br0604L		834R	431R
	Can0610SP			438R, 442R
	Canal-1		746R	405R
	GM0701		852R	436R
	MN0810		087R, 900R	466R, 468R, 531L
	MO0605SPH			435R
	NE-JV-2			462R
	NE-jv-3			431R
	NTS-1		893R	461R, 463R, 529L
	OR0704.3			431R
	TN603		869R	425R
	W10606			457R

[1] Hyaluronan synthase; [2] chitin synthase; [3] chitin binding proteins, except for the chitinase proteins involved in degrading polysaccharides; [a] the recombinant protein has the predicted activity. The numbers refer to the protein names, and the R and L refer to the strand orientation. From Van Etten *et al.* Viruses (2017).

Table 1.2.

Chlorovirus encoded enzymes involved in sugar metabolism

Host	Viruses	GFAT [1]	UDP-GlcDH [2]	GMD [3]	GMER [4]	UGD [5]	AT [6]	D-LD [7]	ADP-RGH [8]	FRD [9]
NC64A	AN69C	109R	384R, 487R	129R	334L		739L	055R		
	AR158	C132R	C413R, C729L	C155R	C344L		C767L			
	CviK1	105R	359R, 662L	122R	312L		742L	055R		
	CvsA1	064R	368R	083R	321L		716L	131R		
	IL-3A	104R	375R, 685L	126R	326L		726L	051R		
	IL-5-2s1	130R	492R, 858L	106L	417L		896L			
	KS-1B			066R	261L		643L			
	MA-1D	481L	355R, 838L	456L	284L		872L			
	MA-1E	116R	396R	134R	355L		806L	045R		
	NE-JV-4			131R	340L		740L	064R		
	NY-2B		473R, 836L	185R	408L		881L			
	NY2A	B143R	B465R	B163R	B395L		B853L			
	NYs-1	143R	483R, 846L	167R	404L		879L			
	PBCV-1	[a] **A100R**	[a] **A609L**	[a] **A118R**	[a] **A295L**		A654L	A053R		
SYN	OSY-NE-5	039R		045R	139L		340L	015R	308R	
Pbi	AP110A	071R	146L				893R	053L		
	CVA-1	056R	144L				900R	040L	205R	
	CVB-1	071R	172L				856R	056L	221R	
	CVG-1	050R	812L				857R			
	CVM-1	069R	159L				893R	052L	222R	
	CVR-1	062R	151L				906R	046L	210R	
	CZ-2	059R	718L				865R	048L		917L
	Can18-4	061R	859L				908R	048L	212R	
	FR483	N035R	N712L				N747R		N170R	
	FR5L	087R	145L				863R	076L		
	MT325	M036R, M037R	M719L				M758R	M026L		
	NE-JV-1	081R	291R				861R	810L		
	NW665	046R	846L				889R		189R	
	OR0704.2.2	062R	722L				862R	051L		

Table 1.2. *Cont*.

Host	Viruses	GFAT [1]	UDP-GlcDH [2]	GMD [3]	GMER [4]	UGD [5]	AT [6]	D-LD [7]	ADP-RGH [8]	FRD [9]
	ATCV-1		Z571L	[a] **Z804L**	Z282L	**Z544R**	Z147L	Z295L		
	Br0604L		667L, 839R	934L	332L	631R	173L	350L		
	Can0610SP		687L	965L	338L	658R	170L	355L		
	Canal-1		605L, 751R	847L	329L	576R	188L	343L	898L	886R
	GM0701		664L, 856R	954L	337L	629R	180L	354L		
	MN0810	887L	720L, 904R	992L	365L	689R	204L	379L		
SAG	MO0605SPH		656L	897L	341L	625R	181L	359L		943R
	NE-JV-2		708L	981L	367L	672R	186L	383L		
	NE-JV-3		679L	935L	332L	648R	175L	347L		981R
	NTS-1		714L, 898R	1012L	378L	681R	188L	391L		
	OR0704.3		676L	960L	335L	639R	179L	354L		
	TN603		659L, 873R	966L	326L	625R	179L	342L		
	W10606		679L	916L	360L	651R	185L	375L		962R

[1] Glutamine-fructose-6-phosphate aminotransferase; [2] UDP-glucose-6-dehydrogenase; [3] GDP-D-mannose dehydratase; [4] GDP-4-keto-6-deoxy-D-mannose epimerase/reductase (GDP-L-fucose synthase 2); [5] UDP-D-glucose 4,6-dehydratase, 6 acetyltransferase; [7] D-lactate dehydrogenase; [8] ADP-ribosylglycohydrolase; [9] fumarate reductase; [a] the recombinant proteins have the predicted activities. The numbers refer to the protein names, and the R and L refer to the strand orientation. From Van Etten *et al*. Viruses (2017).

Table 1.3.

Chlorovirus encoded enzymes involved in the synthesizing glycans attached to virus major capsid proteins

Host	Viruses	EXT [1]	GT-A [2]	GT-GT4 [3]	GT-GTA [4]	CESA CelA-Like [5]	CSCS-2 [6]	GT [7]	GT17 [8]
NC64A	AN69C	078L	123R	559R	065R	104R	255R		
	AR158	C093L	C150R	C661L			C265R		C559R
	CviK1	080L	117R	594L		518L	242R		
	CvsA1	039L	077R	611L		535L	247R		
	IL-3A	071L	120R	606L	060R	099R	240R		
	IL-5-2s1	175R	109L	773L			313R		649R
	KS-1B	024L	060R	528L	009R	046L	170R		
	MA-1D	531R	4549L	753L			194R		637R
	MA-1E	533R	128R	702L		626L	281R, 210R		
	NE-JV-4	085L	124R	631L	074R	108L	250R		
	NY-2B	116L	180R	754L		160R	323R		633R
	NY2A	107L	B159R	B736L					B618R
	NYs-1	098L	162R	760L			187R		641R
	PBCV-1	A075L	A111/114R	A546L	A064R	A473L	A219/222/226R		
SYN	OSY-NE-5	025L	044R	283L			097L		
Pbi	AP110A	013L	548R			226R		970R	
	CVA-1	016L	532R			220R		971R	
	CVB-1	025L	538R			232R	811L	918R	
	CVG-1	019L	520R			217R	815L	920R	
	CVM-1	022L	550R			237R		953R	
	CVR-1	020L	545R			225R		977R	
	CZ-2	012L	532R	380L			822L	932R	
	Can18-4	020L	557R			229R	862L	971R	
	FR483	N012L	N472R			N191R	N715L	N805R	
	FR5L	046L	537R				819L	926R	
	MT325	M009L	M467R			M186R	M721L	M813R	
	NE-JV-1	079L	464L				801R	930R	
	NW665	015L	532R				849L	955R	
	OR0704.2.2	017L	549R	382L			823L	923R	

Table 1.3. *Cont.*

Host	Viruses	EXT [1]	GT-A [2]	GT-GT4 [3]	GT-GTA [4]	CESA CelA-Like [5]	CSCS-2 [6]	GT [7]	GT17 [8]
	ATCV-1	Z830R	Z120R	Z667L		Z178L, Z823R, Z417L	Z425R	Z347R	
	Br0604L	959R	137R			225L, 952R, 483L	489R	399R	
	Can0610SP	1007R	140R	789L		210L, 1002R, 978L, 487L	495R	407R	
	Canal-1	874R	141R	164L		871R	447R	380R	
	GM0701	977R	141R	752L		228L, 975R, 486L	493R	405R	
	MN0810	1009R	165R			244L		424R	
SAG	MO0605SPH	932R	138R	164L		230L, 926R, 479L	488R	412R	
	NE-JV-2	1020R	150R	804L		1015R, 992L, 516L	523R	438R	
	NE-JV-3	970R	145R	778L		215L, 210L, 964R, 484L	493R	407R	
	NTS-1	1044R	156R			1016L, 516L		441R	
	OR0704.3	1006R	146R	944L		1001R, 972L, 485L, 219L	493R	404R	
	TN603	991R	141R			226L, 986R, 475L	483R	400R	
	WI0606	951R	141R			233L, 945R, 506L	514R	430R	

[1] Exotosin glycosyltransferase; [2] glycosyltransferase family A; [3] glycosyltransferase GT4-type super family; [4] glycosyltransferase GTA-type super family; [5] CESA CelA-like cellulose synthase catalytic subunit (UDP-glucose as substrate); [6] cellulose synthase catalytic subunit; [7] glycosyltransferase; [8] glycosyltransferase family 17. From Van Etten *et al.* Viruses (2017).

1.3.A. Polysaccharide Synthesizing Enzymes

Three PBCV-1-encoded enzymes, UDP-glucose dehydrogenase (UDP-GlcDH, gene *a609l*) [98], glutamine:fructose-6-phosphate amino transferase (GFAT, gene *a100r*) [98], and hyaluronan synthase (*has* gene) [99] contribute to the synthesis of hyaluronan (hyaluronic acid), a linear polysaccharide composed of alternating *β*-1,4-glucuronic acid (GlcA) and *β*-1,3-*N*-acetylglucosamine (GlcNAc) residues. UDP-GlcDH and GFAT are involved in the biosynthesis of hyaluronan precursors. Until the *has* gene (*a098r*) was discovered in PBCV-1, hyaluronan was only thought to occur in vertebrates and a few pathogenic bacteria, where it forms an extracellular capsule, presumably to avoid the immune system [100, 101]. Hyaluronan is an essential constituent of the extracellular matrix in vertebrates and consists of ~10,000 or more alternating GlcA and GlcNAc residues. Typically, the HAS enzyme is located on the inner surface of the plasma membrane. The newly-synthesized hyaluronan then moves through the membrane and cell wall to the extracellular matrix. The identification of these three genes, transcribed early in virus infection, led to the discovery that hyaluronan lyase-sensitive hair-like fibers begin to accumulate on the surface of PBCV-1-infected host cells by 15 min p.i. By 4 h PI, the infected cells are covered with a dense fibrous hyaluronan network [102].

The *has* gene is present in about 30% of the chloroviruses. Surprisingly, many chloroviruses that lack a *has* gene have a gene encoding a functional chitin synthase (*chs*) gene. Furthermore, cells infected with these viruses produce chitin fibers on their external surface [103]. Chitin, an insoluble linear homopolymer of *β*-1,4-linked GlcNAc residues, is a common component of insect exoskeletons, shells of crustaceans, and fungal cell walls. A few chloroviruses encode both *chs* and *has* genes and form both chitin and

hyaluronan on the surface of their infected cells [103, 104]. Finally, some chloroviruses lack both genes and cells infected with these viruses produce no known extracellular polysaccharides [102]. The fact that many chloroviruses encode enzymes involved in extracellular polysaccharide biosynthesis, which require huge amounts of ATP for their synthesis, suggests that the polysaccharides are important in the viral life cycles. However, they do not appear to be required for chlorovirus replication in the laboratory. Furthermore, because all of the chloroviruses, including ones that lack the *has* and *chs* genes, have been isolated from natural sources in recent years, the genes do not appear to be essential for replication in native environments. Currently, it is not known how or why the chloroviruses acquired these polysaccharide-synthesizing genes.

1.3.B. Nucleotide Sugar Metabolism Enzymes

Many chloroviruses also encode enzymes involved in nucleotide sugar metabolism, as well as other sugar metabolic enzymes (Table 1.2). Two enzymes involved in synthesizing GDP-L-Fuc are encoded by all of the NC64A, SAG and Syn chloroviruses, GDP-D-mannose (GDP-D-Man) 4,6 dehydratase (GMD) and GDP-4-keto-6-deoxy-D-Man epimerase reductase (GMER, also referred to as Fuc synthase) (Table 1.2), comprise a highly-conserved three-step pathway in bacteria, plants and animals that converts GDP-D-Man to GDP-L-Fuc [105]. In vitro reconstruction of the pathway using recombinant PBCV-1 GMD and GMER proteins resulted in the synthesis of GDP-L-Fuc as expected. Unexpectantly, PBCV-1 GMD is unusual because it can also convert the GMD product into GDP-D-Rha in the presence of NADPH, that is, the enzyme performs two activities that produces both sugars inside the infected cell [105]. Interestingly, the chlorovirus ATCV-1 GMD homolog lacks the GDP-Rha forming activity [106]. Both

Fuc and Rha are in the glycans attached to the PBCV-1 MCP. The Pbi chloroviruses lack both *gmd* and *gmer* genes even though the glycans attached to the MCP from the three evaluated Pbi viruses have Fuc [107]. Other chloroviruses encode additional sugar metabolizing enzymes including a UDP-D-Glc-4,6-dehydratase [108] and a putative Man-6-phosphate isomerase.

Besides the chlorovirus-encoded enzymes described above, the viruses have four additional genes predicted to encode enzymes involved in sugar metabolism (Table 1.2). Recombinant proteins have not been produced from any of these genes, and so, it is unknown if they encode functional enzymes. These putative enzymes include an acetyltransferase (AT) encoded by all 43 chloroviruses, a D-lactate dehydrogenase (D-LD) encoded by 32 chloroviruses, fumarate reductase (FRD) encoded by five chloroviruses and ADP-ribosyl glycohydrolase (ADP-RGH) encoded by nine chloroviruses, all but two of which are Pbi viruses. The roles these putative enzymes play in the viral life cycles are unknown.

1.3.C. Unusual Attachment of Glycans to the Chlorovirus Major Capsid Proteins

Structural proteins of many viruses, such as rhabdoviruses, herpesviruses, poxviruses and paramyxoviruses, are glycosylated. Glycans contribute to the protease resistance and the antigenicity of these viruses. Most virus glycans are linked to Asn in the protein via GlcNAc, although some viruses also have O-linked glycans attached to either Ser or Thr residues via an amino sugar, usually GlcNAc or *N*-acetylgalactosamine (GalNAc). Typically, viruses use host-encoded GTases and glycosidases located in the ER and Golgi apparatus to add and remove N-linked sugar residues from virus glycoproteins either co-translationally or shortly after translation of the protein. The virus

glycoproteins are then transported to the host plasma-membrane where progeny viruses bud though the membrane and they only become infectious as they leave the cell [109]. However, unlike the process described above, glycosylation of the chlorovirus MCPs differs from that scenario because the viruses encode most, if not all, of the machinery for the process. Furthermore, the glycosylation process occurs in the cytoplasm [96, 97].

The conclusion that the chlorovirus PBCV-1 MCP (Vp54, gene *a430l*) is glycosylated by a different mechanism than that used by other characterized viruses originally arose from antibody studies [110]. Rabbit polyclonal antiserum [111] particles inhibited virus plaque formation by agglutinating the virions. Spontaneously derived antiserum-resistant, plaque-forming variants of PBCV-1 occur at a frequency of 10^{-5} to 10^{-6}. Originally, four serologically distinct classes were identified plus wild type virus; subsequently two additional antigenic variants have been isolated for a total of six antigenic variants. Polyclonal antisera prepared against members of each of these antigenic classes react predominately with the Vp54 protein equivalents from the viruses in the class used for the immunization. Each of the Vp54 proteins from the antigenic variants migrated faster on SDS-PAGE than those from the strains from which they were derived, indicating a lower molecular weight. However, all of the de-glycosylated MCPs migrated at the same rate on SDS-PAGE, indicating that the size differences were due to the attached glycans. In addition, the *a430l* nucleotide sequence in each of the antigenic variants was identical to the wild-type gene; these results verified that the polypeptide portion of Vp54 was not altered in the antigenic variants. Western blot analyses of Vp54 proteins before and after removing the glycans with trifluoromethane-sulfonic acid or

altering the glycan with periodic acid, also supported the notion that the antigenic variants reflected differences in the Vp54 glycans, not the Vp54 polypeptide [110].

Additional observations supported the concept that PBCV-1 Vp54 glycosylation was unusual: (1) unlike viruses that acquire their glycoproteins(s) by budding through a plasma membrane, which results in infectious particles, plaque-forming PBCV-1 particles accumulate inside the host 30–40 min before virus release [112]; (2) all of the antigenic variants were grown in the same host so the glycan differences are not due to the host; (3) polyclonal antibodies to Vp54, the MCP, do not react with host glycoproteins; (4) the PBCV-1 Vp54 protein, like the PBCV-1 encoded GTases, lacks an ER and Golgi signal peptide; (5) unlike most glycoproteins that exhibit size micro-heterogeneity, PBCV-1 Vp54 appears homogeneous on SDS-PAGE; in addition, mass spectrometry analysis only revealed one satellite peak that differed from the main peak by 140 Da, the approximate weight of either one arabinose (Ara) or xylose (Xyl) residue [113]; and (6) the ability to easily crystallize Vp54 as a homotrimer provided additional evidence that the protein is essentially homogeneous [114, 115].

Evidence that the N-linked Vp54 glycans are not attached to the Vp54 protein by a N-linkage was initially obtained from the X-ray crystal structure of the protein [115]. The structure revealed that the protein had four N-linked glycans at Asn positions 280, 302, 399 and 406 [114]. None of these Asn were located in an Asn-X-(Thr/Ser) sequon sequence commonly recognized by ER located GTases [116]. This finding also explained why prior attempts to remove Vp54 glycans with enzymes that cleave traditional N-linked glycans were unsuccessful. Nandhagopal *et al.* [114] also reported that Vp54 contained two O-linked glycans. However, re-examination of the X-ray crystal data

(Figure 1.7) indicated that no O-linked glycans were present in the protein [115], which agrees with our unsuccessful attempts to detect them by chemical procedures.

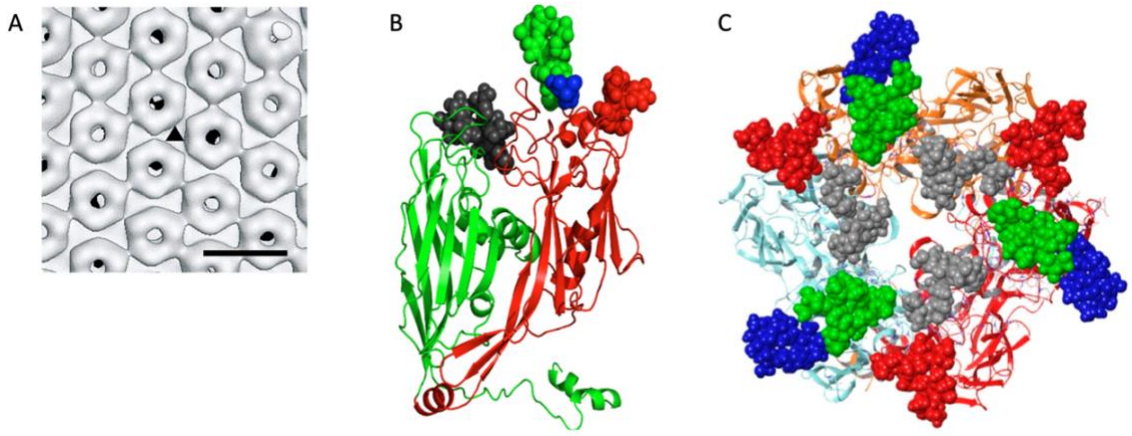

Figure 1.7. Structure of a PBCV-1 major capsid protein (Vp54) N-glycans. (a) Cryo-EM close-up view of PBCV-1 trimers. Shaded-surface representations of the reconstructions viewed down an icosahedral three-fold axis (triangle). Bar, 100 Å. (b) Structure of a single PBCV-1 MCP (Vp54). The four glycans attached to the MCP at amino acids Asn 280 (green), Asn 302 (black), Asn 399 (red) and Asn 406 (blue) are on the outer surface of the virus. (c) Three of the MCPs are assembled into a capsomer, viewed from the top. (a) is from [62] and (b,c) are from [117].

1.3.D. Glycan Structures Attached to Chlorovirus Major Capsid Proteins

The structures of the PBCV-1 Vp54 N-linked glycans were reported recently, and they consist of 8–10 neutral monosaccharide residues, producing a total of four glycoforms (Figure 1.8) [115]. These structures do not resemble any structure previously reported in the three Domains of Life. Among their most distinctive features are: (1) the four glycoforms share a common core structure, and the four glycoforms are related to

the non-stoichiometric presence of two monosaccharides, L-Ara and D-Man; the most abundant glycoform consists of nine neutral monosaccharide residues organized in a highly-branched fashion; (2) the glycans are attached to the protein by a β-Glc linkage, which is rare in nature and has only been reported in glycoproteins from a few organisms [118-121]; and (3) the glycoform contains a dimethylated Rha as the capping residue of the main chain, a hyper-branched Fuc residue and two Rha residues with opposite absolute configurations.

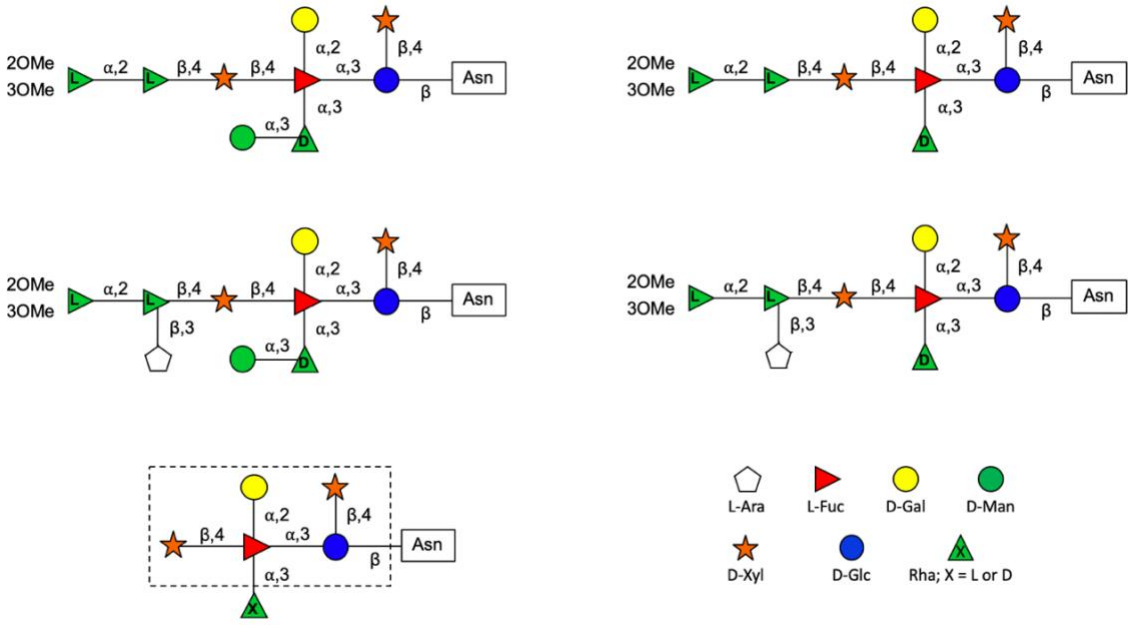

Figure 1.8. Structures of PBCV-1 Vp54 N-glycans. Ara and Man are not stoichiometric substituents and create four different glycoforms. The two on the left are the most abundant, and both have Man. The structure at the bottom represents the conserved core oligosaccharide that is present in all of the chloroviruses studied to date. Residues within the box are those strictly conserved, while D-Rha (outside the box) is a semi-conserved element because its absolute configuration is virus dependent. Figure modified from [97].

Attempts to fit the Vp54 glycan structures into the original Vp54 X-ray crystal structure [114] were unsuccessful and led to a re-examination of the original structure. This re-examination produced a structure that was compatible with the four N-linked glycan structures (Figure 1.7b). As mentioned above, the revised structure lacks the two O-linked glycans reported originally.

The PBCV-1 Vp54 has a molecular weight of 53,790 Da. The *a430l* gene encodes Vp54 with a predicted molecular weight of 48,165 Da so the combined sugars have a molecular weight of 5625 Da, which is about the weight of the four glycans. Vp54 was also reported to have a myristic acid attached to the carboxyl portion of the protein [122]. However, myristic acid has not been observed in any of the recent Vp54 structural experiments, and so, its status is currently unknown. The structures of the Vp54 glycans from the PBCV-1 antigenic variants, referred to above, are currently being determined, and as expected, the structures are truncated forms of the wild-type PBCV-1 glycans. PBCV-1 particles were reported to have two additional glycoproteins in addition to Vp54 [122]. Both of these glycoproteins react with the PBCV-1 antibody, and so, the glycan structures are predicted to be or identical to the glycans associated with Vp54. The gene encoding one of these proteins (Vp260) was identified (gene *a122r*). Gene *a122r* homologs are common in the chloroviruses, and some of the viruses have as many as five copies of the gene [123]. The role that Vp260 plays in the PBCV-1 virion is unknown.

The glycan structures of the MCPs from seven more chloroviruses, which represent all four chlorovirus types, were recently reported (Figure 1.9); collectively, all of the glycans have a common core region (outlined in Figure 1.8). The common core region consists of a pentasaccharide with a β-Glc linked to an Asn residue, which is not

located in the typical sequon Asn-X-(Thr/Ser). The Glc has a terminal Xyl unit and a hyperbranched Fuc, which is in turn substituted with a terminal Gal and a second Xyl residue. The third position of the Fuc unit is always linked to a Rha, which is a semi-conserved element because its confirmation is virus dependent. Additional decorations occur on this core N-glycan and represent a molecular signature for each chlorovirus.

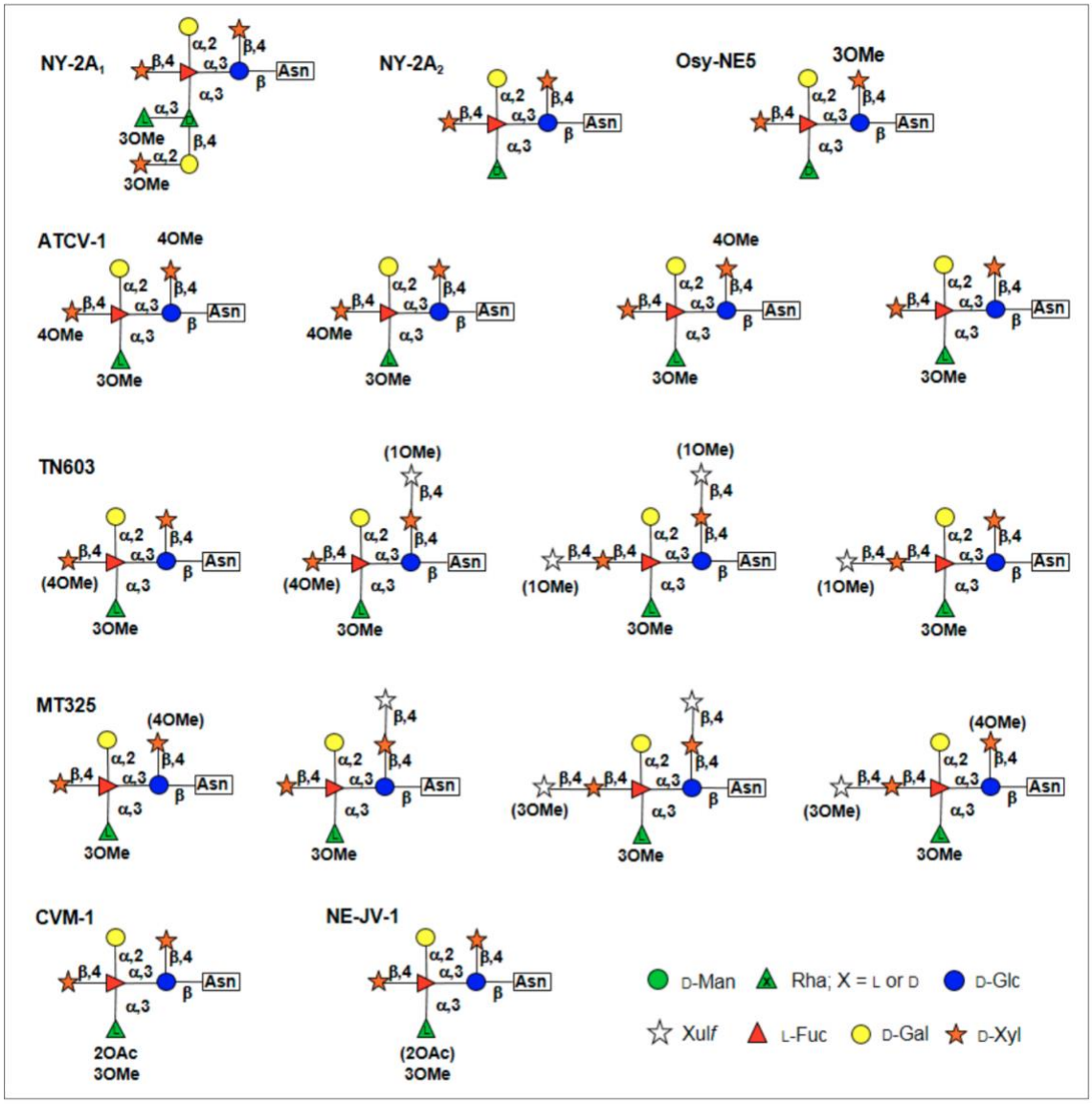

Figure 1.9. N-glycan structures from seven chloroviruses representing four chlorovirus types. Substituents in brackets are not stoichiometric. All sugars are in the pyranose form, except where specified. Virus NY-2A is an NC64A virus, virus Osy-NE5 an Osy virus, viruses ATCV-1 and TN603 SAG viruses and MT325, CVM-1 and NE-JV-1 Pbi viruses. Figure from [97].

1.3.E. Glycosyltransferases

In addition to the two GTases, hyaluronan synthase and chitin synthase previously described, the 43 chloroviruses collectively encode at least eight putative GTases (Table 1.3). PBCV-1 encodes at least eight putative GTases, some of which, if not all, participate in glycosylating the virus Vp54 protein [96, 111, 113]; they are scattered throughout the PBCV-1 genome (not shown in Table 1.3; A071R, and HAS and CHS proteins). Ongoing experiments have identified the role of three PBCV-1-encoded GTases with four GT activities that are involved in the synthesis of the glycan (Figure 1.10). Two of the three GTases were previously annotated as GTases, but the third protein was only identified in a recent study [111]. None of these eight PBCV-1 encoded GTases have an identifiable signal peptide that would target them to the ER. Furthermore, with the exception of PBCV-1 GTases A473L (six transmembrane domains; CESA CelA-like) and A219/222/226R (nine transmembrane domains; CXCX-2), none of the remaining PBCV-1 encoded GTases are predicted to have transmembrane domains. Therefore, these enzymes are expected to be soluble proteins. The genes for the PBCV-1 encoded GTases are expressed early during PBCV-1 infection [124]. Thus, assuming the enzymes are stable, they would be available for adding sugars to the Vp54 glycans during virus replication.

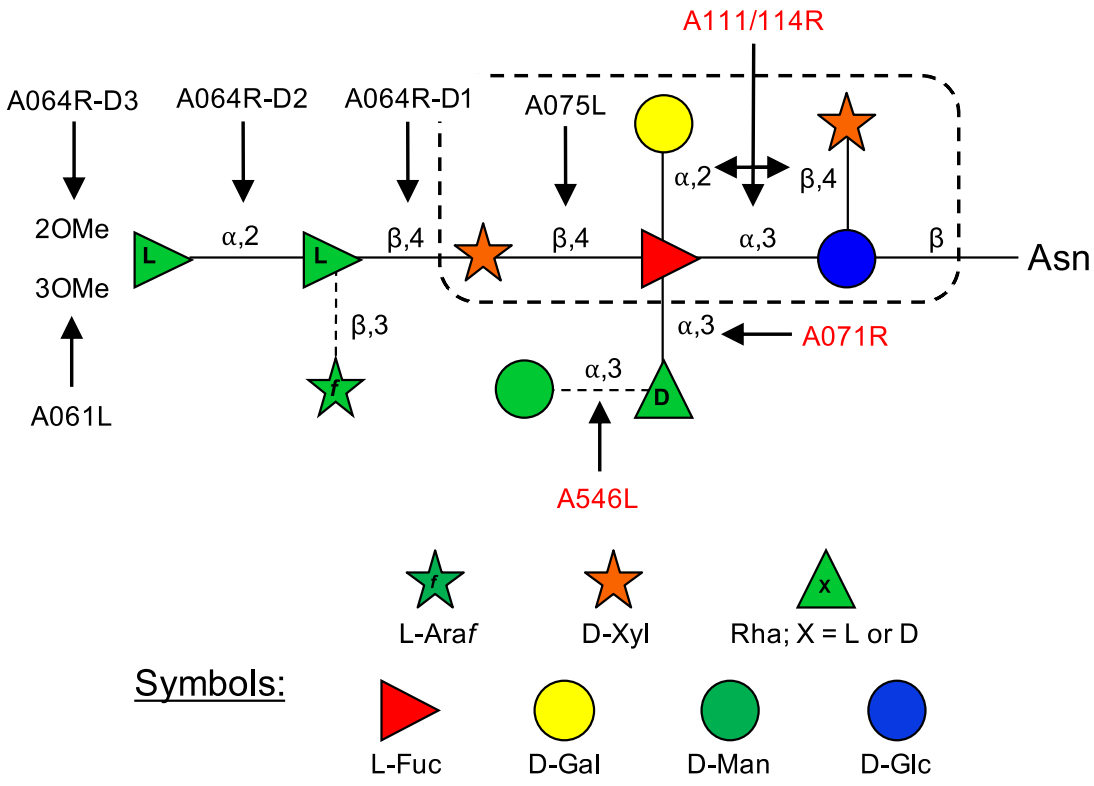

Figure 1.10. Structure of the glycan attached to the chlorovirus PBCV-1 major capsid protein (Vp54) with the predicted PBCV-1 encoded glycosyltransferases. The sugar Man and Ara connected by the dashed lines are nonstoichiometric substituents. The black box encloses the conserved pentasaccharide core structure common to all chloroviruses analyzed to date. The black labeled enzymes have been shown to carry out the listed reactions by examining the end products by NMR. The red labeled enzymes have been shown to carry out the listed reactions by the Glo™ assays. Note that protein A064R has 3 domains, each with an independent activity.

Now that structures of the glycans from the chlorovirus MCPs are becoming available, one can begin to characterize the viral encoded GTases biochemically. One

question that needs to be addressed is: Are the sugars added sequentially to the Vp54 protein backbone or are the glycans initially synthesized independently of Vp54, possibly on a lipid carrier and then attached to the protein in a single step? A slight variation of these two possibilities is that the core glycan is synthesized independently of the protein and then attached to Vp54. Additional sugars could then be added sequentially to these core glycans [96]. We suspect that this viral encoded glycosylation pathway represents a previously undescribed pathway, possible even a pathway that existed in eukaryotes prior to the ER and Golgi glycosylation pathway [102].

In the chapters to follow, this thesis will address individual PBCV-1 encoded GTases and MTases that participate in the synthesis of the unique glycoforms that decorate the virus Vp54 protein independent of the host microalga *Chlorella*. One by one, we will unpack each virus-encoded enzyme with evidence supporting its confirmed or predicted role in glycan synthesis (Figure 1.10). Of the eight GTases encoded by the 43 chloroviruses, only two of them, homologs of PBCV-1 A111/114R and A075L, are present in all of the viruses, and so, they are predicted to be involved in the synthesis of the core glycan structure. A111/114R is especially interesting because it is predicted to have three GT catalytic domains. Similarly, PBCV-1 encodes another three-domain protein (A064R) that performs three different functions; however, it is not coded by any other chlorovirus for which the glycan structures are already known. The following results strongly suggest that GT diversity is in its infancy of discovery. Ultimately, further investigation of chlorovirus glycosylation will provide the field of glycobiology with new discoveries and may ultimately answer the central question about what benefits these viruses have by maintaining their own glycosylation machinery through evolution.

1.4. GENETIC MODIFICATION OF CHLOROVIRUSES

Generating recombinant viruses require a pliable host. Unfortunately, reverse genetic manipulation of NC64A genomes is currently not possible and therefore directed modification of chloroviruses has yet to be achieved. To overcome this transformation bottleneck, we tried to develop a transformation system to generate stable site-directed chlorovirus mutants by targeting virus GT-genes to investigate their potential involvement in chlorovirus glycosylation. GT-genes were targeted simply because we developed an antibody-based selection scheme, described below, that can discriminate between wildtype and variant glycans attached to the MCP. Having the ability to carry out molecular genetic modifications of chloroviruses, with genomes of 290 to 370 kb, would help elucidate the function of both known and unknown, novel virus-encoded proteins. However, to date, only a few algal species and essentially none of their viruses have been genetically manipulated due to inefficient genome editing methods. Of particular note has been the recalcitrance of NC64A to genome modifications. Therefore, the development of heterologous gene expression tools for the viral *Chlorella*-like green algal hosts would be an important accomplishment.

More broadly speaking, genetic modification of chloroviruses would provide functional insight into the unusual chlorovirus encoded proteins mentioned previously such as hyaluronan synthase, potassium ion channel protein, five polyamine biosynthetic enzymes, and as addressed in this thesis, GTases. Furthermore, adoption of a reverse genetics system would also allow the exploration of formerly characterized proteins having potential scientific and economic benefit. For example, chlorovirus genes encode commercially important enzymes such as DNA restriction endonucleases and contain

elements for genetically engineering other organisms. Examples include viral promoter elements that function well in both monocots and dicots of higher plants, as well as bacteria [125, 126], and a translational enhancer element that function in *Arabidopsis* [127]. Chloroviruses have some of the smallest, most primitive forms of highly complex proteins that exist in higher organisms serving as a simplified fossil template to study biochemical models for mechanistic and structural studies [128]. Therefore, the development of a successful and reproducible chlorovirus procedure for achieving genetic modification would accelerate the exploration of these microalgae and their viruses for a broader range of scientific investigations and biotechnological applications. Such an achievement would be a major step forward.

The studies described here were based on harnessing the RNA-programmable CRISPR-Cas9 reverse-genetics pipeline, a bacterial immune system that has been repurposed for genome engineering, to induce mutagenicity and recombination in chloroviruses. This fast and easy genome editing strategy has accelerated research in many fundamental and applied disciplines by allowing the dissection of complex gene interactions and the construction of new pathways through synthetic biology. If we were successful in using the CRISPR-editing system in chloroviruses, we could knock out a variety of virus-encoded genes to observe what effect they would have on virus replication.

1.4.A. Transformation of Green Microalgae

Microalgae and their viruses are receiving increasing attention for their potential usefulness in a number of industrial applications, including bioremediation and the production of high-value specialty chemicals [129, 130]. The ability to genetically

engineer chloroviruses, in order to study and eventually manipulate their metabolic pathways, would greatly enhance the utility of microalgae-chlorovirus counterparts as scientifically and industrially important organisms. Considering no current reverse genetics system exists in either *Chlorella* or chloroviruses, we are equally limited in the capacity to either characterize gene function or exploit unique virus-encoded proteins. With the advent of CRISPR technology and the ongoing discovery of new giant viruses and their annotated genomes, we are armed with resources that have yet to be married. A significant barrier to genetic transformation in chloroviruses is the inaccessibility of its host green alga to genetic manipulation.

Since microalgal cells are not able to take up exogenous DNA by nature, several genetic techniques have been developed for this purpose. Classical mutagenesis has been used for the creation of random mutations and a more directed approach to generate specific insertions, deletions or substitutions into the host strain in order to produce the desired phenotype [131]. Among transformation methods for the delivery of exogenous DNA, the most common techniques are electroporation, *Agrobacterium tumefaciens*-mediated transformation, ballistic systems and agitation with glass beads [132-135]. Although, most of these techniques have been proven to work with great success in model strains such as *Chlamydomonas reinhardtii*, *Phaeodactylum tricornutum*, *Scenedesmus*, *Ankistrodesmus* and some *Chlorella* sp. [136, 137], there is a lack of efficient and stable transformation techniques that can be applied to a broader range of microalgae strains.

Genetic engineering of microalgal strains is difficult due to the great diversity of species with a variety of cell sizes, cell wall structures and composition and, likely,

unique responses to foreign DNA [138]. Like plant cells, *Chlorella* cells are surrounded by a rigid outer cell wall composed of polysaccharides with a variety of sugars as well as lesser amounts of protein and lipid that presumably makes them more difficult to transform [139]. DNA delivery can be challenging since DNA has to be transferred through the cell wall, plasma membrane and nuclear membrane Moreover, the cells have to be able to survive the chemical or mechanical treatment. Therefore, particular methods are needed for specific strains and, thus, a broader range of genetic tools have to be developed.

In order to transform *Chlorella*, various methods including glass beads [140], *Agrobacterium tumefaciens*-mediated transformation [141-143], PEG [144], protoplasting [145, 146], and electroporation [147-149] have been developed till now. Among these, as a preferred and efficient method, electroporation has also been applied to transform various microalgae, such as *Chlamydomonas reinhardtii* [132], *Scenedesmus obliquus* [150], and *Nannochloropsis* sp. [151]. Electroporation became a successful genetic tool for the transformation of several microalgal species [132, 152], however the protocol optimization is often challenging, time-consuming and most importantly, only proven in selected *Chlorella* species (*C. ellipsoidea*, *C. vulgaris, C. minutissima, C. zofingiensis* and *C. pyrenoidosa*) [153]. Microalgal species have a different resistance to transformation and thus are more or less prone to take up and incorporate exogenous molecules into their genomes. Moreover, viability can decrease rapidly when high voltages are applied and variations in DNA fragment lengths or macromolecule sizes can also decrease the transformation efficiencies [152].

Less traditional methods of microalga-transformation are being explored with some success. Karas *et al.* (2015) [154] and Diner et al. (2016) [155] showed that episomal plasmids containing a yeast-derived centromeric sequence CEN6-ARSH4-HIS3 can be transferred by conjugation from *E. coli* strains to the diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*. Recently, Munoz *et al.* (2019) [156] reported an efficient and stable transformation of the green microalgae *Acutodesmus obliquus* and *Neochloris oleoabundans* by transferring exogenous DNA from *E. coli* via conjugation.

1.4.B. CRISPR-CAS9 Genome Editing of Microalgae

The CRISPR-Cas9 system is a bacterial immune system that has been repurposed for genome engineering enabling researchers to manipulate the genome of cells [157-159]. This fast and easy genome editing system has accelerated research in many applied disciplines allowing the dissection of complex gene interactions and the construction of new pathways through synthetic biology. Presently, most industrial researchers prefer CRISPR technology for algal genome engineering, since it can be used as versatile tool for editing a single gene or even a complete metabolic pathway in shorter time than other classical genome engineering approaches [160]. Despite algal genome editing being in early-stage of development, major progress has been made in demonstrating proof-of-concept of genome editing using the type-II CRISPR system [161].

Targeted genome editing using RNA-guided endonucleases is an emerging tool in algal biotechnology. After successful establishment of the CRISPR-Cas9 system in animal and plant cells, this technology is being evaluated in a variety of algal cells. Recently, CRISPR-Cas systems have been widely used to manipulate the genome of both

freshwater and marine microalgae [162]. A general overview of different algae and genes edited by the CRISPR-Cas9 technology, transformants screening procedures, and Cas9-induced double-stranded break repair mechanisms is provided in Table 1.4 and discussed in detailed as follows. CRISPR technology has been demonstrated in microalgae including diatoms to manifest the function of the particular gene(s) and developing industrial traits, such as improving lipid content and biomass productivity [163]. However, application of Cas9 for algal genome engineering is in preliminary stages. As noted above, initial success has been achieved with a few model green algae (*Chlamydomonas*, *Nannochloropsis*, *Phaeodactylum* etc. [163-165]). Cas9-driven systems have been widely used in most of the studies for targeted knock-in, knock-out and knock-down of targeted genes in these algae [164, 166]. Development of methods for targeted genome editing of new strains of microalgae through CRISPR–Cas9 technology requires significant effort in optimizing DNA delivery methods and improving editing efficiencies.

Table 1.4.

Overview of CRISPR-Cas9-based genome editing of green algae and diatoms

Algae	CRISPR/Cas9 Targeted gene	Cas9 and gRNA delivery method	Screening methods of transformants	Cas9-induced break repair mechanism
C. reinhardtii	Endogenous FKB12 gene (rapamycin sensitivity), three exogenous genes; GFP (mGFP) gene, luciferase (Gluc) gene, hygromycin resistance (Hygro) gene	Vector mediated	Genotyping, marker/or reporter assay	NHEJ/KO
C. reinhardtii	MAA7, CpSRP43 and ChlM	RNP based	Whole genome sequencing, western and southern blot analysis, semi-qPCR.	NHEJ/KI
C. reinhardtii	CpFTSY and ZEP	RNP based	Zeaxanthin measurement, genotyping SDS-PAGE and western blot analysis	Double gene KO
C. reinhardtii	Argininosuccinate lyase (ARG), acetolactate synthase (ALS) gene, Endogenous FKB12 gene (rapamycin sensitivity),	Vector mediated	Arginine prototrophy, herbicide and antibiotic resistance, genotyping	HDR and NHEJ/KO KI
C. reinhardtii	Photoreceptor genes COP½, COP3, COP4, COP5, PHOT, UVR8, VGCC, MAT3 and aCRY	Vector/RNP both	Whole-cell qPCR, Protein analysis, Phototaxis Assay,	HDR and NHEJ/KO KI
N. oceanica	Nitrate reductase (NR)	Vector mediated	nPCR/RE analysis and deep sequencing	HDR
N. gaditana	20 TFs related to down-regulation of lipid production under nitrogen deprive condition	Vector mediated, Cas9 + editor cell line was developed, transient	Genotyping, q-PCR, FAME/TOC ratio analysis	HDR/KO
Volvox carteri	glsA and regA	Vector mediated	Hygromycin resistance, microscopy, western blot	NHEJ
Phaeodactylum tricornutum	CpSRP54 gene	Vector mediated	High-resolution melting (HRM) based PCR assays, qRT PCR	NHEJ
Thalassiosira pseudonana	Urease	Vector mediated	Growth on urea, genotyping	NHEJ

NHEJ non-homologous end joining, *HDR* homology directed repair, *KO* gene knock-out, *KI* gene knock-in, *qPCR* quantitative PCR, *qRT-PCR* quantitative real-time PCR, *RNP* ribonucleoprotein (Cas9–sgRNA complex). From Patel *et al*. Mol. Biotechnol. (2019).

Different CRISPR-Cas-systems-mediated transformation have been reported in the model green alga, *C. reinhardtii*. The first attempt to modify target genes in the genome of *C. reinhardtii* was made in 2014 [167]. Unfortunately, the implementation of CRISPR-Cas9 in *C. reinhardtii* was not as straightforward or as efficient as in other model organisms. Previous studies showed highly inefficient gene editing in *C. reinhardtii* using conventional Cas9 and single guide RNA (sgRNA) genes which yielded only one colony with a modified target locus per 1.5×10^9 initial cells used for transformation [167]. From there, several research teams have developed different strategies and methods to transform *C. reinhardtii* cells either by experimenting with CRISPR systems [168], transformation techniques [169], and even by modifying CRISPR-associated protein used for the purpose [170]. Altogether, these studies showed that different factors and parameters need to be considered for successful CRISPR modification in green algae.

The use of circular recombinant plasmids for CRISPR transformation has been a standard method in related research. This plasmid-based approach (Figure 1.11a) employees a suitable plasmid backbone into which has been inserted the *Cas9* gene encoding for its protein and a gene encoding a sgRNA that will hybridize with the target gene. After that, the plasmid is digested and linearized with a restriction endonuclease enzyme(s) and integrated into the microalgae's genome following genetic transformation (usually via electroporation). Cas9 protein, guided by its sgRNA, then induces a DSB (double-strand break) and potential disruption of the targeted gene. An alternative approach that avoids many complications inherit in using Cas9 and sgRNA genes is the use of ribonucleoprotein (RNP)-mediated CRISPR genome editing or the also called

DNA-free CRISPR (Figure 1.11b). This direct delivery of RNPs alleviates difficulties with protein expression in cells in which common eukaryotic promoters are not expressed well. Indeed, this method does not require the delivery of foreign DNA. In addition, the Cas9-sgRNA RNP is degraded over time - thus limiting the potential for off-target effects.

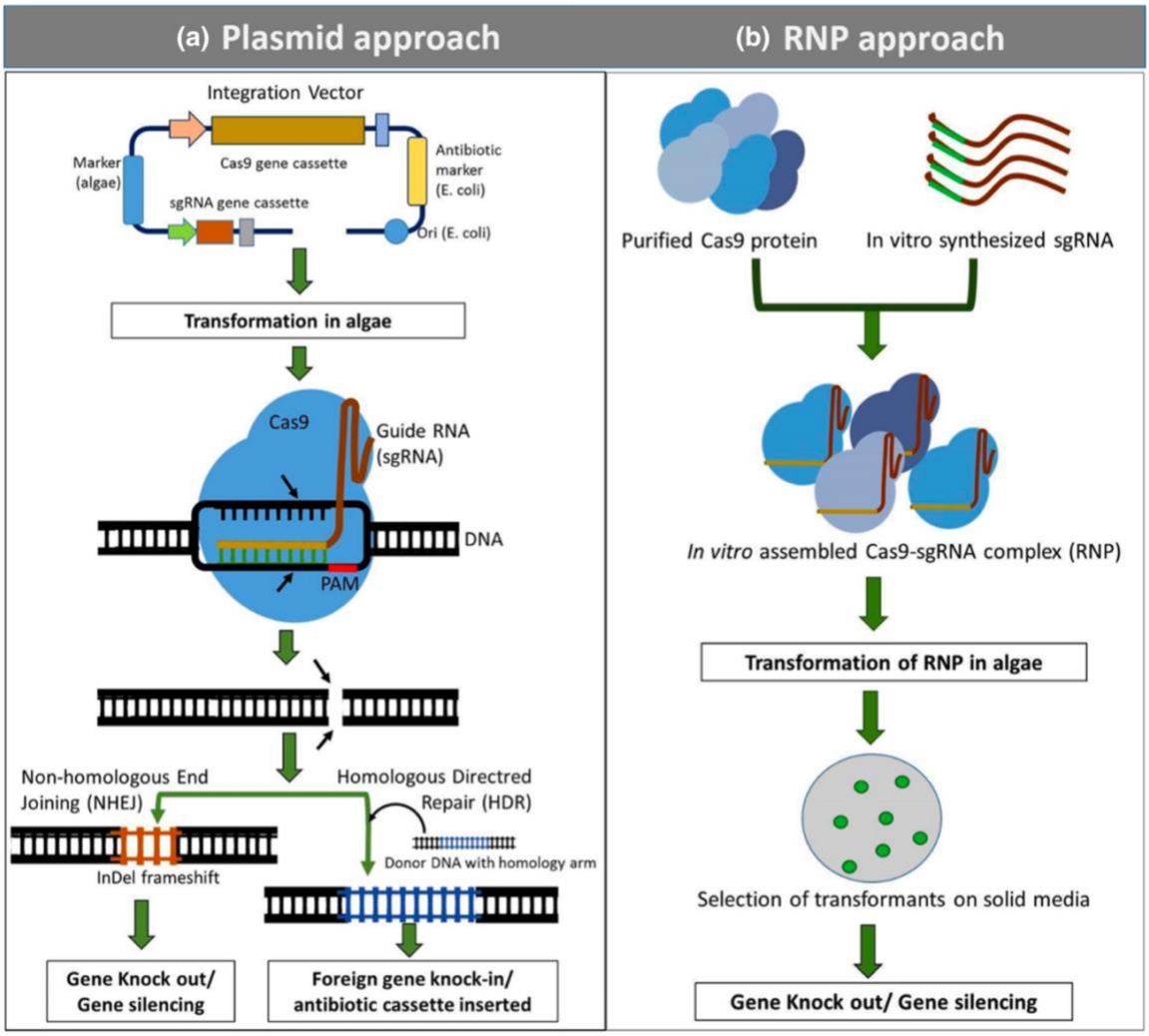

Figure 1.11. Schematic representation on approaches used for CRISPR-Cas9 system-based algal genome engineering. (a) Integration vector-based approach that results in Cas9–sgRNA nuclease complex formation within an algal cell followed by double strand DNA cleavage and gap repair by NHEJ or HDR. NHEJ-based gap repair most often results in insertion or deletion of nucleotide bases (indels) and this mutation in the target gene often results in gene silencing. HDR-mediated DNA repair requires donor DNA with homology to the target DNA that allows homologous recombination and insertion of new DNA sequences into the target gene (i.e., knock-in mutations). (b) In the

RNP-based approach, purified Cas9 protein and laboratory synthesized sgRNA are assembled *in vitro* to form a CRISPR-RNP complex. These complexes are then transformed in competent cells to create gene knock-out mutants. If DNA fragments homologous to the target gene are supplied along with the Cas9/RNP, HDR-directed gene modification (i.e., gene knock-in) can occur. Figure from [162].

High-efficiency protocols for transformation using CRISPR-Cas9 plasmid-based approaches and Cas9/RNPs have been developed for genome engineering of photosynthetic microalgae [164, 171]. Recently, a plasmid-based gene-within-a-gene (Cas9 gene containing an artificial intron with an inserted sgRNA encoding gene) construct was designed and found to be functional in both *C. reinhardtii* and tobacco cells [168]. The authors reported comparatively higher (3×10^{-7}) transformation efficiency for the *FKB12* gene. An arginine auxotroph (*arg7-8*) was co-transformed with Cas9-intron sgRNA and 80 nucleotides ssDNAs complementary to the argininosuccinate lyase (ARG) gene, and arginine prototrophs were isolated at much higher efficiency. When the authors used a ssDNA oligonucleotide designed to appropriately modify the acetolactate synthase (ALS) gene and supplied it along with a Cas9 gene-within-a-gene construct targeting the ALS gene, they were able to create a sulfometuron methyl herbicide-resistant strain of *C. reinhardtii* [168].

There are a number of reports in which Cas9-RNP-based approaches have been used for algal genome engineering. This technology utilizes purified Cas9 protein and *in vitro* synthesized sgRNA. The Cas9 protein and sgRNA are pre-assembled *in vitro*, and directly delivered to algal cells either via electroporation or by biolistic method [166,

172]. In *C. reinhardtii*, Cas9 RNPs comprising Cas9 protein and sgRNAs were directly delivered to the cells by electroporation and created targeted mutations in multiple loci [166, 173]. RNP-based approaches have also been used to generate more robust strains of the industrial alga *Coccomyxa* as a biofuel cell factory [174]. RNP-based approaches have unique advantages and are comparatively simpler than plasmid-based approaches. For example, vectors having strong constitutive promoters for Cas9 expression which often results in continuous expression and higher production of Cas9 inside the cells. Higher concentration of Cas9 may kill the chassis organism by creating a toxic environment (e.g. by creating an abundance of non-specific DNA cleavages and/or causing a lethal level of off-target mutations). Cas9-sgRNA complex used in this approach requires stringent quality control steps. It is especially important to demonstrate *in vitro* DNA cleavage activity at the desired target site on a PCR amplified target DNA. Optimization of sgRNA and Cas9 concentration for assembly of functional RNP and development of a robust transformation/RNP delivery system in cells are critical in using this approach. Indeed, transformation of *in vitro* assembled RNP complex into algal cells requires series of optimization to achieve high transformation efficiency. One of the key considerations for algae competent cell preparation is growth phase and density of cultures which may vary from strain to strain. To our knowledge, overnight grown cultures at 0.3–0.5 OD_{600} are comparatively suitable to achieve higher transformation efficiency in *Chlamydomonas* and *Chlorella*. Optimization of electroporation methods is also specific for different strains. In *C. reinhardtii*, electroporation is mostly used for successful and high-efficiency transformation [167].

CHAPTER II

CHLOROVIRUS GENETIC TRANSFORMATION

2.1. OVERVIEW

The overall goal of this research aim is to develop a transformation system that supports the genetic modification of the large dsDNA chloroviruses that infect unicellular *Chlorella*-like green microalgae. Specifically, we aim to leverage CRISPR, an RNA-guided gene editing technology, to investigate the progression of chlorovirus glycosylation that takes place independent of its host cell's biosynthetic machinery. The ability to genetically engineer chloroviruses, in order to study and eventually manipulate their metabolic pathways, would greatly enhance the utility of microalgae-chlorovirus counterparts as scientifically and industrially important organisms. Considering no reliable current reverse genetics system exists in either *Chlorella* NC64A or chloroviruses, we are equally limited in the capacity to either characterize gene function or exploit unique virus-encoded proteins. In the advent of CRISPR technology and the ongoing discovery of new giant viruses and their annotated genomes, we are armed with resources that have yet to be married.

A significant barrier to genetic transformation in chloroviruses is the inaccessibility of its host green alga to DNA or protein uptake. Like plant cells, *Chlorella* cells are surrounded by a rigid outer cell wall composed of polysaccharides with a variety of sugars as well as lesser amounts of protein and lipid that presumably makes them more difficult to transform. To address these limitations, we developed a series of protocols with an original selection scheme designed to introduce exogenous DNA and protein in an effort to generate chlorovirus GT-gene mutations. We chose to

target another NC64A virus for gene editing, CA-4B. CA-4B-encoded *034r* is a homolog of PBCV-1 *a064r*, that encodes for a highly characterized GT with three domains involved in protein glycosylation: domain 1 has a *β*-L-rhamnosyltransferase activity, domain 2 has an α-L-rhamnosyltransferase activity, and domain 3 is a MT that decorates one position in the terminal α-L-Rha unit [175]. Based on the low methylation levels of CA-4B DNA [32], we predict there is less likelihood the CA-4B virion packages restriction enzymes unlike PBCV-1 and other chloroviruses. Therefore, upon infection of NC64A cells, we believe delivery of any exogenous cargo, like CRISPR editing-RNA and Cas protein, is less susceptible to virus-encoded restriction endonucleases.

The overall strategy to modify chlorovirus DNA involved testing a variety of transformation methods that could support the delivery of preassembled Cas9 protein-sgRNA RNPs to generate a targeted gene cleavage event in the CA-4B putative GT-gene *034r*. Transformation methods that were investigated included protocols with cell wall-targeting enzymes, electroporation, silicon carbide (SiC) whiskers, and cell-penetrating peptides (CPP). The RNA-directed selection of a specific 20–22 bp nucleotide sequence within the target gene *034r* by the Cas9/sgRNA complex allows the two nuclease domains of Cas9 to create a DSB at a predetermined site within the gene of interest. Repair of the DSB by the error-prone nonhomologous end joining (NHEJ) DNA repair system can result in gene inactivation (i.e., gene knockout). Alternatively, replacement of the cleaved DNA segment with a closely related DNA fragment via homologous-directed recombination (HDR) can result in gene replacement (i.e., gene knockin) or nucleotide(s) substitution. Previously, single-stranded oligodeoxynucleotides (ssODNs) reportedly provide 100-fold lower levels of nonhomologous integrations compared with double-

stranded counterparts [176] and have demonstrated scarless genomic editing capabilities with reduced unwanted off-target cutting [170]. Adopting this strategy should increase efficiency of targeted DNA editing and replacement in chloroviruses.

Here, we explored whether targeted DNA cleavage could be used to facilitate homology-mediated mutagenesis by using a DNA repair template complimentary to the *034r* gene. We used tandem co-delivery of a DNA template in the form of ssODN with Cas9/RNPs to achieve HDR. The ssODN is approximately 80 nt long, designed with homology arms extending 40 nt upstream and downstream of the sgRNA target in CA-4B *034r*, respectively (Figure 2.1). The ssODN contained two critical elements that permitted selection when incorporated into a mutant virus: (1) a MscI point mutation (T to A) that converts the native restriction site TGGCCA to AGGCC, and (2) a nucleotide substitution (C to A) that introduces a premature stop codon (TAA). The HDR event would remove the MscI restriction site and allow us to verify HDR-mediated insertion by treating amplicons with the restriction enzyme to test for DNA replacement at the site of Cas9/RNP-mediated DNA cleavage. Amplicons were subsequently sequenced to confirm RNP-cleavage. Further, the premature stop codon would eliminate translation of the protein resulting in a shortened glycan at the MCP surface of newly formed viruses when infected with wild-type PBCV-1. In turn, these virus mutants would be subject to antibody selection as described in detail in the subsequent Materials and Methods section.

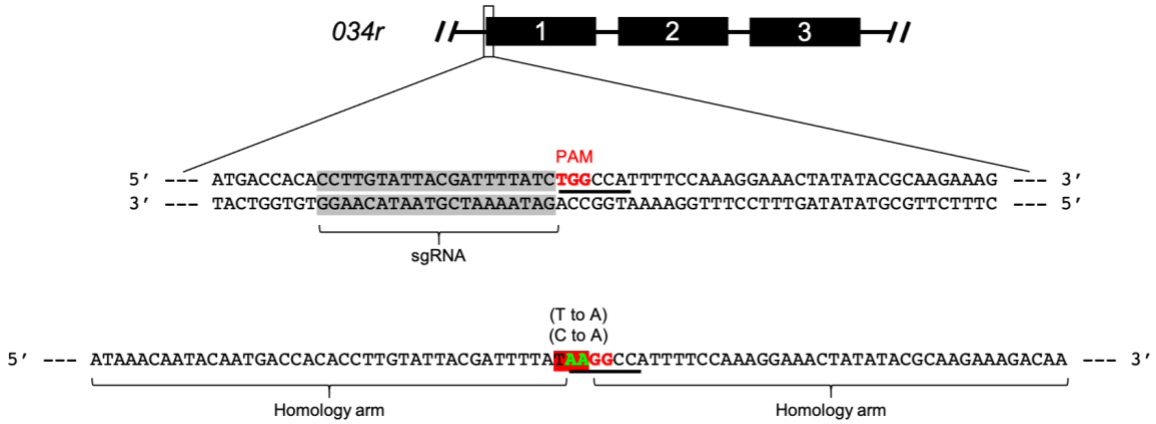

Figure 2.1. sgRNA and ssODN designs for targeting CA-4B putative GT-gene *034r*. Top: Cartoon rendering of CA-4B putative GT-gene *034r* composed of three domains. Middle: Design of a Cas9/sgRNA gene targeting CA-4B gene *034r* to produce mutant CA-4B with truncated surface glycans. Shaded gray, sgRNA target site; red TGG, PAM site; underlined, MscI restriction site. Bottom: Design of ssODN to replace a dinucleotide in the target gene. The homology arms specific to the gene target are flanking the nucleotides to be changed (green text). Shaded red, stop codon.

2.2. ALGA AND VIRUS, RNP, AND SELECTION

2.2.A. Materials and Methods

Alga growth conditions. *Chlorella variabilis* NC64A cells were grown in Bold's basal medium (BBM) (3 mM $NaNO_3$, 170 μM $CaCl_2$ $2H_2O$, 304 μM $MgSO_4$ $7H_2O$, 431 μM K_2HPO_4, 1.3 mM KH_2PO_4, 428 μM NaCl, 12 μM Na_2EDTA, 2.2 μM $FeCl_3$ $6H_2O$, 1.2 μM $MnCl_2$ $4H_2O$, 220 nM $ZnSO_4$ $7H_2O$, 50 nM $CoCl_2$ $6H_2O$, 99 nM Na_2MoO_4 $2H_2O$, 6.4 μM $CuSO_4$ $5H_2O$, 184 μM H_3BO_3) modified by the addition of 0.5% sucrose and 0.1% peptone (MBBM) [94]. All experiments were performed with cells grown to early

log phase (4 - 7 × 10^6 cells/mL). Cell cultures were shaken (200 rpm) at 26 °C under continuous light.

***In vitro* cleavage assay.** Production of chloroviruses PBCV-1, CA-4B and DNA isolation were performed as described [177, 178]. Target loci were PCR-amplified and purified by 2% agarose gel electrophoresis. Purified *Streptococcus pyogenes* strain Cas9 (SpyCas9) and *Lachnospiraceae* bacterium Cpf1 (LbCpf1) (200 nM) were independently preincubated with sgRNA (600 nM) in cleavage buffer [1× NEBuffer 3 (New England Biolabs), 10 mM DTT, 10 mM $CaCl_2$] at 37 °C for 15 min. Target DNA (20 nM) was added to a final volume of 20 μL. Reactions were incubated at 37 °C for 1 h. Cleavage reactions were purified by using a MinElute PCR Purification Kit (Qiagen) and then resolved on 2% agarose gels and imaged on a Gel Doc XR+ and ChemiDoc XRS+ systems (Biorad) (Figure 2.2).

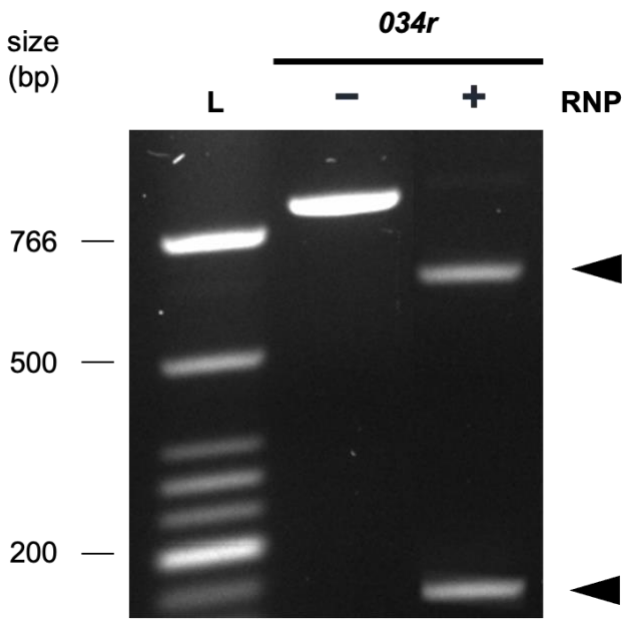

Figure 2.2. *In vitro* cleavage assay. The target *034r* locus from NC64A virus CA-4B was PCR-amplified and incubated with preassembled Cas9 and sgRNA RNP complexes *in vitro*. The complete *in vitro* cleavage of the target locus confirmed active RNP formation. Arrowheads indicate cleaved products. L, ladder.

CA-4B mutant selection using antigenic mutant antibody. NC64A cells were infected with CA-4B (MOI 5) and incubated for 12 h. Site-directed mutant viruses were selected with anti-Rabbit IgG magnetic beads (RayBiotech) according to the manufacturer's protocol (Figure 2.3a). In summary, rabbit polyclonal antiserum prepared against wild-type PBCV-1 collected from previous antibody studies [110], which also binds CA-4B, was added to the viral lysate and allowed to bind wild-type virus for 1 h. Goat anti-Rabbit IgG magnetic beads were incubated with the rabbit antibody solution for 30 min and then separated by magnets. The unbound viruses were collected and subsequently incubated

for 1 h with rabbit polyclonal antiserum derived from serologically distinct PBCV-1 mutants that have a mutation in gene *a064r*, homologous to CA-4B gene *034r*, that in turn produces truncated surface glycans. Goat anti-Rabbit IgG magnetic beads were incubated again with the rabbit antibody solution for 30 min and then separated by magnets. After the unbound particulates were washed from the beads, the bound mutant antibodies were eluted from the beads using the elution buffer. The beads were then magnetically separated from the eluted solution, and the eluted antibodies coupled to mutant viruses were removed manually and plaque assayed. Individual plaques were selected, and target DNA regions were PCR amplified with specific primers. The PCR products were verified by agarose gel electrophoresis, eluted from the gel and sequenced using the Sanger method (GENEWIZ).

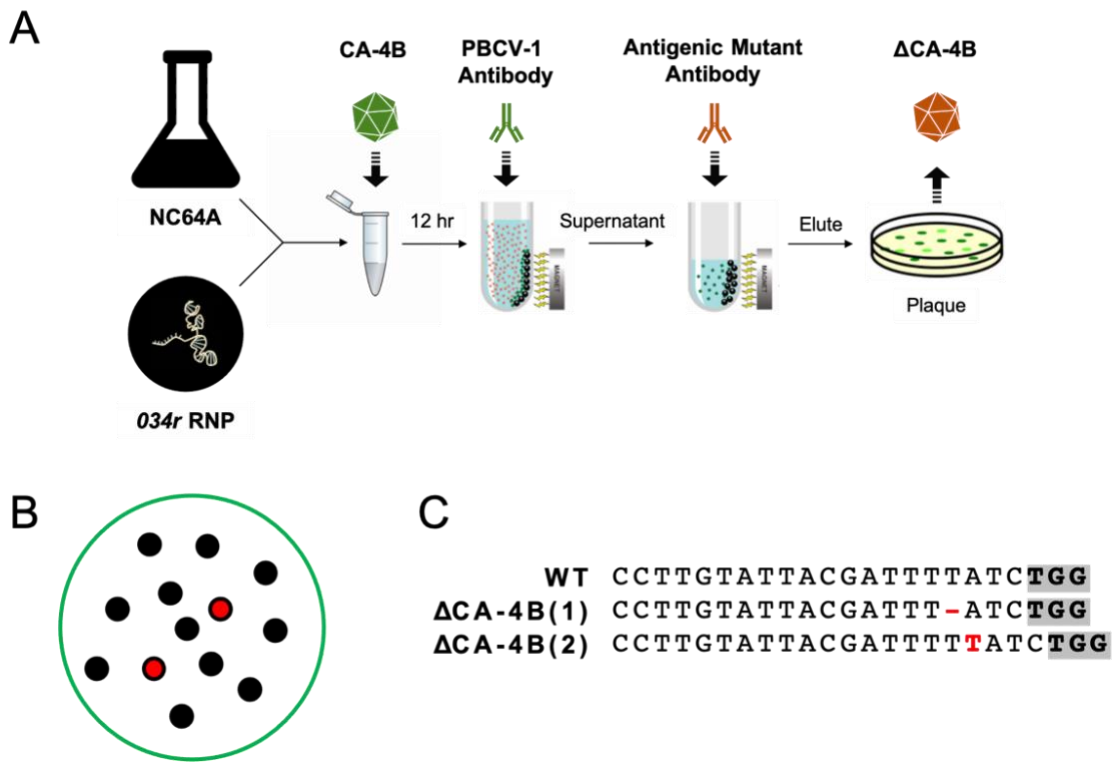

Figure 2.3. Recovery of CA-4B site-directed mutants by RNP-targeting. (a) Workflow of antibody selection for mutant chloroviruses. Following incubation with *034r*-targeting RNP, macerozyme-treated NC64A cells were infected with CA-4A (MOI 5). Rabbit polyclonal antiserum (PBCV-1 antibody, green) was added to the viral lysate. Goat anti-Rabbit IgG magnetic beads were incubated with the rabbit antibody solution and then separated by magnets. The escaped viruses were collected and subsequently incubated with antibody derived from serologically distinct PBCV-1 mutants (antigenic mutant antibody, red) that have shortened surface glycans. Magnetic beads were incubated again with the antibody solution and then separated by magnets. Bound mutant viruses were eluted from the beads and plaque assayed. (b) Following antibody selection for mutant CA-4B, 13 plaques were recovered. Domain 1 of *034r* from each virus isolate

was sequenced. (c) Two indels (red) in *034r* from two CA-4B variants (1 and 2) were detected and sequence verified for site-directed mutagenesis by preassembled Cas9 protein-sgRNA RNPs in the PBCV-1 *a064r* homologous gene. PAM site, gray. WT, wildtype sequence.

2.3. ENZYMATIC CELL WALL DEGRADATION

2.3.A. Introduction

Microalgal cell walls are complex and poorly understood [179]. The *Chlorella* intraspecies variation in cell walls as well as variations observed in a single strain grown under different conditions can be dramatic, and thus it is difficult to predict which of the compounds noted below will be present in any one strain. While some *Chlorella* species have only a single microfibrillar layer, others have two layers with the microfibrillar layer proximal to the cytoplasmic membrane and a mono or trilaminar outer layer [180]. The cell walls of *C. vulgaris* and other green microalgae are known to have rigid wall components embedded within a more plastic polymeric matrix [181]. This matrix is defined as the fraction that is hydrolyzable in 1 N NaOH or 2 M triflouroacetic acid (TFA) and contains uronic acids, Rha, Ara, Fuc, Xyl, Man, Gal, and Glc. The rigid, TFA-resistant cell wall is either glucosamine or a Glc-Man polymer [182]. The predominant amino sugar found in the rigid cell wall of *Micractinium conductrix* Pbi is *N*-acetylglucosamine present as a chitin-like glycan. Chitin (poly-*β*-1,4-D-*N*-acetylglucosamine), a polymer commonly found in shellfish, insects, fungi, worms, and mushrooms which can be deacetylated, either partially or fully, to produce chitosan (poly-*β*-1,4-D-glucosamine). Strong acid (6 N HCl) can hydrolyze the rigid cell wall of a

Chlorella sp., which removes the acetyl group from *N*-acetylglucosamine leaving glucosamine while hydrolysis with strong sulfuric acid fails to release detectable neutral sugars. This suggests that this strain of *Chlorella* does not have cellulose in the rigid cell wall.

Digestion of the rigid wall with chitinase releases *N*-acetylglucosamine yet complete digestion requires a chitosanase, which indicates that glucosamine polymers are also present in the rigid wall of *Micractinium conductrix* Pbi [183]. Based on these previous studies, the genus *Chlorella* was in-part defined as containing glucosamine as the common dominant cell wall polymer [184]. The neutral cell wall sugars of symbiotic *Chlorella* strains contain in decreasing amounts the sugars Glc, Rha, Ara, Man, Xyl, Fuc, and Gal, respectively [185]. Other than the neutral sugars (25 %), the cell wall contained 15–20 % uronic acids, 10–15 % glucosamine, and 6–10 % protein. This cell wall composition was reported to be similar to the glucosamine containing cell wall composition of the non-symbiotic algae *C. vulgaris, C. sorokiniana*, and *C. kessleri*; however, the authors were unable to determine the composition of 30–44 % of the cell wall [185]. Previously [186], isolated *Chlorella* NC64A cell walls were analyzed for component monosaccharides (Table 2.1).

Table 2.1.

Monosaccharide composition of Chlorella variabilis NC64A cell wall from Meints et al. Virol. J. (1988).

Sugar	**Mole percent**	**Percent of Soluble fraction**
Arabinose	5.3	2.6
Fucose	0.7	0.3
Galactose	9.1	4.5
Glucose	51.3	25.5
Mannose	5.3	2.6
Rhamnose	16.0	8.0
Xylose	5.5	2.7
Glucosamine	6.9	3.4
Total	100.0	49.6

Enzymatic digestion of algal cell walls has been previously studied as a methodology to make protoplasts of algal cells usually in cell preparation for genetic transformation. Protoplasts of *C. saccharophila* and *C. ellipsoidea* were made using an enzyme mix containing cellulase, hemicellulase, and pectinase activities [187]. However, these green algae, which are presumed not to have an outer layer or sporopollenin (now known as algaenan), required up to 90 h to display any effect from the enzymes. The cells became very sensitive to physical pressures, bursting simply from the pressure of a cover slip placed on a microscope slide. Similarly, protoplasts were made from *C. vulgaris* K-73122 using acromopeptidase, cellulase, chitosanase, gluczyme, and uskizyme; however, only 88 % of the cells were considered osmotically labile [188]. Protoplasts were also made in *C. ellipsoidea*, *C. vulgaris*, and *C. saccharophila* using three different enzyme mixes containing cellulase, macerozyme, pectinase, cellulysin, and driselase, although 7

of the 12 strains tested showed no protoplast formation with any of the formulations [180].

In the work presented here, we sought to define the extent of degradation that a variety of enzymatic activities have on the cell walls of NC64A. As the cell wall provides the main protective barrier between the algae and its surroundings, a detailed characterization of enzymes that act on the wall could enable the rational design of an enzyme cocktail that would allow efficient degradation of the wall and delivery of CRISPR-targeting cargo. Removal of the wall could allow more efficient transformation technologies to be developed. Finally, our findings could provide broader insight on the conversion of these cell wall sugars into a more usable substrate for downstream processes, or potentially ease lipid extraction from these organisms.

2.3.B. Materials and Methods

Cell wall-degrading enzymes. Achromopeptidase, alginate lyase, α-amylase (960 FAU/ mL), chitinase, chitosanase (25.9 U/mL), chondroitinase ABC (5 U/mL), drieselase, β-galactosidase (1,000 U/mL), β-glucosidase, hemicellulase, heparinase II (10 U/mL), hesperidinase, hyaluronidase, kitalase, laminarinase, lyticase, lysing enzymes from *Aspergillus* sp., lysozyme, mutanolysin (10 U/mL), pectinase (3,000 U/mL), pectolyase, phospholipase A_1 (10 kU/mL), phospholipase A_2, proteinase K, sulphatase (3.37 mg/mL), and trypsin were purchased from Sigma. β-glucuronidase (140 U/mL) was purchased from Roche. Neuraminidase (50 kU/mL), β-*N*-acetylhexosaminidase (5 kU/mL), β-N-acetylglucosaminidase (4 kU/mL), α-1-6 mannosidase (32 kU/mL), α-2-3 neuraminidase (50 kU/mL), 1-2,3 mannosidase (40 kU/mL), and endo-α-*N*-acetylgalactosaminidase (40,000 kU/ mL) were purchased from New England Biolabs (Ipswich, MA). Glusulase

was purchased from NEN Life Science Products (Boston, MA) (10,000U sulfatase and 90,000 U β-glucuronidase/mL). Zymolyase (10 mg/mL) was purchased from ZymoResearch (Irvine, CA). Macerozyme was purchased from RPI corp (Mt Prospect, IL). Cellulase was purchased from Sigma Aldrich (St. Louis, MO). Stock concentrations of enzymes were 20 mg/mL unless otherwise noted.

Enzyme inhibition assay. To assay growth inhibition due to enzymatic activity, 200 μL of algal cells normalized to an OD_{750} of 1.0 was mixed with 4 mL of media containing 7.5 g/L agar, which was at 42 °C and poured on a petri plate containing 15 g/L agar. Once hardened, 10 μL of enzyme stock was then spotted on the top agar and plates were incubated in the light at 26 °C for 5 days. As a negative control for enzymatic activity, enzymes were heat denatured at 100 °C for 10 min and spotted on plates.

Flow cytometry. Cell permeability assays were performed on the BD FACSAria cell sorter using ATTO™ 550 labeled Alt-R™ Cas9 tracrRNA (IDT 1075928). One μL of ATTO™ 550 labeled tracrRNA (200 uM) was added, incubated for 2 min, and loaded on the BD FACSAria where 20,000 cells were imaged. Samples were excited using a 488 nm laser and 660–740 nm (chlorophyll) and 480–560 nm (ATTO) emission data as well as bright field image data were collected. Populations were gated for in-focus cells and analyzed for permeability.

NC64A transfection by enzymatic digestion. One mL of cell culture of NC64A (5×10^6 cells) was taken, centrifuged at 8,000 rpm for 5 min, and the pellet was resuspended in the same volume of a MBBM in the presence of cell wall-degrading enzymes (Table 2.2) or MBBM only. The culture was incubated for 24 h at 25 °C in continuous light. NC64A cells were centrifuged and resuspended in a solution (0.8 M NaCl and 0.05 M

$CaCl_2$). Prior to NC64A cells being tested for transforming activity with RNPs, all treated cells were previously infected with virus and analyzed by plaque assay to confirm treated cells retained their permissive qualities required for virus attachment and infection. To generate target-specific site-directed mutants using RNP complex in NC64A, enzyme-treated and nontreated cells were transformed with either Cas9 protein or Cpf1 (200 μg) premixed with *034r*-targeting sgRNA (140 μg). Cas9 and Cpf1 proteins in storage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol) were mixed independently with sgRNA dissolved in nuclease-free water and incubated for 10 min at room temperature. The cell mixtures were incubated for 15 min at room temperature. Enzyme-treated NC64A cells were then infected with CA-4B and incubated overnight. Virus mutants were selected using the same antibody selection assay previously described. Virus plaques were analyzed by PCR and sequenced to detect Indel mutations by NHEJ, while *034r* amplicons were also incubated with MscI to screen for ssODN-mediated HDR.

Table 2.2.

Growth inhibition of Chlorella variabilis NC64A by a variety of enzymes

Enzyme	**NC64A**
Cellulase	–
Chitinase	+ + +
Chitosanase	+
Driselase	–
β-Glucosidase	–
β-Glucuronidase	–
Hyaluronidase	–
Laminarinase	+ + +
Lysozyme	+ + +
Lyticase	+ + +
Macerozyme	+ + +
Pectinase	+ + +
Pectolyase	–
Sulphatase	+
Trypsin	–
Zymolyase	–

–, no growth inhibition; +++, complete growth inhibition

2.3.C. Results and Discussion

To access enzyme activity against algal cell walls, we developed a reasonable throughput growth inhibition assay on agar plates, where the ability to inhibit growth was compared to heat denatured enzyme. Growth inhibition suggests the enzyme is either degrading the cell wall during construction or the enzyme interferes with precursor generation prior to assembly into the cell wall. From a variety of enzymes tested, chitinase, laminarinase, lysozyme, lyticase, macerozyme, and pectinase had the broadest effect on *Chlorella* NC64A (Table 2.2) based on the zone of inhibition around the spotted

area while chitosanase and sulphatase demonstrated minor growth inhibition by comparison. Heat denatured enzymes did not impact growth in our assay, indicating the growth inhibition was due to the enzymatic activity alone and not from other potentially toxic components of the enzyme preparations (data not shown). In one study [181], 14 different strains of *Chlorella* were challenged with cell wall-degrading enzymes and no two strains had the same pattern of inhibition. This large range of sensitivity to various enzymatic activities illustrates the wide diversity of cell wall composition amongst the alga *Chlorella* species. Our results show that in methodologies or processes using intact algal cells or residual algal biomass, enzymatic treatment can have large impacts on the permeability of the algal cell walls and may be useful in optimization.

After finding enzymes that inhibited growth, we examined their effects on cell walls of actively growing cultures. To process a large number of enzymes individually and in combination, we developed a rapid assay that would give statistically relevant results. We postulated that degradation of the cell wall may have dramatic effects of the cell membrane as well, which could lead to an altered permeability barrier. Cell permeability due to enzymatic disruption of the wall was visualized using an imaging flow cytometer, the BD FACSAria. To evaluate translocation of the RNP across the cell wall/membrane barrier, we used the 67mer Alt-R™ CRISPR-Cas9 tracrRNA (IDT) that has a ATTO™ 550 fluorescent dye attached to the 5' end. The fluorescent dye allows for a visual analysis of transfected cells and cell sorting by FACS. During data acquisition, algal cells were positively defined by their chlorophyll autofluorescence. A minimum chlorophyll autofluorescence was set to eliminate potential false positives from bacteria and debris present in the culture. The resultant data were analyzed by setting a minimum

ATTO™ 550 fluorescent intensity threshold, such that the majority of untreated cells (Figure 2.4a) had fluorescence intensity lower than the threshold. Cells with a fluorescent intensity above this threshold were considered permeable to the tracrRNA dye as indicated by cells pretreated with macerozyme (Figure 2.4b), or vortexed with SiC whiskers or electroporated prior to FACS (Figures 2.4c and 2.4d, respectively).

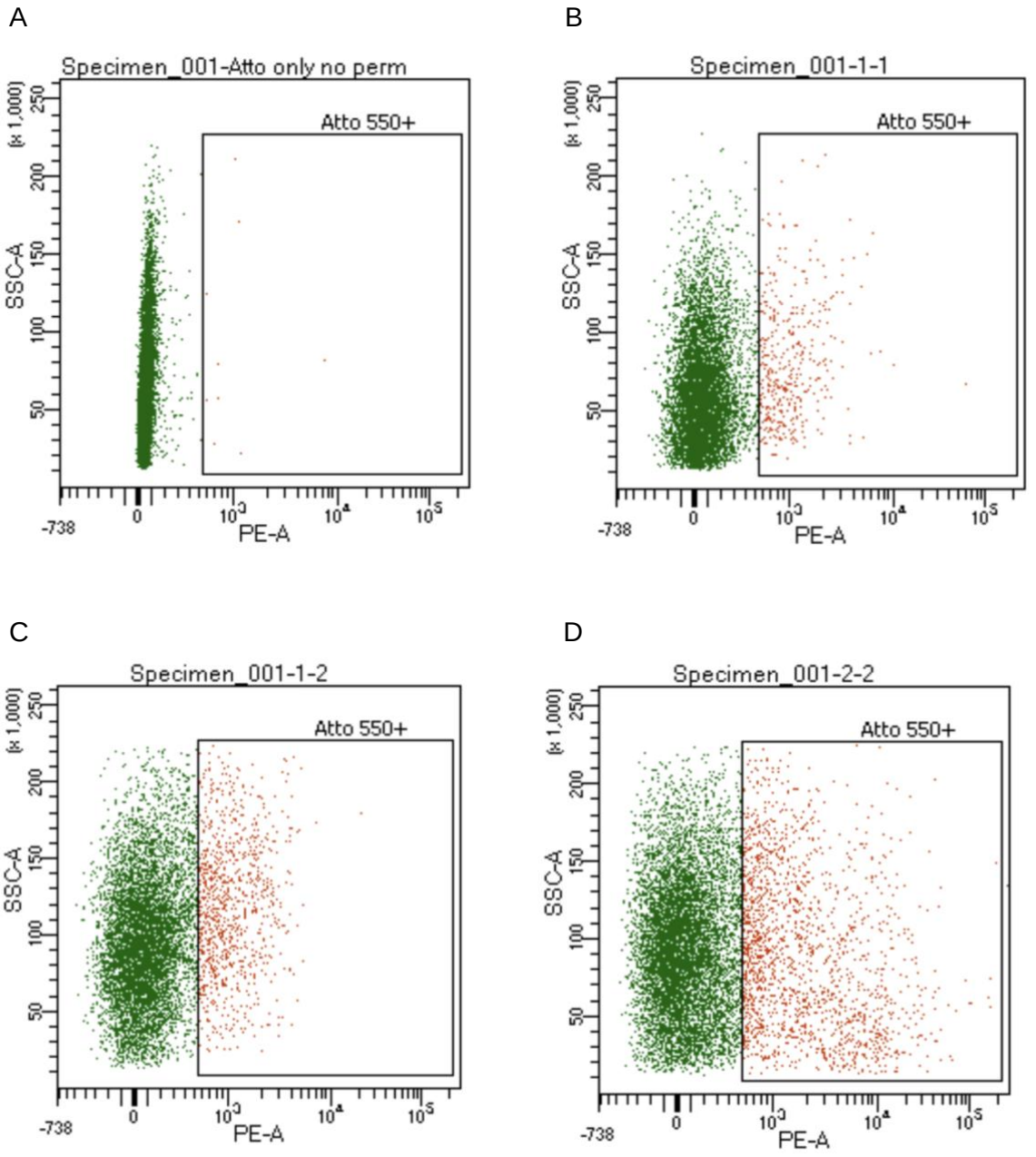

Figure 2.4. FACS dot plot analysis of fluorescently labeled NC64A cells. NC64A cells were visualized by intracellular incorporation of fluorescently labeled tracrRNA (ATTO^{TM} 550), (a) without enzyme treatment, or prior to FACS were (b) macerozyme digested, (c) vortexed with SiC whiskers, or (d) electroporated. Cells were illuminated with a white LED for bright field and a 488 nm laser for fluorescence. Images were

collected in the bright field and 480–560 nm emission (ATTO^{TM} 550 fluorescence) channels. Cell populations with a high fluorescent intensity were gated for in-focus cells and analyzed for permeability.

We have shown previously that CA-4B reacts with PBCV-1 antibody making it useful to precipitate wild-type CA-4B virus [32]. In addition, studies involving spontaneous mutant viruses have demonstrated previously that disruption of PBCV-1 *a064r* leads to viruses with an altered glycan attached to the MCP and that they are antigenically different from wild type PBCV-1 [107]. Polyclonal antisera were prepared against these spontaneous mutants. This allowed us to screen for the suspected *a064r* disrupted gene with antibody that specifically recognized viruses with this mutation. Following incubation with the serologically distinct antigenic antibody, the putative mutant precipitate was eluted, and plaque assayed (Figure 2.3a). The resulting plaques were subsequently verified by DNA sequencing of domain 1 from *034r* using appropriate PCR primers. Two isolated plaques showed separate indels at the target site supporting evidence of site-directed mutagenesis by RNP-targeting in the PBCV-1 *a064r* homologous gene, CA-4B *034r* (Figure 2.3b and 2.3c). These results demonstrate successful delivery of preassembled Cas9 protein-sgRNA RNPs that resulted in frameshift mutations in the CA-4B-encoded gene *034r*. However, despite exhausting a number of trails and alternative enzyme combinations, we were unable to duplicate these results. It is unlikely detection of the mutant virus DNA was due to erroneous sequencing given all samples were sequenced in triplicates with consistent readings.

All attempts at producing HDR using ssODNs to remove the native MscI restriction site in *034r* were unsuccessful. Amplicons of the putative GT-gene from recovered CA-4B plaques after antibody selection were incubated with MscI. Digestion of virus DNA produced DNA fragments consistent with an intact MscI site (restriction digestions not shown). The conservation of the restriction site demonstrates no evidence of ssODN-mediated HDR.

One possible explanation for obtaining the RNP-mediated NHEJ event in CA-4A DNA is macerozyme successfully eroded the alga cell wall, promoting cell permeability, allowing the entrance of the RNP-targeting cargo (Figure 2.4b). Despite targeting cell-wall polymers, the crude multi-component enzyme mixture containing pectinase, hemicellulase, and cellulase (amounts decreasing in order) did not compromise the host receptor enough to prevent virus attachment as evident by a productive virus infection. The host receptor, which is likely a carbohydrate [81], is uniformly present over the entire surface of the alga. Once inside the cell, deposited viral DNA could be recognized by the GT gene-targeting RNPs. However, this doesn't explain why SiC and electroporation procedures were unsuccessful in obtaining mutant viruses given evidence of RNP delivery inside NC64A was greater using these delivery techniques compared to macerozyme-treated cells (Figure 2.4c and Figure 2.4d). Perhaps the fluorescently labeled cells, although in greater quantity, were either dead or compromised in health that prevented the desired viral DNA editing.

2.4. SILICON CARBIDE WHISKERS

2.4.A. Introduction

We were interested in developing a simple, inexpensive technique for introducing DNA into walled algal cells. In 1990, Kaeppler [189] reported using SiC whiskers to mediate the introduction of DNA into plant cells. We adopted a technique based on the SiC method of Kaeppler that uses SiC whiskers to transform the microalgae *C. reinhardtii* with its intact cell wall [190]. This SiC method reported transformants at an efficiency of up to 10^{-5} per cell for walled cells and up to 10^{-4} per cell for *cw*-15 strains (protoplast). In contrast to agitation of the cells with glass beads, agitating the cells with SiC whiskers for up to 10 min resulted in little loss in cell viability. The gentle nature of the SiC protocol may allow more flexibility in the development of protocols for the transformation of *Chlorella* strains.

The SiC whiskers method involves the mixing (e.g., by vortex treatment) of cells in liquid medium with whiskers and exogenous DNA and/or protein. The resulting collisions between cells and whiskers appear to lead to cell penetration and DNA delivery [191]. Here, we aimed to utilize FACS again to evaluate the delivery of fluorescently labeled CRISPR-targeting cargo inside NC64A using the SiC whiskers method.

2.4.B. Materials and Methods

NC64A transfection by SiC whiskers. For each transformation, cells were mixed with sterile SiC whiskers and preassembled Cas9 protein-sgRNA RNP +/- ssODNs armed with the protospacer sequence targeting domain 1 of CA-4B *034r*. Samples were agitated by a vortex mixer at top speed for 2 min, stopping briefly every 10 s. Immediately after vortexing, samples were diluted with PEG and infected with CA-4B. Following overnight

infection, virus mutants were selected using the same antibody selection assay previously described. Virus plaques were analyzed by PCR and sequenced to detect Indel mutations by NHEJ, while *034r* amplicons were also incubated with MscI to screen for ssODN-mediated HDR.

2.4.C. Results and Discussion

FACS analysis of the SiC whiskers treated NC64A cells showed between 10-20% cell fluorescence (Figure 2.4c) suggesting positive delivery of the RNPs inside the alga cells. We speculate that cell wall perforations allowed RNP complexes brief access inside the cell. Immediately following electroporation, cells were infected with CA-4B. In theory, once viral DNA was deposited inside the alga cell, the naked nucleic acid would be available for RNP-binding and subsequent *034r* cleavage. The edited viral DNA would be subsequently replicated and packaged. Following antibody selection, we expected to recover an NHEJ indel event in CA-4B at the RNP-designated target site, however, we were unsuccessful. Although virus plaques were recovered, sequencing results confirmed the wildtype *034r* was intact. Given these sequencing results, these plaques were likely caused by escape viruses that evaded wildtype antibody binding.

2.5. ELECTROPORATION

2.5.A. Introduction

Electroporation is widely used for transformation of competent and non-competent bacteria, yeast, mammalian and microalgae cells [192]. This technique involves the use of high-intensity electric pulses to induce the permeabilization of the cell membrane, thus allowing the uptake of DNA directly from the surrounding environment. The effectiveness of transformation through this method is mediated by length of the

electric pulses, the composition of the medium, presence or absence of a cell wall, concentration of the heterogeneous DNA and strength of the electric field. The medium often demonstrates an increase in temperature after the procedure and as a result, a portion of the cells present in the suspension will be lost. Electroporation has already been successfully used for the transformation of many microalgae species such as *C. reinhardtii*, *Nannochloropsis* sp*., Dunaliella* sp*.,* and some *Chlorella* sp. not including NC64A [145].

Transformation of *Chlorella* cells via the method of electroporation alone must be performed under extreme conditions (high voltage and long pulse) that can cause damage to cells, affecting their ability to function normally. Even under such extreme conditions, mostly transient transgene expression has been achieved [193]. Widely used in bacteria, it can be used for microalgae that have a weakened cell wall [194]. In the microalgae*, C. ellipsoidea* and *C. reinhardtii*, mutant strains without cell walls are easy to transform [135, 146], suggesting that other *Chlorella* species such as NC64A can have the same transformation potential if the physical barrier of the cell wall is eliminated. The thickness of the cell wall of the free-living *C. variabilis* is between 11–17 nm and that of the symbiotic *C. variabilis* is 6–10 nm [195]. A major wall component is chitin, a glucosamine polymer not normally present in green algae likely contributes to the cell's impermeability. The total makeup of the NC64A cell wall isn't completely known, however Glc is the main structural component [186], and therefore, its dilution or elimination by use of digestive enzymes prior to electroporation could be a crucial step for reliable transformation.

2.5.B. Materials and Methods

NC64A transfection by electroporation. Purified Cas9 (100 μg) was preincubated at a 1:3 molar ratio with three *034r*-targeting sgRNAs at 37 °C for 15 min to form RNP complexes. For transfection, 250 μL cell culture (2.5×10^8 cells/mL) supplemented with sucrose (40 mM) were mixed with preincubated RNPs, and ssODNs. Cells were electroporated in 2-mm cuvettes (600 V, 50 μF, 200 Ω) by using Gene Pulser Xcell (Biorad). Immediately after electroporation, 800 μL of MBBM with 40 mM sucrose was added to the sample and the cells were infected with CA-4B (MOI 5). Following overnight infection, virus mutants were selected using the same antibody selection assay previously described. Virus plaques were analyzed by PCR and sequenced to detect Indel mutations by NHEJ, while *034r* amplicons were also incubated with MscI to screen for ssODN-mediated HDR.

2.5.C. Results and Discussion

FACS analysis of the electroporated NC64A cells showed the highest cell fluorescence of all delivery methods (approximately 20%) suggesting positive delivery of the RNPs inside the alga cells (Figure 2.4d). We speculate that electroporation-induced cell wall perforations allowed RNP complexes brief access inside the cell. Immediately following electroporation, cells were infected with CA-4B. In theory, once viral DNA was deposited inside the alga cell, the naked nucleic acid would be available for RNP-binding and subsequent *034r* cleavage. The edited viral DNA would be subsequently replicated and packaged. Following antibody selection, we expected to recover an NHEJ indel event in CA-4B at the RNP-designated target site, however, we were unsuccessful. Although virus plaques were recovered, sequencing results confirmed the wildtype *034r*

was intact. Given these sequencing results, these plaques were likely caused by escape viruses that evaded wildtype antibody binding.

2.6. CELL-PENETRATING PEPTIDES

2.6.A. Introduction

CPPs are short peptides that facilitate cellular intake and uptake of molecules. This group of short peptides with at most 30 to 35 amino acid residues, has the ability to translocate across the cell membrane [196]. In animal system, the direct protein delivery strategy using CPPs has been extensively developed [197-199]. However, the peptide-mediated protein delivery system in plants and algae are still scarcely developed. CPPs are of different sizes, amino acid sequences, and charges but all CPPs have one distinct characteristic, which is the ability to translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle; thus, they act as molecular delivery vehicles. There has been no real consensus as to the mechanism of CPP translocation, but the theories of CPP translocation can be classified into three main entry mechanisms: direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure.

Previous studies demonstrated that CPP-fused to a polycationic peptide (an alternate copolymer of lysine and histidine, $(KH)_9$), which was termed as fusion peptides, is a more efficient nucleic acid carrier compared with that of the CPP alone in intact plants [200]. By using the fusion peptides, the negatively charged cargo preferentially interacts with the polycationic peptide through ionic interactions, whereas the CPP interacts with fewer cargo molecules and is preferentially present on the surface of peptide-cargo complexes. The higher number of CPPs at the surface of the complexes

leads to a higher efficiency in gene delivery [201]. Based on the concept of fusion peptide, two protein delivery carriers, $(BP100)_2K_8$ and $BP100(KH)_9$, were designed by fusing a CPP, i.e., BP100 or BP100 dimer, and a polycationic peptide, i.e., alternate copolymer of lysine and histidine, $(KH)_9$ or eight consecutive lysine residues, K_8. The polylysine is known to destabilize cell-membrane via electrostatic interaction between the protonated amine and the negatively charged cell membrane [202]. On the other hand, the copolymer of lysine and histidine has the ability to promote the buffering effect on the prelysomal vesicle [203]. In this current study, we assumed that the polycationic peptides [$(KH)_9$ and K_8] also play an important role in the electrostatic interaction with negatively charged protein cargo, similarly to the ionic interaction with plasmid DNA as described in previous studies [200, 204].

2.6.B. Materials and Methods

Peptides. Peptides were gifted by Dr. Heriberto Cerutti (UNL): $(BP100)_2K_8$ (KKLFKKILKYLKKLFKKILKYLKKKKKKKK, theoretical pI/Mw: 10.75/3851.13 Da) and $BP100(KH)_9$ (KKLFKKILKYLKHKHKHKHKHKHKHKHKH, theoretical pI/Mw: 10.81/ 3809.71 Da).

Complex formation. Peptide/RNP complexes were prepared by adding different amounts of each peptide to RNP at various ratios (0.1, 0.5, 1, 2, 5, 10, and 20). The solution was thoroughly mixed by repeated pipetting and allowed to stabilize for 30 min at 25 °C.

NC64A transfection by CPPs. Referring to a previous study [173], 10 μg of Cas9 and 12 μg of sgRNA were incubated at room temperature for 15 min. A prepared cell sample of 100 μl at 3×10^8 cells/mL was added to the incubated RNP, ssODNs, and gently

mixed. Tested independently, $(BP100)_2K_8$ or $BP100(KH)_9$ was added to the sample and mixed immediately. In a series of experiments, we observed that there was no significant difference between the two types of treatment with CPPs: (1) pre-incubation of RNP and CPP for the formation of complex, and (2) treatment of CPP and RNP to the cell simultaneously. After incubation of the cells mixed with the RNP and CPP for 30 min at 25 °C, trypsin was added, and the mixture was incubated for 15 min at 37 °C. The sample was washed by MBBM media and transferred to 10 mL of MBBM media and incubated for 16 h under dim light without shaking as a "recovery" step. Following incubation, cells were infected with CA-4B (MOI 5) and incubated overnight (24 h). Following overnight infection, virus mutants were selected using the same antibody selection assay previously described. Virus plaques were analyzed by PCR and sequenced to detect Indel mutations by NHEJ, while *034r* amplicons were also incubated with MscI to screen for ssODN-mediated HDR.

2.6.C. Results and Discussion

We examined the effect of parameters, such as the peptide/RNP ratio and serum addition, on peptide-mediated transfection. Properties of peptides (DNA binding stability and condensation capacity) and of peptide/RNP complexes (size and surface charge) were investigated because these are known to vary as a function of the peptide/RNP ratio. Following antibody selection, we expected to recover an NHEJ indel event in CA-4B at the RNP-designated target site. However, we were unsuccessful. Although virus plaques were recovered, sequencing results confirmed the wildtype *034r* was intact. Given these sequencing results, these plaques were likely caused by escape viruses that evaded wildtype antibody binding.

2.7. *AGROBACTERIUM*-MEDIATED

2.7.A. Introduction

Agrobacterium tumefaciens is a gram-negative bacterium species found in soil, more commonly known for infecting a wide variety of dicotyledonous plants, causing the growth of tumors on their stems near the surface of the soil (a disease referred to as crown gall disease). The infection is triggered by the release of phenolic compounds from wounds in plant roots and is controlled by the "*vir* regulon" present in the Ti plasmid. This leads to the transfer of the T-DNA from the bacterium's Ti plasmid (tumor inducing plasmid) into the plant cells where it permanently integrates into the host genome causing the aforementioned tumors [133]. It was later discovered that disarmed Ti plasmids could be used to deliver foreign DNA into plant cells in order to produce genetically engineered plants with improved features [205].

More recently it was demonstrated that this same system could also be used to transform microalgae [133]. This technique holds several advantages over other transformation methods. Not only does it allow the transfer of large DNA molecules with little rearrangements, but also demonstrates a high rate of integration into potentially transcribed regions and usually leads to the integration of a singular transgene in the target's genome [141]. In contrast with other transformation procedures, transformation frequency and gene stability do not appear to be frequent complications and can also result in homologous recombination facilitating the localization of the integrated DNA segments. *Agrobacterium tumefaciens*-mediated transformation is a form of conjugation as it also involves the physical contact between the organisms involved in the process.

Kumar *et al.* (2004) [147] were able to implement this method for the first time in an algal system through the nuclear transformation of *C. reinhardtii*. When compared with other approaches like the glass bead method, transformation mediated by *Agrobacterium* was shown to have an increase in transformation frequency of 50-fold when transforming *C. reinhardtii*. In addition, it allows for the nuclear transformation of microalgae without the need for the removal of the cell wall, a trait only shared with particle bombardment.

These aforementioned alga transformation protocols except particle bombardment often utilize cell wall less strains or require an enzymatic step to facilitate the removal of the cell wall prior to transformation. In light of these stepwise procedures, we were interested in the development of a transformation system that doesn't require alga cell wall treatment. Here, we transferred T-DNA of *A. tumefaciens* strain LBA4404 harboring the binary vector pCAMBIA1304 containing the gfp:gusA fusion reporter, a hygromycin phosphotransferase (*hpt*) selectable marker driven by the CaMV35S promoter, and the transgene a064r-D1 from PBCV-1 antigenic mutant EPA-1. NC64A Transformation was assessed by monitoring transient *β*-glucuronidase (GUS) and *gfp* expression 2 days p.i. In theory, if NC64A transformation is achieved, the hygromycin-resistant recombinant alga expressing the truncated form of A064R (domain 1 only per nonsense mutation) would produce some shortened EPA-1 glycans at the MCP surface of newly formed viruses when infected with wild-type PBCV-1. In turn, these virus mutants would be subject to antibody selection as described above.

2.7.B. Materials and Methods

***Agrobacterium* strain and vectors.** The binary vector pCAMBIA1303 containing a gfp:gusA fusion reporter, domain 1 of gene *a064r* from a PBCV-1 antigenic mutant referred to as EPA-1, and a selectable marker for hygromycin B resistance driven by the CaMV 35S promoter were used for transformation. The binary vector was mobilized into *A. tumefaciens* strain LBA4404 by using the Gene Pulser Xcell electroporator (Biorad) according to manufacturer's protocol. Transformed cells were aliquoted and maintained at -80°C in 25% (v/v) glycerol.

Antibiotic sensitivity test. The sensitivity of *A. tumefaciens* towards the antibiotic cefotaxime was tested by inoculating 200 μL of *Agrobacterium* culture (OD_{600} = 1.0) in 5 mL LB broth supplemented with varying concentrations of cefotaxime (0, 50, 100, 150, 200, 300, 400 and 500 mg/L) and the growth of *Agrobacterium* in each concentration was measured spectrophotometrically at OD_{600} after 2 days. The effect of the antibiotic cefotaxime on the viability of NC64A was accessed by plating a serially-diluted microalgae culture on solid MBBM supplemented with different concentrations of cefotaxime (0, 100, 200, 300, 400 and 500 mg/L). The agar plates were incubated in the dark for 2 days at 25 °C before exposure to light and the number of surviving colonies from the dilution that produced less than 100 colonies was counted in duplicates after 2 weeks. In order to determine the minimum inhibitory concentration of hygromycin B, 1 x 10^6 NC64A cells were plated on solid MBBM medium supplemented with 500 mg/L cefotaxime and varying concentrations of hygromycin B (6, 8, 10, 12, 14, 16, 18, 20, 23 and 26 mg/L) and each treatment was tested in triplicates. The agar plates were incubated for 2 days in the dark at 25 °C before exposure to light and the number of surviving colonies was accessed after 20 days.

NC64A transfection by *Agrobacterium*. A general transformation procedure for NC64A was established based on work done by Kumar *et al.* (2004) [133] on transformation of *C. reinhardtii* with some adjustment as described here. Single colonies of *Agrobacterium* initiated from a frozen stock were used to inoculate 10 mL of LB supplemented with 5 mM Glc, 100 mg/L streptomycin and 50 mg/L kanamycin and grown overnight in a rotary shaker at 27 °C with shaking at 200 rpm in the dark. Five mL of this overnight culture was used to inoculate 50 mL of the same medium and it was grown in the dark at 27 °C with shaking at 200 rpm until OD_{600} = 0.8–1.2. The bacterial culture was harvested by centrifugation and washed once with induction medium (MBBM plus 150 μM acetosyringone, pH 5.6) and diluted to a final density of OD_{600} = 0.5. Prior to co-cultivation, a total of 5 x 10^6 NC64A cells from a log-phase culture (OD_{600} = 0.5–1.0) were pre-cultured for 5 days on MBBM agarose plates at 25 °C and harvested with induction medium on the day of co-cultivation. The algal cell pellet was mixed with 200 μL of the bacterial suspension and plated on induction medium solidified with 1.2% (w/v) bacto-agar. Co-cultivation was performed for 3 days at 25 °C in the dark. Following co-cultivation, cells were harvested with MBBM supplemented with 500 mg/L cefotaxime in a total volume of 7 mL and incubated in the dark at 25 °C for 2 days to eliminate *Agrobacterium*. Visualization of *gfp* expression was performed using a confocal fluorescent microscope (Nikon A1R-Ti2 confocal system, UNL Microscopy Core). Subsequently, the remaining cells were plated on selective media containing 20 mg/L hygromycin B and 500 mg/L cefotaxime and incubated at 25 °C in the dark for 2 days before exposure to light. Resistant colonies were propagated on non-selective media

and utilized for PCR analysis. Detection of contaminating *Agrobacterium* was performed by growing cells on LB agar plates for at least 7 days at 25 °C in the dark.

2.7.C. Results and Discussion

A. tumefaciens was suppressed at a cefotaxime concentration as low as 50 mg/L (Figure 2.5a) whereas the growth of NC64A was found to be uninhibited in cefotaxime-supplemented media up to 500 mg/L (Figure 2.5b). Thus, the highest concentration (500 mg/L) of cefotaxime was selected for all experiments to ensure thorough elimination of *Agrobacterium* post-transformation. The lowest concentration of hygromycin B which completely inhibited the growth of NC64A was 20 mg/L (Figure 2.5c), and this concentration was used for subsequent selection of transformants.

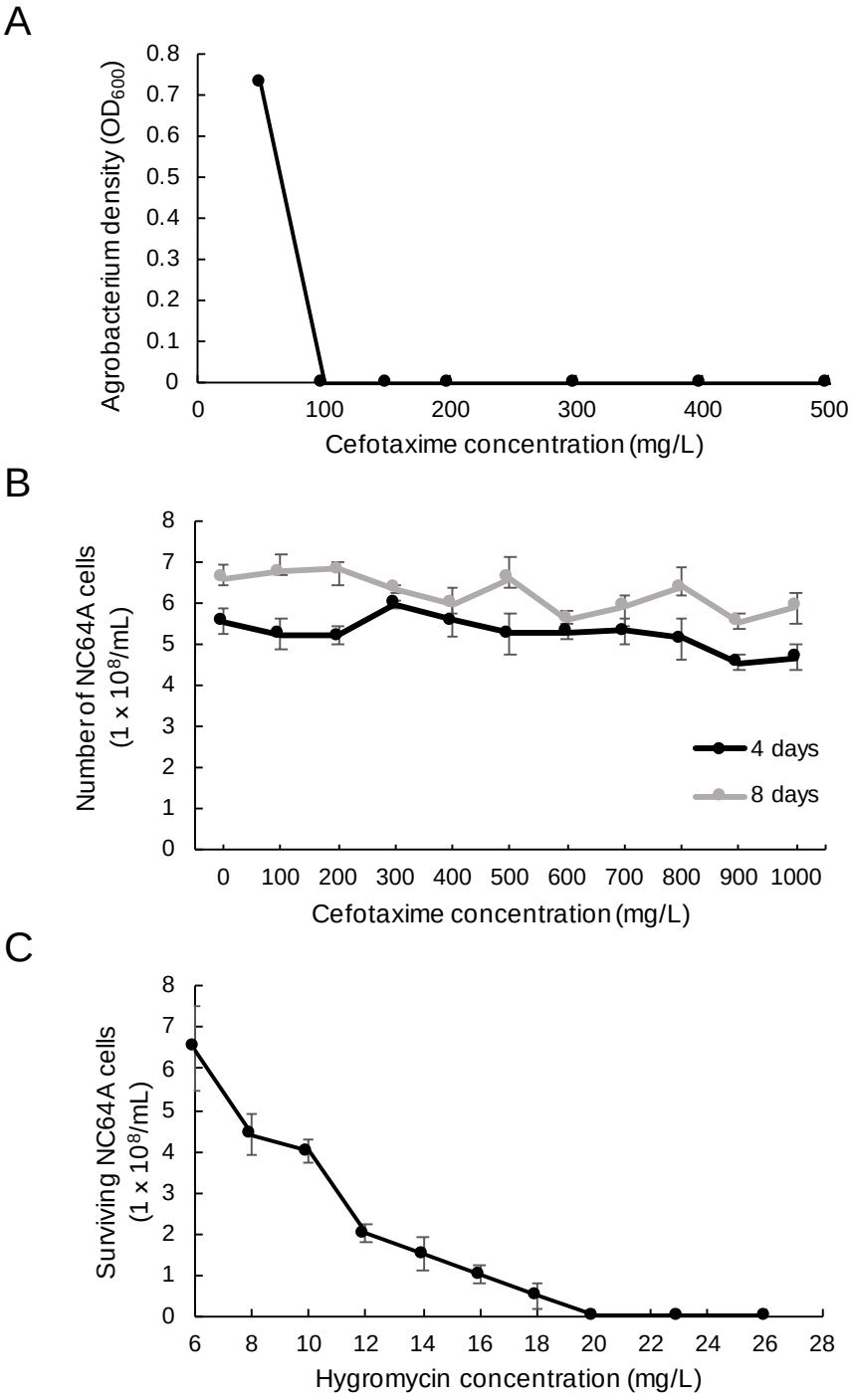

Figure 2.5. The effects of antibiotics on *Agrobacterium* and NC64A. (a) The effect of cefotaxime on *Agrobacterium* viability. The effect of (b) cefotaxime and (c) hygromycin on NC64A viability.

Colony PCR of *Agrobacterium* colonies post electroporation with the recombinant pCAMBIA1303 validated the binary vector was successfully mobilized LBA4404 (Figure 2.6a). A total of eight putative hygromycin-resistant single colonies that appeared within 20 days on selection media (containing 20 mg/L hygromycin and 500 mg/L cefotaxime) were randomly selected and grown in liquid media (containing 500 mg/L cefotaxime but lacking hygromycin) before the DNA was extracted and used in PCR analysis. Amplification with EPA-1 *a064r* domain 1 gene-specific primers successfully detected the 633 bp *a064r* gene fragment (25 kDa) from seven putative transgenic lines (Figure 2.6a), which represents 87.5% of the total number of screened colonies. To rule out *Agrobacterium* contamination, amplification with Kan^R gene-specific primers only produced the expected 795 Kan^R gene in LBA4404 (positive control) but not in the seven putative NC64A transgenic lines (Figure 2.6b). The presence and absence of these gene-specific fragments in seven putative transgenic lines with no amplification product detected in non-infected wild-type NC64A indicated the successful transfer of both the hygromycin marker and the GFP-GUS reporter gene as a single T-DNA unit flanked by the left and right borders.

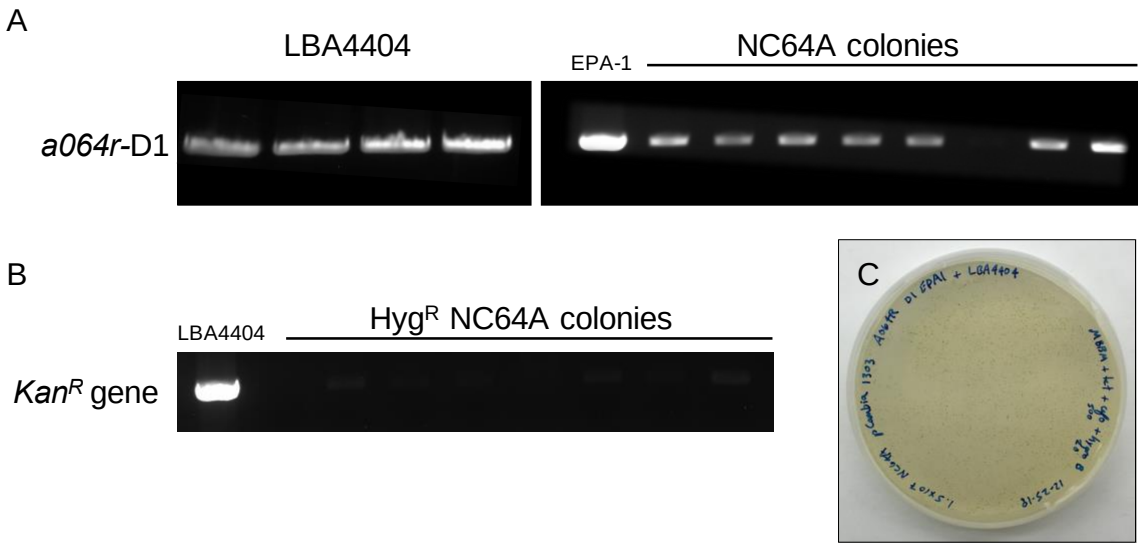

Figure 2.6. PCR analysis of putative NC64A transformants transformed with recombinant pCAMBIA1303 *a064r* domain 1. (a) Amplification of the 633 bp fragment of the *a064r* domain 1 gene fragment from four LBA4404 colonies post-electroporation, and eight NC64A colonies after agro-mediation transfection. (b) Amplification of the *Kan*^{R} gene from LBA4404 (positive control) but not in the seven putative NC64A transgenic lines. (c) Transgenic NC64A colonies on selective media (containing 20 mg/L hygromycin and 500 mg/L cefotaxime).

The identity of the PCR amplicons derived from transgenic NC64A colonies (Figure 2.6c) were further confirmed to be positive by DNA sequencing where sequence alignment showed 99 and 98% identity to the *hpt* and gfp-gusA genes in pCAMBIA1303 respectively (Gene Bank accession no. AF234300). Visualization of GFP using a fluorescent microscope confirmed weak expression of the *gfp* gene in transformed cells as bright green fluorescence (Figure 2.7). The failure of many hygromycin-resistant colonies to produce any amplification product for either primer pair indicates the incidence of 'escapes' through antibiotic selection. The seven transgenic NC64A lines

that were previously confirmed positive for *a064r* domain 1 by PCR could not be maintained or subcultured on hygromycin-supplemented media. No GUS-positive cell was detected in transgenic lines post-selection with hygromycin, although PCR amplification confirmed the presence of the gfp-gusA genes (data not shown). This may indicate the possibility of the gfp-gusA genes silencing in transgenic lines.

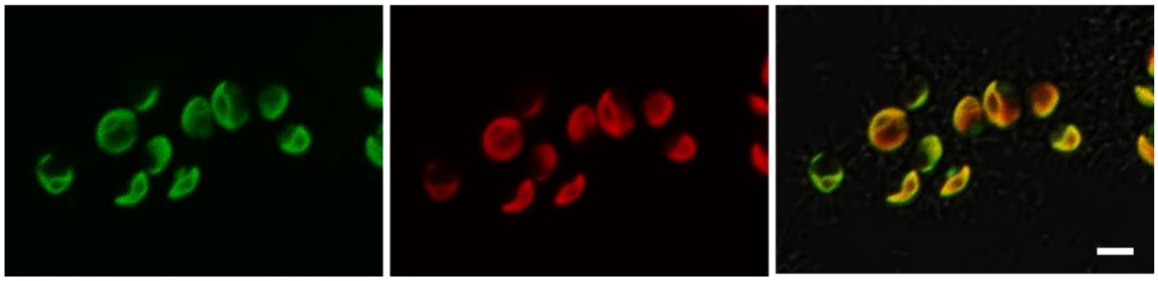

Figure 2.7. Transformation of NC64A with pCAMBIA1304 vector carrying the gfp–gusA fusion reporter genes. Confocal fluorescent microscopy of live NC64A cells transformed with pCAMBIA1304 by *Agrobacterium tumefaciens* imaged with green, red, and GFP channels (from left to right). The GFP florescent channels shows the appearance of *gfp* expression. Bar represents approximately 5 μm.

In summary, a potentially simple method for *Agrobacterium*-mediated transformation of *Chlorella variabilis* NC64A was developed. Seven PCR-positive transformants were obtained, and the presence of the viral gene fragment suggests that the T-DNA was integrated into the genome; however, more extensive studies are needed to elucidate possible factors and mechanisms contributing to the loss of the gusA-gfp

expression. Other vector systems and/or promoters might also need to be tested on this microalga to develop a better transformation system. Nonetheless, the finding opens the possibility of further genetic manipulation of this commercially important microalga with other genes of interest.

CHAPTER III

PBCV-1 PROTEINS A064R AND A061L TOGETHER HAVE FOUR TRANSFERASE ACTIVITIES INVOLVED IN GLYCAN SYNTHESIS

3.1. OVERVIEW

PBCV-1 encodes seven putative sugar manipulating enzymes with most of them predicted to be involved in the synthesis of its N-glycan(s) as disclosed by bioinformatic analyses [97, 111]: A064R (638 amino acids), A111/114R (860 aa), A219/222/226R (677 aa), A473L (517 aa), A546L (396 aa), A071R (354 aa), and A075L (280 aa). However, this number is less than the 10 sugars present in the N-glycan attached to the MCP. Several possibilities could explain this discrepancy: 1) One or more of the seven putative GTases might have multiple functional domains, making this restricted repertoire of enzymes sufficient to synthesize these structures; 2) PBCV-1 genes encode enzymes that do not resemble GTases in the databases and, hence, they are not recognized during query searches; 3) a host-encoded GT(s) could contribute to the glycosylation process.

Among the genes mentioned above, the protein coded by gene *a064r* is of interest because it is found in PBCV-1 but not in other chloroviruses where the glycan structures are known [51, 107, 206]. Also, experimental evidence shows that isolated spontaneous mutants (or antigenic variants) of PBCV-1 with mutations in *a064r* have truncated glycan structures [111, 113]. A combination of genetic and structural analyses suggested that the A064R protein has three putative domains of ~200 amino acids each (SI Appendix, Fig. S1 [175]). Domain 1 (A064R-D1) was predicted to encode a β- L-rhamnosyltransferase and domain 3 (A064R-D3) was predicted to encode a S-adenosyl-L-methionine-

dependent MT (SAM)-dependent MT that decorates two positions in the terminal α-L-Rha unit [111]. The A064R domain 2 (A064R-D2) was hypothesized to encode an α-L-rhamnosyltransferase, despite only resembling bacterial proteins with unknown functions, thus, suggesting that it could be a new GT family.

Here we provide evidence that the predictions about the three A064R domains are correct with the exception of A064R-D3, which adds only one methyl group and not two as proposed. It methylates O-2 of the α-L-Rha residue, while O-3 methylation is accomplished by another virus-encoded protein A061L.

3.2. GLYCOSYLTRANSFERASE A064R AND METHYLTRANSFERASE A061L

The chloroviruses are unusual because they are predicted to encode most, if not all, of the machinery to synthesize the glycans attached to their MCP. Here we show that two of the virus-encoded proteins A064R and A061L are functionally active. A064R has three domains: The first two are GTases and the third domain is a MT. A061L has a MT activity. The action of these two enzymes produce the fragment 2,3-di-O-methyl-α-L-Rha-(1→2)-β-L-Rha, which is part of the complex N-linked glycan attached to the virus capsid protein. A064R domain 2 is a member of a new GT family. This provides direct evidence that the synthesis of PBCV-1 glycans are accomplished with virus-encoded enzymes.

3.2.A. Materials and Methods

Recombinant proteins. All recombinant proteins were expressed in *E. coli* cells. The A064R full-length protein was produced with a C-terminal 6xHis-tag (SI Appendix, Fig. S1 A and B [175]) using the pET23a vector as described [207]. Protein purification to homogeneity was performed using the Probond nickel-chelating resin (ThermoFisher,

code R801-01), following the manufacturer's protocol. A064R domains (SI Appendix, Fig. S1 A and B [175]) were cloned and expressed in the pGEX-6P1 vector as described [207] obtaining fusion proteins with the glutathione S-transferase (GST) tag at the N terminus, that were purified using glutathione Sepharose 4B (GE Healthcare, code no. 17–0756-01). The glutathione-Sepharose-bound proteins were used for the enzymatic reactions. Cleavage was achieved for A064R-D1 only by using Prescission protease (GE Healthcare, code no. 27–0843-01) on the column, and the released protein was concentrated using Amicon Ultra-4 (Millipore). The released domain was used for the UDP-Glo™ assay. Purity of full-length A064R-6xHis protein and of the GST-fusion domains was determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE) and Coomassie blue staining (SI Appendix, Fig. S1C [175]). A061L protein (SI Appendix, Fig. S1 [175]) was cloned and expressed in pGEX-5X-1 expression vector (GE Healthcare), producing a GST tag at the N terminus of the target protein. The recombinant protein was purified on glutathione Sepharose as before and used for enzymatic reactions (more details in SI Appendix, Materials and Methods [175]). All eluted proteins were resolved by SDS/PAGE with Coomassie brilliant blue staining solution.

Enzymatic reactions. All of the reactions were performed at 25 °C in phosphate buffer saline ([PBS]; Sigma-Aldrich, code no. P4417, composition: 0.01- M phosphate buffer, 0.0027-M potassium chloride, and 0.137-M sodium chloride, pH 7.4) by using the enzyme still attached to the resin through its GST tag or it's His tag in the case of the whole A064R. In a typical experiment, the reaction volume was about 150 μL and included 50 μL of resin (glutathione Sepharose 4B GE Healthcare or Probond nickel-

chelating resin) with ~250 μg of protein adsorbed unless otherwise specified. All enzymatic reactions were screened via high-pressure liquid chromatography (HPLC), using a C18 column (Phenomenex Kinetex 5 μ, 2,50 × 4.60 mm, code no. 00G-4601-E0) and 70% methanol as an eluent (flow rate at 0.8 mL/min) by injecting 10 μL of the crude reaction after centrifugation. For A064R-D1, the reaction was performed by using oligosaccharide 1 (1.5 mM) (Figure 3.1) as an acceptor and UDP-β-L-Rha (1.5 mM) as a donor in the presence of Mn^{2+} (2 mM). In order to verify the cation dependence of the enzyme, the reaction was repeated by adding EDTA (1 mM) to the protein 30 min prior to the addition of all of the other reagents except the cation. For A064R-D2, three different constructs were produced and tested: D2 (191–405 aa), D2L (191–438 aa), and $D2L_2$ (213–438 aa). All three constructs were tested with oligosaccharide 2 (1.5 mM) as an acceptor and UDP-β-L-Rha (1.5 mM) and with or without cations (Mg^{2+} and Mn^{2+}, 2 mM). To exclude any cation dependency, a reaction with D2L was also tested by adding EDTA (0.5 mM), keeping the other reagents unchanged. The D2L reaction was repeated replacing acceptor oligosaccharide 2 with the L-Rha monosaccharide (1.5 mM). The resin with A064R-D1D2 was suspended in PBS buffer and incubated with oligosaccharide 1 (1.5-mM), UDP-β-L-Rha (3.6 mM) and Mn^{2+} (2 mM). In the second reaction, oligosaccharide 6 (1.5 mM) was used instead of oligosaccharide 1, while all other components were kept unchanged. The resin with the full-length A064R attached was suspended with oligosaccharide 1 (1.5 mM), UDP-β-L-Rha (3.4 mM), the precursor for the MTs SAM (CAS no. 86867–01-8; code no. A7007, Microtech), and with both Mn^{2+} and Mg^{2+} (2 mM each). As for A061L, two reactions were performed by varying the nature of the acceptor. In the first reaction, oligosaccharide 3 (1 mM) was used as an

acceptor along with an excess of SAM (ca. 2 mM), Mn^{2+} and Mg^{2+} (2 mM each); in the second reaction, oligosaccharide 3 was replaced with oligosaccharide 4.

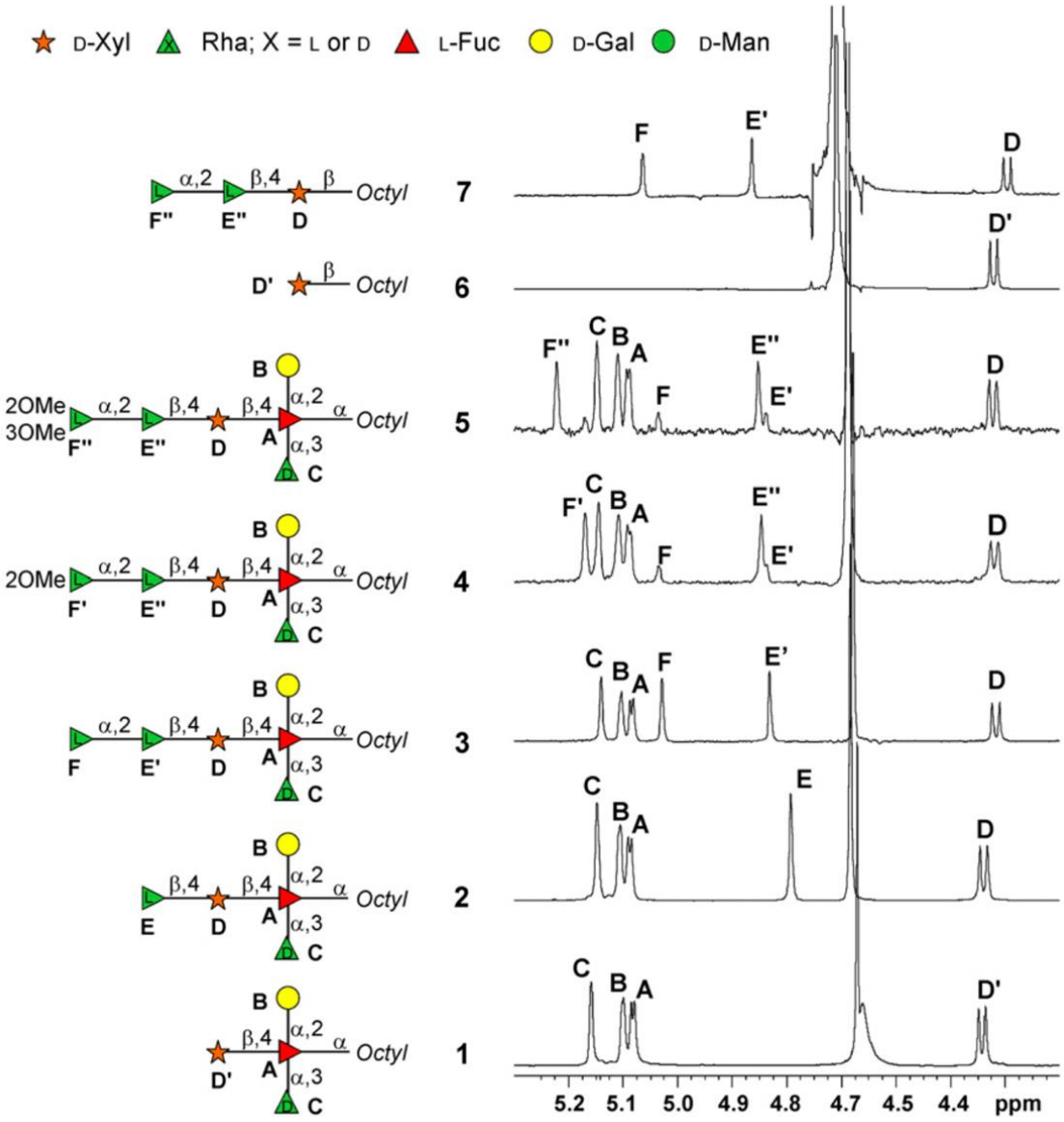

Figure 3.1. (600 MHz and 310 K, D_2O) ^{1}H NMR spectra detailing the anomeric region of the oligosaccharides 1–7, drawn as a cartoon next to each spectrum. Each monosaccharide unit is labeled with a capital letter, and the full NMR attribution of each

compound is reported in the SI Appendix [175]. The peak at ca. 4.7 ppm is the residual solvent signal.

Compounds isolation. The crude reaction solutions were centrifuged using spin columns with collection tubes (code no. H6787, Sigma-Aldrich) to remove the resin, and then, the pure products were recovered by C18 Sep-Pak cartridges (pore size 125 Å, particle size 55–105 μm, code no. WAT051910, Waters) as detailed below. The Sep-Pak cartridge was activated with 10 mL of ethanol, 4 mL of acetonitrile, and 10 mL of water. After being loaded with the supernatant, the following elution was performed: 20 mL of water, 15 mL of acetonitrile/water 1:4, 5 mL of acetonitrile, and 20 mL of ethanol. All fractions were dried and analyzed via NMR spectroscopy.

Bioluminescent assay: UDP-Glo™ assay. Two UDP standard curves were prepared in a 96-well plate, performing a serial twofold dilution as described in the UDP-Glo™ kit (code no. V6961, Promega), obtaining 12 solutions at different UDP concentrations. One reaction contained Mn^{2+} the other contained Mg^{2+} at the same concentration (40 mM). Thereafter, 25 μL of each solution was transferred to a second 96-well plate, and 25 μL of UDP detection reagent (provided by the UDP-Glo™ kit) was added; the solutions were incubated at 25 °C for 1 h, and luminescence was measured with a luminometer. Reactions that involved the A064R-D1 enzyme, oligosaccharide 1 as an acceptor and UDP- β-L-Rha as a donor in separate phosphate buffer solutions (one with Mg^{2+} and the other with Mn^{2+}), were performed maintaining a constant ratio of UDP-L-Rha/1 (2:1 ratio; 2.470 and 1.250 nmol, respectively), while the GT amount decreased from the first solutions to the last where GT was absent. Similar to the UDP standard curves, the

solutions at different GT concentrations were obtained using twofold serial dilutions: 100 µL of the GT solution (5 µL of GT 150 µg/mL; 5 µL Mn^{2+}/or Mg^{2+}, 40 mM; 4.75 µL PBS 20×; 85.25 µL of water) was added in the first well followed by the addition of 50 µL buffer (composed of Mn^{2+}/ Mg^{2+}, 40 mM; PBS 20×, and H_2O) to the additional 11 wells. Then, 15 µL from each well was transferred into 12 wells of another 96-well plate, and 10 µL of the solution containing UDP-L-Rha (3.3 mM) and oligosaccharide 1 (1 mM) was added. The reactions were incubated at 25 °C for 1 h, after that 25 µL of the UDP-detection buffer was added to each well, and the solutions were kept at the same temperature for one additional hour prior to measuring luminescence. Two replicas for each experiment were performed. The same strategy was used for the following two reactions, in which the activated sugar donor or the acceptor substrate were changed, maintaining the 2:1 ratio as before: in one, the UDP-L-Rha was substituted with the UDP-D-Glc (10 mM), and in the other, the D-Xyl monosaccharide (1 mM) was used instead of oligosaccharide 1. Experiments were performed in the presence of both bivalent cations. Luminescence measurements were performed with a Synergy HT multimode microplate reader (BioTek instrument) using a 96-well microplate with standard 128 × 86-mm geometry with an integration time of 1.0 s. Luminescence was measured by a low-noise photomultiplier detector through an empty filter position in the emission filter wheel. Data analysis and the construction of the graphs were achieved with Excel Software.

NMR spectroscopy. All NMR experiments of the oligosaccharides from 1 to 7 in Figure 2 were recorded in D_2O at 310 K on a Bruker DRX-600 MHz (^1H: 600-MHz and 13C: 150-MHz) instrument equipped with a cryoprobe except for the Rha disaccharide for which a temperature of 298 K was used. All chemical shifts are referred to internal

acetone (^{1}H 2.225 and 13C 31.45 ppm). The set of 2D spectra (correlation spectroscopy, TOCSY, T-ROESY, and HSQC) were measured for each substrate, except for oligosaccharides 3, 7, and the α-L-Rha-(1→2)-L-Rha disaccharide for which only the heteronuclear single quantum correlation (HSQC) and the total correlation spectroscopy (TOCSY) experiments were acquired and for structures 1 and 2 where HMBC was also recorded. Homonuclear experiments were recorded using 512 free induction decays (FIDs) of 2,048 complex data points, setting 24 scans per FID for all experiments, whereas for structures 3 and 4, 32 scans per FID were set in order to obtain a better signal to noise ratio. Mixing time of 100 ms was applied for TOCSY and 300 ms for T-ROESY spectra acquisitions. ^{1}H-13C heteronuclear experiments were acquired with 512 FIDs of 2,048 complex points with 40–80 scans per FID, depending on the sample abundance. Standard Bruker software Topspin 3.1 was used to process and analyze all spectra. NMR operating conditions used for the characterization of the synthetic intermediates of oligosaccharides 1 and 6 are reported in the SI Appendix [175].

Phylogenetic analyses. Multiple sequence alignment and phylogenetic reconstructions were performed using the Environment for Tree Exploration3 (ETE3) v3.1.1 program [208] as implemented on the GenomeNet (https://www. genome.jp/tools/ete/). Phylogenetic trees were constructed using maximum likelihood and neighbor-joining approaches using PhyML v20160115 ran with model pa- rameters: -f m–pinv e -o tlr–nclasses 4–bootstrap 100–alpha e [209]. The branch supports are the χ2-based parametric values return by the approximate likelihood ratio test. Maximum-likelihood annotation was generated by Interactive Tree of Life (iTOL) [210]. Viral proteins were used as queries to search the publicly available NCBI nonredundant sequence database using

BLASTp and hits filtered based on general parameters excluding chloroviruses. The characterized GT domains were identified using UniProt [211] and HHpred [212] from selected rhamnosyltransferases in the CAZy database.

Protein modeling. PBCV-1 A064R-D3 and A061L structure predictions were constructed by Phyre2 [213]. Molecular docking of viral proteins complexed with SAM were performed using PatchDock [214], an algorithm based on shape complementarity principles. Protein complexes were inputted into the molecular modeling system, Chimera where 3D protein models were con- structed. SAM-binding residues were generated by SAMbinder [215, 216]. Final drawings and residue analysis were prepared with molecular graphics system PyMol, a molecular graphics system, Version 1.2r3pre, Schrödinger, LLC.

Data availability. All study data are included in the article and supporting information and the UniProt Knowledgebase (UniProtKB), https://www. uniprot.org/ (accession no. A0A516L7Y9).

3.2.B. Results and Discussion

A064R-D1: β-L-Rhamnosyltransferase activity. A064R-D1, amino acid residues 1–212 in the PBCV-1 A064R protein (SI Appendix, Fig. S1A [175]) was originally predicted to be a GT in subfamily 34 (GT34; retaining) [217]. Thereafter, A064R-D1 was expressed as a recombinant protein and crystallized confirming it to be a member of the GT34 subfamily and preferring uridine 5′-diphosphate (UDP)-Glc as a donor substrate. However, this study occurred before the PBCV-1 glycan structure(s) was known, and this first hypothesis was later refuted by Speciale *et al.* [111] who predicted that A064R-D1 added L-Rha to D-Xyl.

Here, we biochemically evaluated this prediction by using the tetrasaccharide 1 (Figure 3.1) as the acceptor substrate. This oligosaccharide is a simplified truncated version of the N-glycan from PBCV-1 in which the D-Xyl is available for glycan elongation. In contrast to the natural glycan, the Fuc in 1 is capped at the reducing end with an octyl group (Figure 3.1 and SI Appendix, NMR Characterization, Figs. S2 and S3, and Table S1 [175]), facilitating the monitoring of the reaction via HPLC and the purification of the products. UDP- β-L-Rha was used as the donor, and the reaction was monitored by either the HPLC or by the UDP-Glo^{TM} assay. When the reaction mixture included Mn^{2+}, the HPLC profile (Figure 3.2) showed the formation of pentasaccharide 2 while oligosaccharide 1 simultaneously decreased. NMR analysis of oligosaccharide 2 revealed an additional signal at 4.79 parts per million (ppm), labeled E, in the anomeric region (Figure 3.1). Comparison of the HSQC spectrum of oligosaccharide 2 (SI Appendix, Fig. S4 [175]) with that of oligosaccharide 1 identified the three common residues: namely, A (2,3,4-substitued α-L-Fuc), B (terminal α-D-Gal), and C (terminal α-D-Rha). The β-D-Xyl unit (D) showed some chemical shift variations because of glycosylation at O-4 as inferred by the downfield displacement of the corresponding carbon chemical shift (79.4 ppm, SI Appendix, Table S2 [175]) compared to the reference (70.7 ppm as found in 1, SI Appendix, Table S1 [175]).

As expected, the newly added residue in oligosaccharide 2 was a β-L-Rha (E) linked at position 4 of D-Xyl (D) as proven by the diagnostic correlations found in both heteronuclear multiple bond correlation (HMBC) and T-ROESY spectra (SI Appendix, Figs. S4 and S5B [175]). The β-configuration of this anomeric center was inferred by the

diagnostic nuclear Overhauser effect (NOE) contacts between the anomeric proton H-1 of E with its H-3 and H-5 protons and confirmed by the $^{1}J_{C,H}$ value (160.3 Hz).

To demonstrate the importance of the cation, the same reaction was repeated without Mn^{2+} along with EDTA; the HPLC profile (SI Appendix and Fig. 3B [175]) revealed that the enzyme required Mn^{2+} to function. Together, these results establish that A064R-D1 acts as a Mn^{2+}-dependent β-L-rhamnosyltransferase, able to form the β-L-Rha-(1→4)-β-D-Xyl linkage. Moreover, the complete conversion of the substrate attests to a lack of a relevant hydrolytic activity versus the donor under the conditions tested, similar to what is observed for the other enzymes (A064R-D2 and A064R-D1D2).

These results were confirmed by the bioluminescent UDP-Glo™ assay in which the specificity of the enzyme for Mn^{2+} and Mg^{2+} was tested, disclosing the enzyme's affinity for Mn^{2+} (SI Appendix, Fig. S6A [175]), in agreement with previous studies [217]. The nature of the acceptor and donor substrates was also investigated, leading to further information (SI Appendix, Fig. S6B [175]): 1) The enzyme did not recognize UDP-α-D-Glc, in agreement with our previous prediction and contrary to previous studies [217]; and 2) the enzyme recognized D-Xyl alone as an acceptor, albeit with a lower affinity with respect to the oligosaccharide acceptor 1.

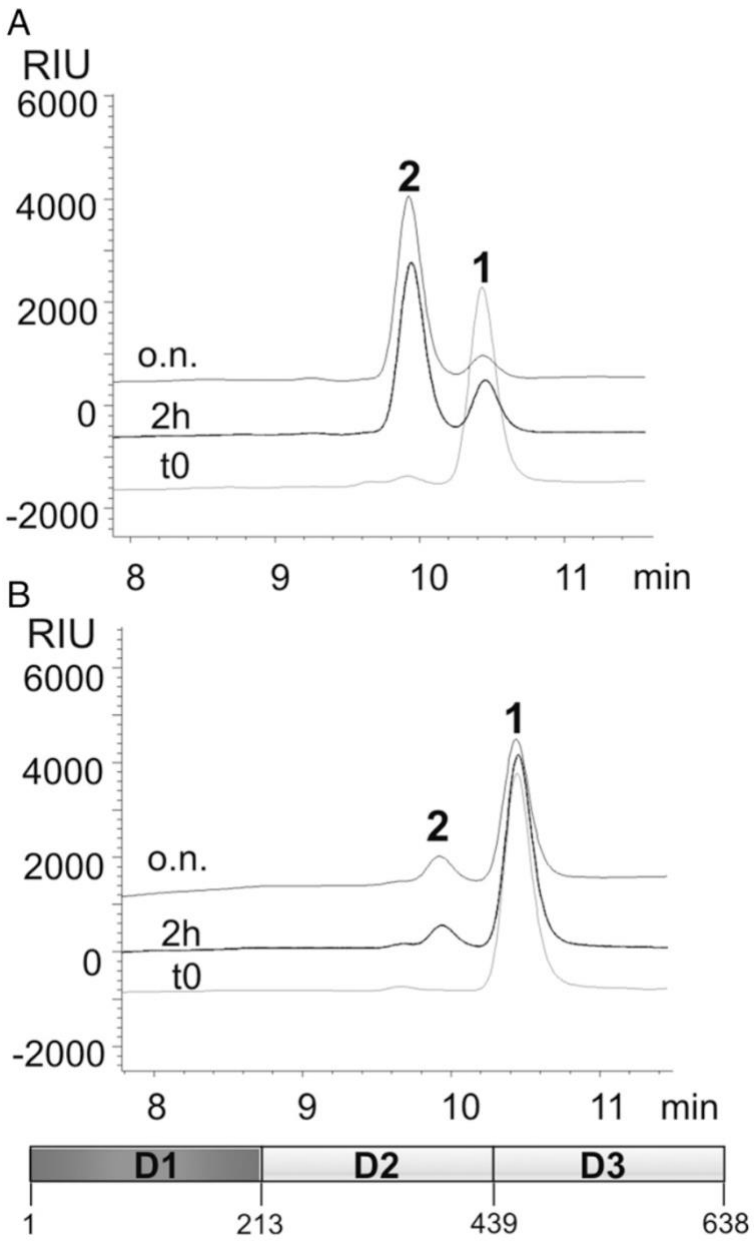

Figure 3.2. A064R-D1 demonstrates β-L-Rhamnosyltransferase activity. HPLC chromatographic profiles obtained from the reaction of the A064R-D1 enzyme (schematic at the bottom indicating its length in aa) with tetrasaccharide 1 (acceptor) and UDP-β-L-Rha (donor) in the presence of (a) Mn^{2+} and (b) with EDTA and without cations. Peak labels refer to oligosaccharide structures in Figure 3.1.

A064R-D2: α-L-Rhamnosyltransferase activity. We previously predicted that A064R-D2 adds the second L-Rha to the first L-Rha [97, 111]. To test this prediction, three versions of the second domain were designed by changing the boundaries, cloned and tested (SI Appendix, Fig. S1A [175]). The first construct (215 aa; named D2) was based on bioinformatic studies. The other two constructs included 33 additional amino acids at the C terminus and differed in length of the N-terminal region: D2L (aa 191–438) and $D2L_2$ (aa 213–438).

C-terminal elongation was investigated because the antigenic variant CME6 [111] has a functional D2 domain longer than the bioinformatic prediction. The first 17 amino acids of the N-terminal region were omitted in the $D2L_2$ construct because they were predicted to form a flexible loop and, therefore, unlikely required for activity. The three proteins were screened for activity using pentasaccharide 2 (Figure 3.1) as an acceptor and UDP-β-L-Rha as a donor with or without cations (Mg^{2+} or Mn^{2+}). Interestingly, the construct initially developed from the bioinformatic studies (D2) was not active (SI Appendix, Fig. S7A [175]), while the other two D2L and $D2L_2$ produced the hexasaccharide 3 (SI Appendix, Fig. S7 B and C [175]).

The same reaction occurred with the D2L enzyme after five days of storage at 4 °C indicating that the enzyme was still active (SI Appendix, Fig. S7D [175]). ^{1}H-NMR spectroscopic investigation showed that oligosaccharide 3 had an additional anomeric signal (5.03 ppm, labeled F) compared to 2 (Figure 3.1). Comparison of the HSQC spectra of 2 and 3 (SI Appendix, Fig. S8A [175]) identified L-Fuc (A), D-Gal (B), α-D-Rha (C), and D-Xyl (D), while the chemical shifts of β-L-Rha (E′) were slightly different due to its glycosylation at C-2, (78.1 ppm instead of 71.8 ppm, SI Appendix, Table S3

[175]). The new unit F was a terminal α-L-Rha; its anomeric signal showed only one correlation in the TOCSY pattern (SI Appendix, Fig. S8B [175]) ascribed to the H-1/H-2 cross peak, whereas the connections up to the methyl group H-6 (1.30 ppm) were visible from H-2, while the C-5 chemical shift value at 69.7 ppm supported the α-configuration of this residue [218].

These results establish that A064R-D2 is an inverting GT, that forms the α-L-Rha-(1→2)-L-Rha linkage. Moreover, amino acids 191–212 at the N terminus do not contribute to the functionality of the protein, but the 33 extra amino acids at the C terminus in D2L$_2$ and D2L are crucial for activity. It is possible that these residues are involved in some dynamic interaction that enables the proper protein folding or confers stability. The reaction was cation independent as confirmed by adding EDTA to the mixture (SI Appendix, Fig. S7E [175]).

Two additional experiments were performed to learn more about the substrate specificity of A064R-D2. First, the octyl xyloside 6 (Figure 3.1) was used as the acceptor with the D2L construct, and no reaction occurred (SI Appendix, Fig. S7F [175]). Next, the reaction was performed using L-Rha monosaccharide as the acceptor. In this case, the reaction could not be monitored via HPLC, but NMR analysis after overnight incubation revealed the formation of a α-L-Rha-(1→2)-L-Rha disaccharide. The ^{1}H-NMR spectrum of the product differed from those of the reagents Rha and UDP-Rha (SI Appendix, Fig. S9A [175]), and the HSQC spectrum (SI Appendix, Fig. S9B [175]) displayed two anomeric signals at ^{1}H/13C 5.22/93.7 ppm and 4.96/103.3 ppm related to the two-linked α-L-Rha at the reducing end (R) and to the terminal α-Rha (T), respectively (SI Appendix, Table S4 [175]). The NMR analysis (SI Appendix, Fig. S9 B and C [175])

confirmed the identity of the disaccharide. In particular, the downfield shift of the C-2 carbon signal of residue R (to 80.3 ppm), confirmed its glycosylation and showed that the enzyme was able to link α-L-Rha at O-2 to another L-Rha unit. A detailed inspection of the spectra detected trace amounts of disaccharide α-L-Rha-(1→2)-β-L-Rha, as well.

Thus, D2L recognizes free Rha as an acceptor. Based on the structure of oligosaccharide 2, which has a β-Rha, we hypothesize that the β-anomer of Rha is the substrate of the enzyme not the α-anomer. Once formed, the reducing end of the disaccharide is converted by mutarotation into the predominant α-form, while some of the free α-Rha replaces the β-form that was consumed. This β-form then reacts with the enzyme and additional UDP-β-L-Rha to give the disaccharide, and this process continues up to the total consumption of the monosaccharide. This hypothesis needs to be validated by future work. Because A064R-D2 resembles several proteins with unknown functions in the GenBank, we predict that these homologs are likely to be undiscovered GTases. Thus, A064R-D2 is a new cation- independent inverting GT that can interact with different Rha-containing molecules.

A064R-D1D2 activity. To confirm our previous findings and expand our understanding about A064R-D1 and A064R-D2 activities, two additional experiments were performed. We made a recombinant protein, A064R-D1D2 (aa 1–438), similar to that encoded by the antigenic variant CME6 virus [111], comprising D1 and D2L active domains. The first reaction used tetrasaccharide 1 as an acceptor and two equivalents of UDP-L-Rha and Mn^{2+}; formation of hexasaccharide 3 was almost complete after 2 h (SI Appendix, Fig. S10A [175]).

The other acceptor used was monosaccharide 6 (Figure 3.1) which disclosed that A064R-D1D2 worked also with a different D-Xyl source. The chromatographic profile after 2 h showed that oligosaccharide 6 was replaced with trisaccharide 7 (SI Appendix, Fig. S10B [175]), whose identity was confirmed by NMR analysis (Figure 3.1 and SI Appendix, Fig. S11, chemical shifts in SI Appendix, Table S5 [175]), namely, that it was 6 with the α-L-Rha-(1→2)-β-L-Rha disaccharide attached to O-4 of the Xyl unit as expected.

A064R-D1 and A064R-D2 bioinformatic analysis. Phylogenetic analysis of annotated GT domains from characterized rhamnosyltransferases revealed that the A064R-D1 was more like members in the GT1 family rather than the GT2 family (Figure 3.3a, detailed list of the organisms used in SI Appendix, Table S6 [175]). However, a previous crystal structure of A064R-D1 indicated a structural fold similar to catalytic domains of retaining GTases in the GT-A family (characterized by a N-terminal Rossmann-like fold) [217], although the amino acid sequence similarity was very low (less than 14% for equivalent Cα atoms), further validating its separation from other characterized rhamnosyltransferase proteins. An A064R-D1 pBLAST query search shows the highest sequence similarity to hypothetical proteins from bacteria and a galactosyl transferase domain from *Burkholderia*.

Interestingly, A064R-D1 shares an ancestral node of origin with eukaryotic homologs (SI Appendix, Fig. S12A [175]). Interestingly, A064R-D2 is the most unusual of the three A064R domains because it least resembles a recognizable GT. Its lack of sequence homology with annotated rhamnosyltransferases (<30% with only 15% coverage) makes it an outlier in a phylogenetic analysis that groups more closely with

other GT2 family members rather than GT1s in the phylogenetic tree (Figure 3.3a). We used BLASTp to further evaluate A064R-D2 (aa 213–439, SI Appendix, Fig. S1A [175]). This domain shares limited homology with several hypothetical and uncharacterized proteins from bacteria, including one GT-family protein from *Lacipirellula parvula* (31.36% identities with 95% coverage) and two homologs in Pithovirus: a hypothetical protein and a GT10 fucosyltransferase (>45% and >35% identities covering the C-terminal domain, respectively). The majority of A064R-D2 homologs are of bacterial origin (SI Appendix, Fig. S12B [175]).

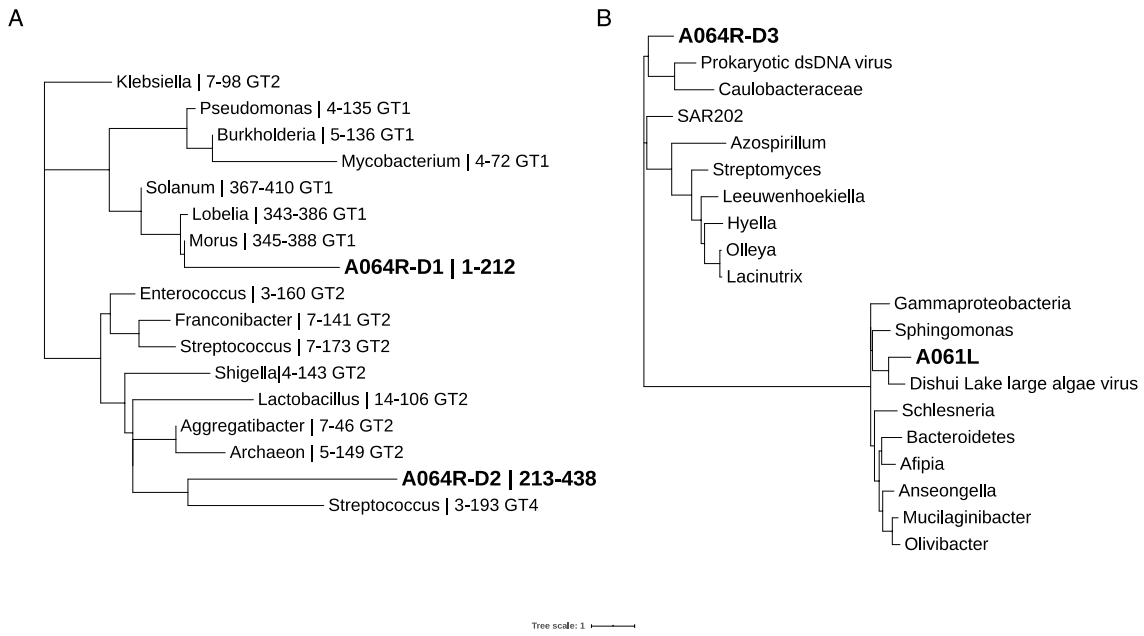

Figure 3.3. Maximum-likelihood phylogenetic trees of some PBCV-1 encoded proteins involved in glycan synthesis. Phylogenetic trees are shown for GT-associated domains from (a) characterized rhamnosyltransferases and (b) selected MTases. (a) Tree construction is based on GT-associated domains of virus-encoded rhamnosyltransferases A064R-D1, A064R-D2, and a variety of annotated rhamnosyltransferases from CAZy. (b) Tree construction is based on MT homologs of virus-encoded proteins A064R-D3 and A061L. The branch lengths indicate evolutionary distance and are labeled with (a) organism, residue range of GT domain, and GT family (a); and organism (b). PBCV-1 encoded proteins are in bold. The list of the organisms used along with their UniProtKB gene accession nos. is reported in SI Appendix, Table S6 [175].

A064R-D3: SAM-Dependent methyltransferase activity. A064R-D3 (aa 439–638) is predicted to have SAM-dependent MT activity. An A064R-D3 pBLAST query search shows the highest sequence similarity to several hypothetical proteins from bacteria and

many class I SAM-dependent MTases of both bacteria and virus origins (SI Appendix, Fig. S13A [175]). Therefore, we previously hypothesized that it might add the two methyl groups to the terminal L-Rha of the glycan [111]. To address this hypothesis, we used the complete A064R protein because the third domain by itself did not form a soluble recombinant protein.

Hence, A064R was incubated with 1 as the acceptor and with two equivalents of UDP-β-L-Rha with Mn^{2+} and Mg^{2+} as the cation requirement of this domain was unknown. Due to the labile nature of SAM, it was not possible to add two equivalents with certainty, thus, we repeatedly added SAM until no increase in new products occurred as judged by HPLC. The reaction produced two products (SI Appendix, Fig. S14A [175]). The less abundant one was hexasaccharide 3 as established by its retention time and by NMR analysis of the mixture. The NMR analysis of the most abundant product 4 (Figure 3.1 and SI Appendix, Fig. S14 B and C and Table S7 [175]) revealed that methylation only occurred at O-2 of the terminal α-L-Rha residue not at both O-2 and O-3 as hypothesized.

^{1}H-NMR and HSQC spectra of oligosaccharide 4 (Figure 3.1 and SI Appendix, Fig. S14 B and C [175]) revealed the same residues found in oligosaccharide 3 with H-1 of E (here, labeled E″) slightly shifted and a new anomeric signal at 5.17 ppm, labeled as F′, which instead replaced F. The anomeric proton of F′ had the TOCSY pattern typical of a Rha unit, α-configured based on its C-5 value (69.5 ppm) and methylated at O-2 based on the diagnostic value of C-2 (81.2 ppm). This finding was confirmed from the NOE contact between H-2 of F′ with the methyl group at 3.45 ppm in the T-ROESY spectrum, which also reported that F′ was linked to O-2 of E'' (SI Appendix, Fig. S14C

[175]). This reaction established that the third domain of A064R only added one methyl group to the C-2 hydroxyl group of the terminal α-L-Rha, leaving open the question about what enzyme methylated the C-3 hydroxyl group.

A061L: Methyltransferase activity. To determine the enzyme responsible for adding a methyl group to O-3 of the terminal α-L-Rha residue, we screened the PBCV-1 genome for another gene encoding an O-MT. This led to the discovery of the gene *a061l* (SI Appendix, Fig. S1A [175]). Phylogenetically distinct from A064R-D3 (Figure 3.3b), the closest homolog of A061L is a macrocin-O-MT from Dishui Lake large algae virus 1. An A061L pBLAST query search shows the highest sequence similarity to several well-characterized class I SAM-dependent MTases from bacteria and, although more distantly related, from cyanobacteria and algae (SI Appendix, Fig. S13B [175]). Moreover, it had a structural homology of 90% of the sequence modeled with 100% confidence with the O-MT NovP from *Streptomyces spheroids*.

Recombinant A061L was tested for activity using hexasaccharides 3 or 4 as acceptors, SAM as a donor of the methyl group, and with both Mg^{2+} and Mn^{2+}. The two acceptors differed only in the presence of the methyl group on O-2 of the terminal L-Rha (Figure 3.1). A061L only methylated hexasaccharide 4, the substrate with the terminal L-Rha already possessing a methyl group at O-2 (SI Appendix, Fig. S15 [175]). No reaction occurred with the unmethylated hexasaccharide 3. The structure of product 5 was inferred via NMR spectroscopic analysis (SI Appendix, Fig. S16 and Table S8 [175]) which confirmed the presence of two methyl groups (at O-2 and O-3) on the terminal L-Rha unit. This last experiment demonstrated that the virus-encoded A061L added a methyl group to the O-3 of the terminal L-Rha only when it was already methylated at O-2.

However, it remains unclear if it has a cation requirement. Additional experiments will be performed to better define all of the characteristics of this enzyme.

Modeling the two methyltransferases. The MT activity of A064R-D3 was supported by bioinformatic information; the protein has 31.5% identity with 91% coverage ($3e^{-19}$ confidence value) with a protein coded by a prokaryotic dsDNA virus sp. (UniProtKB gene accession no. A0A516L7Y9). This viral gene encodes a putative 8-demethyl-8-α-L-rhamnosyl tetraceno-mycin-C2′-O-MT (Figure 3.3b). Moreover, a protein structure prediction by Phyre2 analysis chose the SAM/metal-dependent O-MT MycE from *Micromonospora griseorubida* to predict the 3D structure of A064R-D3 based on 178 residues (89% of the sequence modeled with 100% confidence). The choice by Phyre2 to use this particular MT is consistent with our results.

Previously, crystal structures were determined for MycE bound to the product S-adenosyl-L-homocysteine (SAH) and Mg^{2+}, the first structure of a natural product sugar MT in complex with its natural substrate [219]. The structure of A064R-D3 (aa 446–638) is predicted to accurately trace the catalytic MT domain of MycE (aa 161–399) from residues 161–340 situated in the C terminus comprising an α/β-sandwich with a seven-stranded β-sheet core sandwiched by three α-helices on the front face and two on the back. In agreement with other class I MTases, A064R-D3 possesses a series of conserved motifs shared among these proteins, named motifs I–VI [220].

Notably, A064R-D3 residues 43–51 (Ile43, Ile44, Glu45, Ile46, Gly47, Ile48, Gly49, Asp50, and Phe51) resemble motif I, a nine-residue amino acid block with the consensus sequence (V/I/L)(L/ V)(D/E)(V/I)G(G/C)G(T/P)G. This nine-residue structure contains the glycine-rich ‘‘GxGxG’’ signature sequon, a SAM-binding motif found in

almost all SAM-dependent MTases. Prediction of the A064R-D3 SAM-binding sites was corroborated by the enzyme-ligand structure model (Figure 3.4a and 3.4b). Residues positioned within 5 Å of the SAM ligand are organized into three major regions: 46–51, 74–81, and 121–124. The predicted SAM- binding site for A064R-D3 is located within the loops at the C-terminal end of the central β-sheet as usually occurs for methyl-transferases. Prediction models detect coordination of a metal ion by homologous residues Asp121, Glu149, and Asp150 in accord with the metal dependence of activity observed in MycE (Asp275, Glu303, and Asp304). These residues are strictly conserved among the MycE homologs, implying that all homologs are metal dependent. The SAM amino acid moiety contacts Asp18 through the carboxyl group and with Asp121 and Asp150 through the amino group. As for the metal ligands, the residues that contact SAM are highly conserved. Protein sequence alignment of A064R-D3 and MycE displays homology in critical functional residues (SI Appendix, Fig. S17 [175]). In total, A064R-D3 has 12 of the 23 SAM-binding residues with MycE. Furthermore, active site residues Tyr208 and His278 determined in MycE have homology with A064R-D3 residues Phe51 and His124. The single-residue substitution Y208F MycE resulted in reduced but not abolished enzyme activity [219].

Phyre2 analysis of the SAM/metal dependent sugar O-MT NovP from *S. spheroides* allowed the prediction of the 3D structure of A061L based on 188 residues (modeled with 100% confidence). The A061L sequence contains recognizable motifs I and II that are typical of SAM-dependent MTases (SI Appendix, Fig. S18 [175]). Motif I lies between β1 and α4 and forms the expected interactions with the amino acyl portion of SAM via Glu21 and Gly23, equivalent to residues Glu92 and Gly94 in NovP. Motif II,

defined as an acidic loop, which lies at the C-terminal end of β2, forms the expected interactions with the ribose hydroxyls groups via Asp50, equivalent to residue Asp122 in NovP. The active center of NovP contains a strictly conserved metal-binding site composed of three Asp residues: Asp196, Asp223, and Asp224 [221]. A further conserved Asp198 likely acts as the general base that initiates the methyl transfer reaction. Using in silico docking, we generated models of the A061L-SAM complex that are consistent with this mechanism (Figure 3.4c and 3.4d). A061L possesses a homologous three Asp sequence (Asp143, Asp170, and Asp171) and a general base Asp145, the putative general base of the reaction, is well con- served in A061L structural homologs. We propose this residue initiates the methyl transfer reaction by deprotonating the C-3 hydroxyl group of the L-Rha unit.

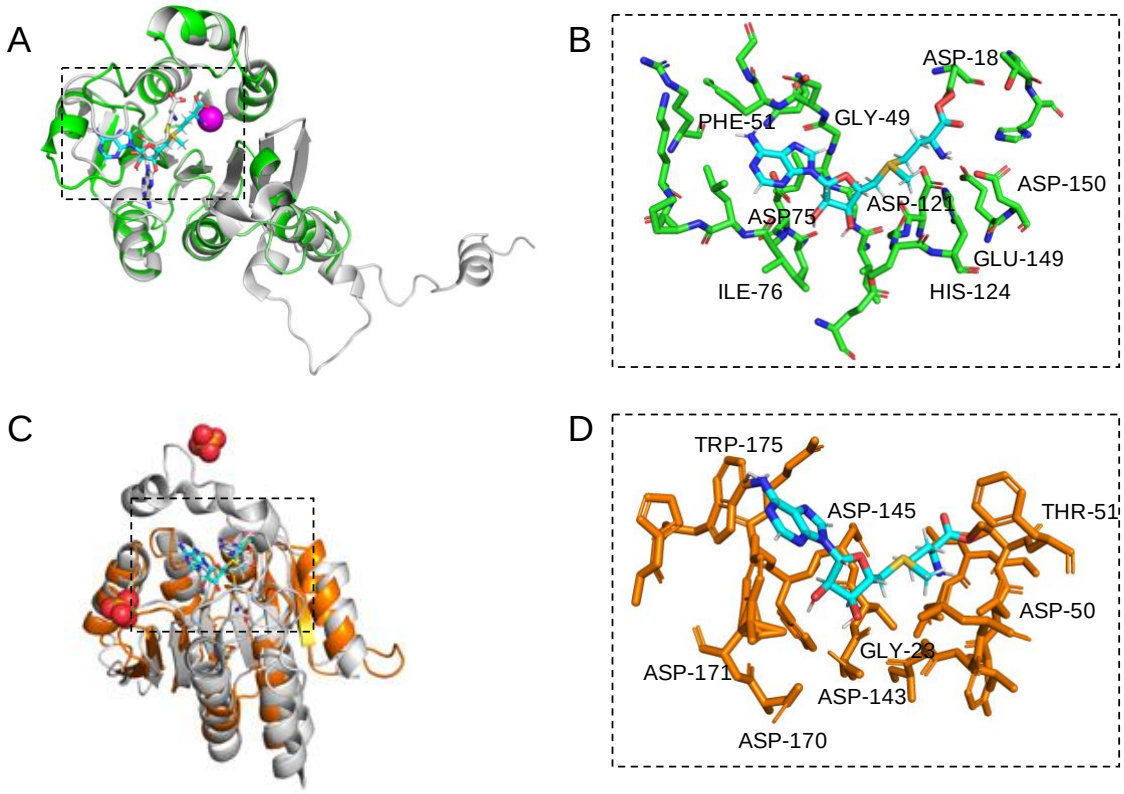

Figure 3.4. Stereoviews showing superpositions of methyltransferases A064R- D3 and A061L with structural homologs. (a) Superposition of predicted A064R-D3 (green) complexed with SAM (cyan) with crystalized SAM/metal-dependent sugar O-MT MycE chain E (gray) from *M. griseorubida* in complex with Mg ion (magenta) and SAH (gray). (b) Magnified view of the predicted A064R-D3 SAM-binding site. (c) Superposition of predicted A061L (orange) complexed with SAM (cyan) with crystalized SAM/metal-dependent sugar O-MT NovP chain A (gray) from *S. spheroids* in complex with sulfate ion (red and yellow) and SAH (gray). (d) Magnified view of the predicted A061L SAM-binding site. Residues represented by stick models are predicted to be positioned within 5 Å of the SAM ligand and have the equivalent residue position of ligand-binding amino acids in characterized sugar O-MT homologs. Labeled residues are

predicted to bind SAM or metal ion via hydrogen bonding or participate in polar contact. Bound SAM and SAH are represented by stick models.

3.2.C. Conclusions

To date, the evidence indicates that chloroviruses decorate their MCP with atypical N-glycans [51, 107, 206] and that the viral genome encodes most, if not all, of the enzymes responsible for synthesizing these glycans. These results support this hypothesis by defining the activity of protein A064R, one of the GTases encoded by chlorovirus PBCV-1, along with the function of the A061L protein encoded by another gene a061l.

A064R (638 aa) has three functional domains of ~200 aa each. The first two are GTases, and the third is a MT that methylates O-2 of the terminal α-L-Rha residue. The finding that the third domain only transfers one methyl group prompted the search for a second virus-encoded MT, which led to the identification of A061L. This enzyme completes the methylation of the terminal α-L-Rha by adding a methyl group to its O-3 position but only after the O-2 position has been methylated.

From a structural viewpoint, *a064r* encodes a rare protein because it has three functional transferase activities: two different rhamnosyltransferases and one MT. To date, few enzymes with multiple GT activities have been identified, and most of them only have two functions. The oldest is KpsC, an enzyme with two GT activities that are involved in the biosynthesis of the *E. coli* K5 polysaccharide [222]. More recently, Clarke *et al*. [223] discovered an O2a polymerase (WbbM) in *Klebsiella pneumoniae* that possesses two domains, a galactopyranosyltransferase resembling known GT8 family

enzymes and a galactofuranosyltransferase defining a previously unrecognized family (GT111). Only a few other multidomain enzymes have been identified with more than one GT active site: WbdA mannosyltransferase involved in the synthesis of the *E. coli* O9, O9a, and O8 lipopolysaccharide O antigens, which is recognized as a bifunctional α-(1→2)-, α-(1→3)-mannosyltransferase in serotype O9a, while its counterpart in serotype O8 is a trifunctional mannosyltransferase (α-[1→2], α-[1→3], and β-[1→2]) [224, 225]. Analyses on the O-antigenic polysaccharide produced by some isolates of *Raoultella terrigena* and *K. pneumoniae* identified another trifunctional GT protein (WbbB), which is able to generate a polysaccharide of [4)-α-Rha-(1→3)-β-GlcNAc-(1→] repeating units capped with a nonreducing terminal residue of β-linked 3-deoxy-D-manno-oct-2-ulosonic acid (β-Kdo) [226]. Other bifunctional microbial GTases are those involved in the biosynthesis of glycosaminoglycans and glycosaminoglycan-like polysaccharides, such as those involved in synthesis of hyaluronan [227], including one from PBCV-1 [99], chondroitin [228], and heparosan [229], along with KpsC from *E. coli* K5, mentioned above. The dGT1 enzyme from *Streptococcus parasanguinis* is a bifunctional protein involved in the biosynthesis of the serine-rich repeat protein adhesin Fimbriae associate protein (Fap1), which is essential for glycosylation of Fap1 [230]. Thus, A064R is part of a restricted pool of GTases with three activities with the qualifier that the third domain is a MT.

Our analysis leads to several significant conclusions: 1) PBCV-1 and probably other chloroviruses encode multifunctional proteins involved in glycan synthesis of their MCP glycoproteins; 2) the PBCV-1 protein A064R possesses a domain (D2) that may represent a new GT family; 3) the A064R protein and probably other chlorovirus encoded

GTases are likely to be soluble due to the lack of N-terminal signal peptides that target them to the ER or the Golgi. This last conclusion suggests they may be easier to express and purify.

Beyond the significance of our findings to chlorovirus biology, we can imagine employing this protein or its individual domains for other purposes, e.g., the A064R-D1D2 protein could be used to produce a different class of rhamnolipid biosurfactants, having the rhamnobiose unit attached to an alkyl xyloside. Rhamnolipids are primarily produced from some strains of *Pseudomonas* [231] and have multiple uses and properties as active agents in skin re-epithelialization in wound healing [232] or used in the cosmetics field for the treatment of wrinkles or to alleviate and/or prevent immunological activities associated with autoimmune diseases.

In conclusion, the results described herein provide direct evidence that the synthesis of the PBCV-1 N-glycan, or, at least, part of it, is accomplished with enzymes encoded by the virus itself. This finding is particularly relevant as it subverts the dogma that all viruses use host enzymes to glycosylate their proteins.

CHAPTER IV

PBCV-1 PROTEIN A111/114R HAS THREE GLYCOSYLTRANSFERASE ACTIVITIES INVOLVED IN GLYCAN SYNTHESIS

4.1. OVERVIEW

Interestingly, all the chloroviruses studied to date, including those with different host specificities, have the capsid protein N-glycosylated with other types of oligosaccharides; however, all chloroviruses share the same pentasaccharide core oligosaccharide [51, 206] composed of an N-linked Glc, a hyperbranched Fuc, a distal and a proximal Xyl, and a Gal (Figure 1.10). Additional monosaccharides decorate this core N-glycan, producing a molecular signature for each chlorovirus [107]. This oligosaccharide core was also found in all mutants (or antigenic variants) of PBCV-1 analyzed to date, except for PIL6 [111]). Mutant PIL6 is a representative of the antigenic class D, characterized by a large genomic deletion spanning genes *a014r* through *a078r* [111, 233]. Its N-glycan is a tetrasaccharide and is significantly truncated compared to that of wild-type PBCV-1. It contains all the units of the chloroviruses N-glycans present in the conserved core region except the distal Xyl attached to Fuc [107]. By analyzing the genomes of all chloroviruses and of the mutants sequenced to date, we noted that the *a111/114r* gene is the only annotated orthologous GT gene found outside of the region of the large deletion mutants that is present in all other chloroviruses. This finding suggests its likely involvement in the synthesis of the initial part of the unusual N-glycan shared by all of these viruses and prompted us to investigate its role in the assemblage of the conserved core oligosaccharide.

4.2. GLYCOSYLTRANSFERASE A111/114R

To predict the A111/114R protein structure and functions, we used a combination of genetic and structural analyses, together with hydrolytic activity assays and in silico evidence by sequence analysis and protein modeling. The combined results suggest that the A111/114R protein is a multi-domain/multi-functional GT likely involved in the attachment of three of the five monosaccharides in the conserved core region of the N-glycan (Figure 1.10). Specifically, this large protein of 860 amino acids is comprised of three putative domains (Figure S1 [234]; Figure 4.1), each with a specific role; the N-terminal domain (1–260 aa) is a galactosyltransferase, the central domain (261–559 aa) is a xylosyltransferase, and the C-terminal domain (560–860 aa) is a fucosyltransferase.

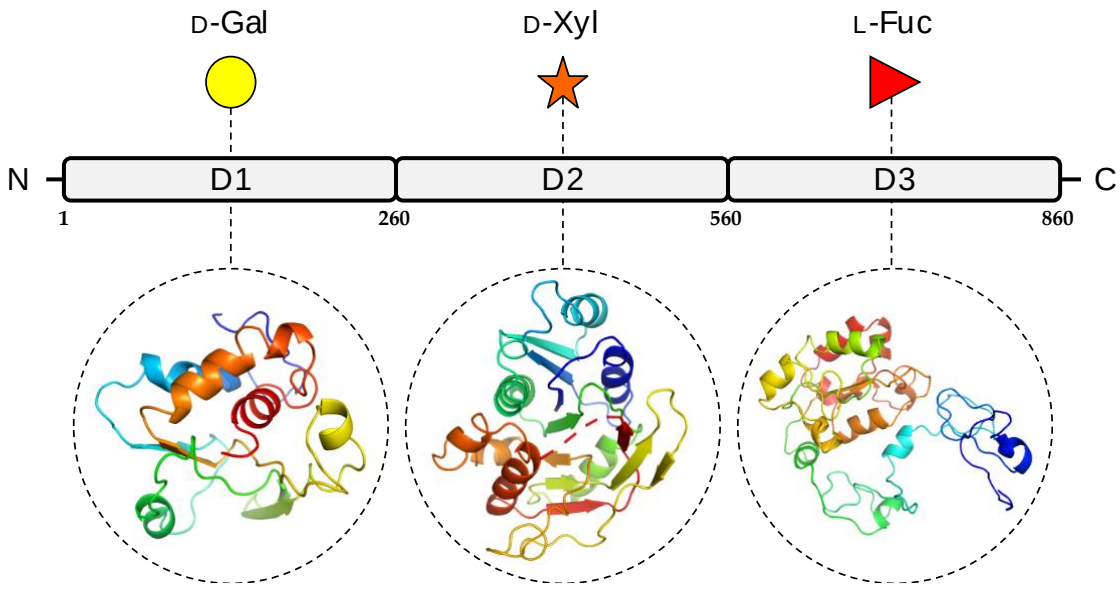

Figure 4.1. Predicted GT domains of PBCV-1 encoded protein A111/114R.

A111/114R domain analysis based on remote homology identified three putative GT domains labeled as D1, D2, and D3, located at the N-terminal (1 to 260 aa), central (261

to 559 aa), and C-terminal (560 to 860 aa) regions, respectively. Below individual domains are the corresponding three-dimensional protein models assigned by Phyre2 [235] based on alignments to known protein structures identified by their PDB entry: 1GA8 Chain A (D1), 2Z86 Chain D (D2), 2NZY Chain A (D3). Protein ribbon models are rendered using rainbow colors from N-terminus (blue) to C-terminus (red). The putative domain, predicted protein model, and sugar substrate are connected by the black dashes.

4.2.A. Materials and Methods

Protein modeling. The prediction of the different domains of A111/114R was performed by HHpred [236] based on remote homology detection. Then, the 3D model of each domain was built by Phyre2 (Protein Homology/analogY Recognition Engine V 2.0) in a normal mode [235]. Each 3D model was based on an alignment generated by HMM–HMM matching. Final drawings and residue analysis were prepared with the molecular graphics system PyMol Version 1.2r3pre, Schrödinger, LLC.

Cloning and expression. PBCV-1 *a111/114r* and domain variants were cloned from PCR-amplified viral DNA using oligonucleotide primers with restriction sites NotI–BamHI. PCR fragments of the expected size were digested and inserted into the restriction sites of the pMAL-c6T expression vector (New England Biolabs, Ipswich, MA, USA). This process produced a maltose-binding protein (MBP) tag at the N-terminus of the target protein. The resulting plasmid was transformed into *E. coli* strain One Shot TOP10 competent cells (Invitrogen) for maintenance. The *E. coli* cells containing positive cloned plasmids were selected with 100 μg/mL carbenicillin. The

cloned structure of each vector was sequence verified. Plasmids were isolated with a QIAprep Spin Miniprep kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions and transformed into NEBExpress competent cells for expression. Viral genes were expressed by growing cells overnight at 37 °C in 10 mL of LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride) containing 100 µg/mL carbenicillin. Then, 5 mL of the over-night culture was sub-cultured into 200 mL LB medium containing 100 µg/mL carbenicillin. The batch culture was grown to an OD_{600} of 0.6 at 37 °C and then induced with 0.1 mM IPTG and incubated at 16 °C overnight. The cells were harvested by centrifugation at 3500× *g*, for 5 min at 4 °C, and resuspended in 35 mL of PBS with 2 mM phenylmethylsulfonyl fluoride (PMSF). After incubation on ice for 30 min, cells were disrupted by sonication for 3 min using a Tekmar sonic disruptor at 30% amplitude, in 5 s pulses. Samples were centrifuged at 10,000 rpm for 15 min to separate soluble and insoluble fractions.

Purification of recombinant enzymes. Amylose resin (New England Biolabs) was loaded onto a 5-mL self-packing column with a 45- to 90-µm-pore-size polyethylene filter (frit) (Life Science Products, Chestertown, MD, USA), and the resin was allowed to settle. The column was equilibrated with 5 column volumes of cold wash buffer (50 mM NaH_2PO_4, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.2). The soluble bacterial fraction was applied to the column and allowed to drain. The column was washed again with 5 column volumes of cold wash buffer. The recombinant proteins were eluted with the MBP moiety using elution buffer (wash buffer plus 10 mM maltose). The recombinant protein concentrations were determined by a NanoDrop spectrophotometer

(NanoDrop Technologies, Wilmington, DE, USA). Eluted proteins were resolved by SDS-PAGE (7.5% acrylamide) with Coomassie brilliant blue staining.

UDP-Glo™ and GDP-Glo™ GT assays. Detection of free uridine diphosphate (UDP) after hydrolysis of the sugar nucleotide was performed using the UDP-Glo™ GT assay kit (Promega Corporation, Madison, WI, USA), which detects UDP after UDP-sugar hydrolysis or transfer by converting UDP to light (measured in Relative Luminescence Units) in a luciferase-type reaction. Detection of free guanosine diphosphate (GDP) from GDP-sugar hydrolysis was evaluated using the GDP-Glo™ GT assay kit (Promega Corporation), which operates by the same principles as above. A standard curve using 0–25 μM of the respective nucleoside diphosphate (NDP) was performed, and the range of measurements was determined to be in the linear range of detection, where the luminescence detected is directly proportional to NDP concentration. Enzymes were diluted with an optimized GT solution (0.1 M MOPS-NaOH pH 7, 10 mM Mn^{2+}) and supplemented with 100 μM of the targeted nucleotide sugar(s). Each sugar–nucleotide hydrolysis reaction was incubated at 16 °C for 16 h. Following the manufacturer's protocol, each reaction was combined in a ratio of 1:1 (25 μL: 25 μL) with the UDP-Glo™ detection reagent in independent wells of a white, flat bottom 96-well assay plate and allowed to incubate at ambient temperature. After 1 h of incubation, luminescence was measured in triplicate, which is directly proportional to UDP or GDP concentration based on the standard curves. Luminescence measurements were performed with a Veritas™ microplate luminometer (Turner Biosystems, Sunnyvale, CA, USA) using a 96-well microplate with standard 128 × 86-mm geometry with an integration time of 1.0

s. Luminescence was measured by a low-noise photomultiplier detector through an empty filter position in the emission filter wheel.

4.2.B. Results and Discussion

***In silico* analysis of A111/114R domain 1.** PBCV-1 encoded protein A111/114R was annotated in NCBI (https://blast.ncbi.nlm.nih.gov/) as a hypothetical protein (NP_048459) of which only one region was predicted to be a GT (262 to 382 aa). PSI-BLAST analysis [237] revealed that A111\114R was conserved among all chloroviruses with 62–94% sequence identity, reinforcing the notion that it could be involved in the assembly of the conserved oligosaccharide core. Therefore, to elucidate the function of A111\114R, we analyzed the full-protein sequence using HHpred tool [236], which predicted three putative GT domains. N-terminal homology with several known GTases (>98% probability) was identified for residues 1 to 256; hence, we assigned residues 1 to 260 aa as domain 1 (A111/114R-D1). The other two domains are discussed in the following sections.

Protein structure prediction by Phyre2 analysis [235] identified six GT crystal structures (Table S1 [234]) to model three-dimensional structures of A111/114R-D1 based on 50–82% protein coverage (>90% confidence and 12–23% sequence identity). Models were ranked according to raw alignment score using the sequence and the secondary structure similarity, inserts, and deletions. Interestingly, the top ranked models were based on two xylosyltransferases (PDB:6BSV, 4WMA), one glucosyltransferase (PDB:1LL2), and three galactosyltransferases (PDB:1GA8, 5GVV, 6U4B), in agreement with our initial hypothesis, as both Xyl and Gal are part of the conserved region of the core glycan (Figure 1.10). The highest ranked model was the xylosyltransferase XXT1

from *Arabidopsis* (PDB: 6BSV) based on 120 residues (19% identity with 96% confidence). However, the homologous region of A111/114R-D1 (9 to 168 aa) that aligned with XXT1 residues 167 to 287 had no homology with the residues in the active site of the XXT1 model, which utilized Lys-382, Asp-317, Asp-318, and Gln-319 to bind the nucleotide sugar for catalytic activity *[238]*. Similarly, important residues for enzyme activity in the xylosyltransferase XXYLT1 (PDB: 4WMA) from *Mus musculus* [239] (His-262, Trp-265, and Gly-325) did not share homologous positions with A111/114R-D1. Based on this evidence, we deduced that A111/114R-D1 has no xylosyltransferase activity, and for this reason we examined the other high top ranked models, all reporting well characterized galactosyltransferases (PDB: 1GA8, 5GVV, 6U4B), except for 1LL2 which is a glucosyltransferase involved in glycogen synthesis [240]. Although glucosyltransferase activity cannot be definitely ruled out, A111/114R-D1 lacks the equivalent 1LL2 residue, Tyr-194, involved in Glc addition.

In order to investigate possible galactosyltransferase activity for A111/114R-D1, we chose, as reference protein, LgtC (1GA8), a galactosyltransferase of *Neisseria meningitidis*, for which the residues involved in the sugar–nucleotide binding and in the catalysis were well characterized. LgtC is a retaining GT that transfers α-D-Gal from UDP-Gal to a terminal lactose. The structure of LgtC was solved in complex with Mn^{2+} and a non-cleavable analog of the donor sugar (UDP-Gal) in which the hydroxyl at the 2′ position of the Gal was substituted by a fluorine for stability purposes [241]. The alignment of the protein sequences of LgtC and A111/114R-D1 (Figure 4.2), along with the structural superimposition of the 3D model of A111/114R-D1 based on LgtC (Table S1 [234]) with 1GA8 (Chain A) in complex with Mn^{2+} and an analog of the donor sugar

(Figure 4.3), clearly revealed that all the residues responsible for binding and catalysis were preserved. In detail, A111/114R-D1 shares four sugar binding residues with LgtC, namely Asp-85, Asn-149, Asp-193, and Gln-194, which correspond to Asp-103, Asn-153, Asp-188, and Gln-189 in LgtC (1GA8). Mutagenesis experiments have established that LgtC Gln-189, contained within the invariant D/EQD motif found in all members of GTases in family 8 (GT8), plays a crucial role in binding and probably in catalysis as well [241]. A catalytic mechanism leading to the retention of the anomeric carbon configuration for A111/114R-D1 is consistent with the presence in the oligosaccharide core of an α-D-Gal (Figure 1.10). It has also been proven that LgtC is a cation-dependent GT [241]. Indeed, LgtC exhibits the typical DXD motif (^{103}DXD^{105}), common in a wide range of GTases, both in prokaryotes and eukaryotes [242]. In addition, a crucial role has been attributed to Asp-103, as LgtC D103E and D103N mutants present a dramatic reduction in their activity compared to the wild-type [241].

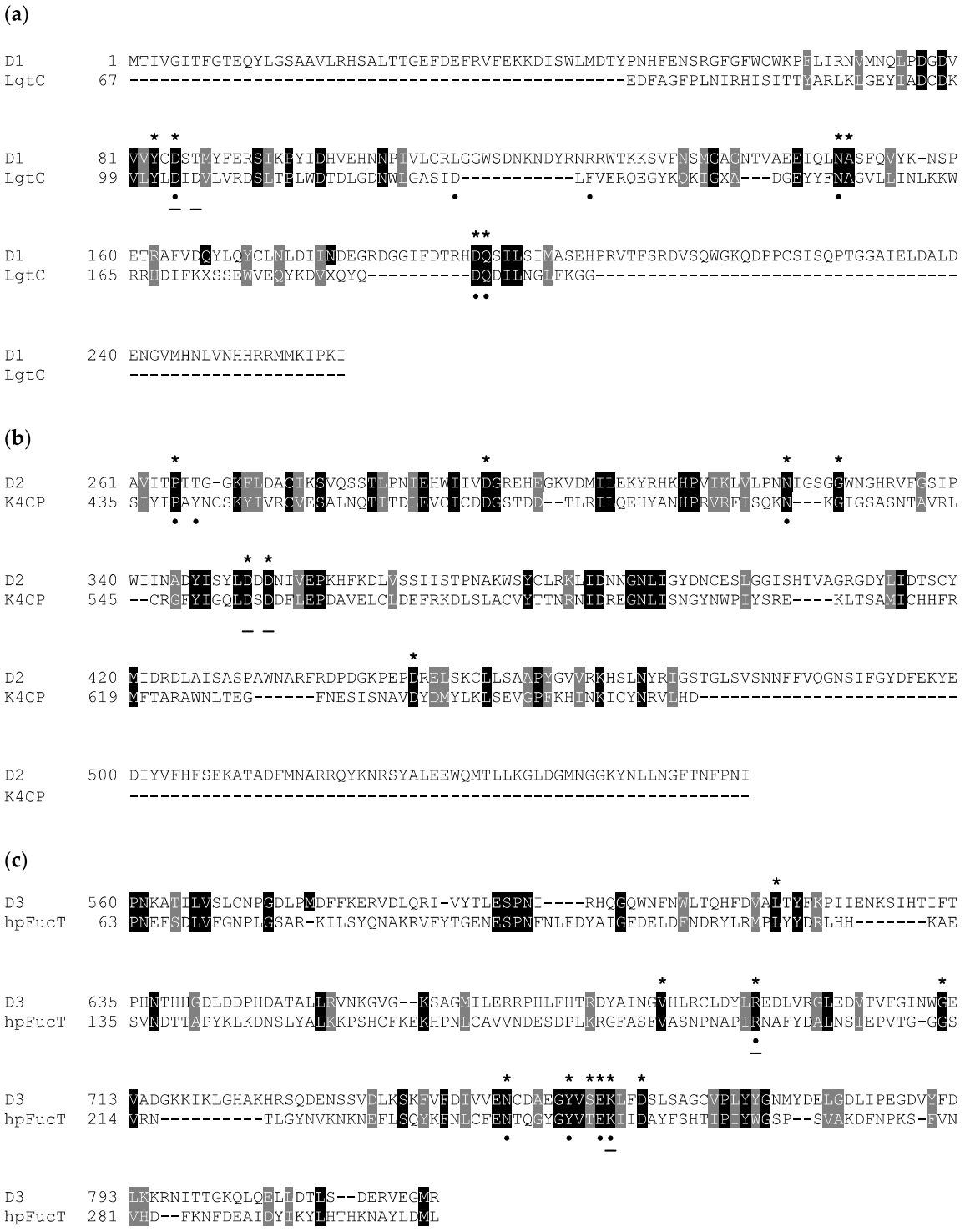

Figure 4.2. Amino acid sequence alignment of A111/114R domains and known GTases. Invariant and similar residues are highlighted in black and gray, respectively.

Sequences from the following organisms were used (PDB code) for three individual domains (D1, D2, and D3), respectively: (a) *N. meningitidis* LgtC galactosyltransferase (1GA8), (b) *E. coli* K4CP dual GalNAc-transferase and GlcA-transferase (2Z86), and (c) *H. pylori* fucosyltransferase (2NZW). Homologous residues positioned less than 4 Å from the nucleotide sugar in the three-dimensional model analysis are marked with an asterisk (*****). Known residues from annotated enzymes that interact with the nucleotide sugar donor and ion are marked with a dot (•) and underline (–), respectively. Multiple alignment was performed by Phyre2 using structural information and homology extension. File output was compiled by BOXSHADE.

However, the ^{103}DXD^{105} motif of LgtC is not preserved in position with A111/114R-D1, except for LgtC Asp-103, which corresponds to A111/114R-D1 Asp-85. Homologous to residue LgtC Asp-103, Asp-85 of A111/114-D1 is positioned in close proximity to a divalent cation and to the nucleotide sugar, as denoted with the characteristic ligand distance < 4 A° (Table 4.1). It is likely that Asp-85 is sufficient in providing one side-chain oxygen ligand in coordination with Mn^{2+}, as evident from the structural superimposition (Figure 4.3). The evidence that A111/114R-D1 possesses all residues involved in binding and catalysis is further supported by the fact that other well-noted galactosyltransferases exhibit similar homologies. The GT GlyE from *Streptococcus pneumoniae* TIGR4 (PDB: 5GVV) binds UDP-Gal [243], and it shares the same conserved sugar binding residues as LgtC and A111/114R-D1 (Asp-103, Asn-142, Asp-177, and Gln-178) (Figure S2 [234]). Mutation of these key residues in GlyE completely abolished the hydrolytic activity [243]. Additionally, the bifunctional domain

polymerase WbbM from *Klebsiella pneumoniae* (PDB: 6U4B) possesses a C-terminal galactopyranosyltransferase [244] with similar homologous residues, Asp-486 and Gln-487, resembling the known GT8 family enzyme signature likely to bind UDP-Gal.

Table 4.1.

Predicted H-bond distances between homologous A111/114R residues and nucleotide–sugar ligand

	A111/114R residues	**H-bond distance of nucleotide-sugar (Å)**
Domain 1	Tyr-83	<4
	Asp-85	2.6, 3.0, 3.6
	Asn-149	<4
	Ala-150	<4
	Asp-193	2.4, 2.5
	Gln-194	<4
Domain 2	Pro-265	2.8, 3.3
	Asp-294	<4
	Asn-323	<4
	Gly-328	<4
	Asp-351	3.2, 3.3
	Asp-353	<4
	Asp-447	<4
Domain 3	Leu-617	<4
	Val-684	<4
	Arg-693	<4
	Gly-711	<4
	Asn-749	1.6, 3.1, 3.3
	Tyr-755	1.9, 3.4
	Ser-757	<4
	Glu-758	2.2
	Lys-759	2.5
	Asp-762	<4

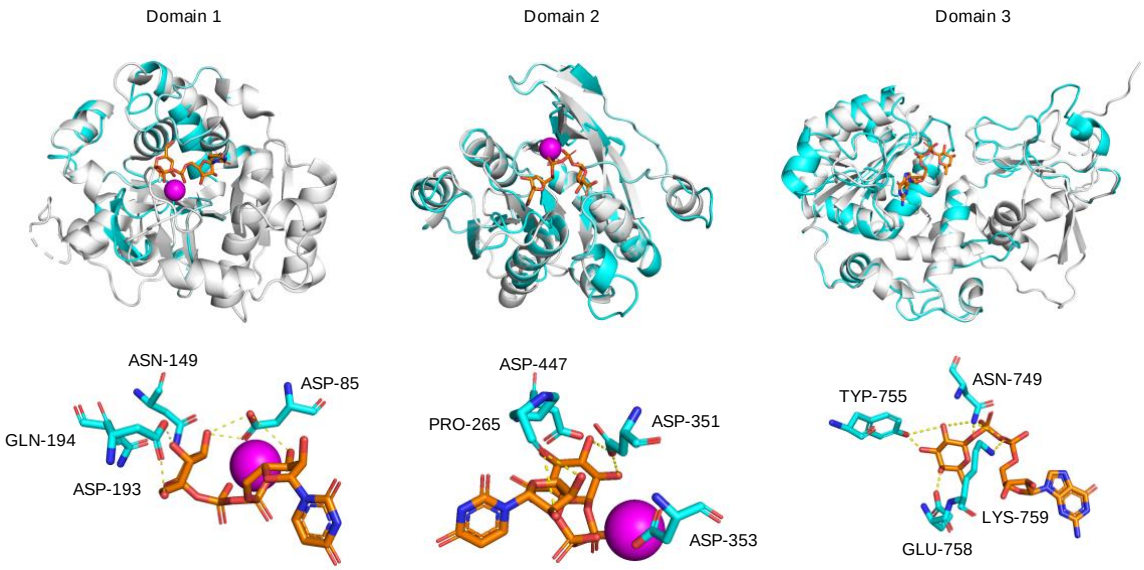

Figure 4.3. Superpositions of A111/114R-D1, -D2, and -D3 with structural homologs. Individual domains of A111/114R (cyan) are shown independently as ribbon diagrams superimposed with known GTases (gray) bound to respective nucleotide sugars drawn as stick models (orange) and Mn^{2+} ion (magenta): *N. meningitidis* LgtC with UDP-fluorogalactose (PDB: 1GA8), *E. coli* K4CP with UDP-glucuronic acid (PDB: 2Z86), and *H. pylori* fucosyltransferase with GDP-Fuc (PDB: 2NZY) are superposed with D1, D2, and D3, respectively (left to right). The corresponding active sites of D1, D2, and D3 are shown magnified below the complexed stereoviews with labeled residues proposed to be involved in sugar and ion coordination. Hydrogen bonds are represented as yellow dotted lines.

***In silico* analysis of A111/114R domain 2.** Three-dimensional renderings of the central domain of A111/114R (261 to 559 aa), referred to as A111/114R-D2, were assembled based on protein alignment and secondary structure similarities. Phyre2 analysis assigned

protein model predictions based on as high as 97% coverage with 100% confidence (11–18% sequence identity) from multiple N-acetylgalactosaminyltransferases (Table S1 [234]). The top ranked model was based on the second domain of the chondroitin polymerase K4CP from *E. coli* (PDB: 2Z86), namely 197 residues (66% of the protein sequence) of A111/114R-D2 were modelled with 100% confidence. K4CP is a bi-functional enzyme organized into two GT-A domains (A1 and A2) that catalyzes elongation of the bacterial chondroitin chain [245]. K4CP A1 (1–417) and A2 (418–682), located respectively to the N- and C-terminal, are engaged in the transfer of GalNAc and GlcA residues alternatively from UDP-GalNAc and UDP-GlcA. The 3D model of A111/114R-D2 (261–474) is based on the second domain (A2) of K4CP (435 to 631 aa), which binds UDP-GlcA, thus excluding GalNAc-transferase activity. The GlcA, absent in the oligosaccharidic core, has the same stereochemistry of Xyl and differs from this monosaccharide by a carboxyl function attached to carbon 5. This finding suggests that the A111/114R-D2 could be a xylosyltransferase, and for that reason we used K4CP-A2 as a reference to assess the conservation of the residues implicated in binding and in catalysis. Sequence alignment of A111/114R-D2 with K4CP-A2 validated the conserved residues involved with GlcA binding and divalent cation coordination (Figure 4.2). In detail, A111/114R-D2 residues involved in the sugar-nucleotide binding are Pro-265, Asp-294, and Asn-323, which correspond to Pro-439, Asp-469, and Asn-496 in K4CP-A2 [245]. Superposition of K4CP Chain D with the predicted A111/114R-D2 structure (Figure 4.3) showcase these residues aligning in close proximity (<4 Å) to the nucleotide sugar (Table 4.1), strengthening their participation in sugar-binding and catalysis. K4CP,

in analogy with other GT-A fold GTases, has a ^{519}DSD^{521} motif coordinating the Mn^{2+}. This DXD motif is preserved in A111/114R-D2 and corresponds to ^{351}DDD^{353}.

Together, the corresponding catalytic sites and DXD motif support orthology in enzyme activity. Of further note, two additional DXD motifs (Figure S1 [234]) are present downstream of A111/114R-D2 (^{422}DRD^{424} and ^{439}DPD^{441}) that potentially could play an active role in substrate recognition or catalysis, but they do not exhibit homology with K4CP (Figure 4.2). Should A111/114R-D2 have functional xylosyltransferase activity, it would be the first xylosyltransferase sandwiched between two domains specific for different nucleotide sugars. The placement of the proximal Xyl in the core glycan structure, positioned closely to Gal and Fuc, is agreeable with the proposed domain organization of A111/114R.

***In silico* analysis of A111/114R domain 3.** The C-terminal domain (560 to 860 aa), referred to as A111/114R-D3, was clearly predicted as a putative α-1,3-fucosyltransferase (Table S1 [234]) modelled from *Helicobacter pylori* (PDB: 2NZW) with 100% confidence and 15% sequence identity (86% coverage). The prediction is that Fuc is linked to the O-3 of a Glc, which, like the monosaccharide contributions of domains D1 and D2, is a component of the overall virus core glycan structure. Fucosyltransferase belongs to the GT B family, in which the protein contains N- and C-terminal domains binding to the acceptor and donor substrates, respectively. Normally, these GTases do not have a recognizable DXD motif responsible for the Mn^{2+}/Mg^{2+} binding; however, A111/114R-D3 has a ^{642}DLD^{644} signature (Figure S1 [234]) that must be evaluated for activity. A sequence alignment of A111/114R-D3 and *H. pylori* fucosyltransferase (hpFucT) revealed conserved residues involved in binding of the donor substrate, GDP-

Fuc (Figure 4.2); Asn-240, Tyr-246, Glu-249, and Lys-250 correspond to Asn-749, Tyr-755, Glu-758, and Lys-759 in A111/114R-D3, respectively (Table 4.1). Classified as an inverting GT, the proposed catalytic mechanism of hpFucT incorporates Glu-95 as a general base in catalysis, while Lys-250 and Arg-195 in hpFucT share a key role in the neutralization of the negative charged phosphate groups from GDP-Fuc to facilitate the glycosidic bond cleavage [246]. Glu-249 acts to stabilize in part the positive charge developed in the transition state as well as to form two hydrogen bonds with both the ribose and the Fuc residues of GDP-Fuc. The Glu-95 residue of hpFucT is located in the N-terminal domain, presumably associated with the acceptor substrate, and it has no equivalent in A111/114R-D3. This finding could be related to alternative acceptors in the two fucosyltransferase enzymes compared here, for example, hpFucT binding GlcNAc or A111/114R-D3 binding Glc.

Comparison of superimposed proteins disclosed that the second half of A111/114R-D3 traces the C-terminal domain of hpFucT Chain A (160 to 320 aa), illustrating similarities in secondary structures. In contrast, structure resemblance decreases towards the N-terminal regions, which could be a result of different acceptor substrates. The open structure of the N-terminal region of the A111/114R-D3 model could be a consequence of a larger acceptor.

To evaluate the sugar-binding and catalytic sites of A111/114R-D3, we superimposed the domain with hpFucT (Chain A) in complex with GDP-Fuc (PDB: 2NZY). This overlap confirmed that all the expected residues involved in the GDP-Fuc binding were preserved and in good orientation (Figure 4.3), in agreement with the sequence alignment data (Figure 4.2). A111/114R-D3 residues Asn-749, Tyr-755, Glu-

758, Lys-759 appeared <4 Å from the nucleotide sugar in agreement with their equivalent residues in hpFucT (Figure 4.3). This observation suggests these A111/114R-D3 residues likely participate in GDP-Fuc binding and are critical for A111/114R-D3 activity. Specifically, Asn-749, Tyr-755, and Glu-758 likely stabilize the positive charge on the Fuc moiety like their equivalent residue in hpFucT. Positioned close in proximity to GDP-Fuc is A111/114R-D3 residue Arg-693 that corresponds to Arg-195 in hpFucT, an essential residue for enzyme activity. In fact, Ala mutants of Arg-195 or Lys-250 resulted in no detectable activity, supporting the idea that the two residues provide positive charges to interact with negatively charged GDP-Fuc [246]. The important role of Arg-693 is supported by its conservation amongst different chloroviruses.

***In vitro* evidence of A111/114R hydrolytic activity.** In order to test A111/114R for GT activity, we used the bioluminescent UDP/GDP-Glo^{TM} assays to detect free UDP/GDP released by GT-mediated hydrolysis of the nucleotide sugars. This allowed us to screen for nucleotide donor specificities without their target substrates (acceptor). A time course experiment with the full-length recombinant A111/114R protein in the presence of the core monosaccharides UDP-Gal, UDP-Xyl, GDP-Fuc, and Glc displayed evidence of UDP-hydrolysis (Figure 4.4). Glc was supplemented to simulate the Glc-Asn acceptor located on the nascent glycoprotein. The release of the UDP increased steadily over time, suggestive of GT activity by A111/114R-mediated hydrolysis.

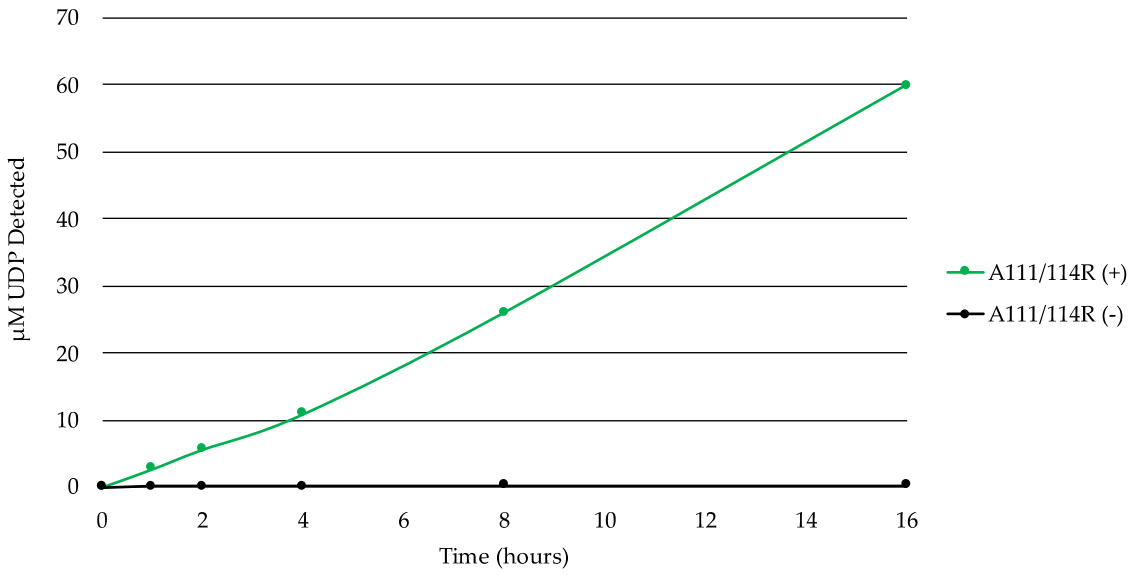

Figure 4.4. A111/114R-catalyzed hydrolysis of UDP-sugars. Time-course experiment with the full-length recombinant A111/114R protein (6 µg) in the optimized buffer consisting of 0.1 M MOPS-NaOH (pH 7.0), 10 mM $MgCl_2$, and 100 µM each of UDP-Gal, UDP-Xyl, GDP-Fuc, and Glc for 16 h at 16 °C. The release of UDP was detected by the UDP-Glo™ assay. Data are representative from three independent replicates, and error bars represent standard deviation.

We recombinantly expressed the full-length A111/114R protein (1 to 860 aa) and three variants with omitted regions (Figures 4.5a and 4.5b). A111/114R (1–397 aa) contains the complete domain 1 and approximately the first half of domain 2, A111/114R (266–860 aa) contains the complete domains 2 and 3, and A111/114R (391–860 aa) contains the second half of domain 2 and the complete domain 3. Constructs containing only individual domains were originally designed; however, their proteins did not exhibit activity. This may reflect inaccurate residue boundaries or an incorrect folding of the single domains in the absence of adjacent ones. Representative data from each assay are

shown in Figure 4.5 (4.5c and 4.5d) and are represented as a ratio of the UDP or GDP measured from reactions containing the indicated GTases relative to the full-length protein-catalyzed hydrolysis, and negative controls in which no enzyme was added.

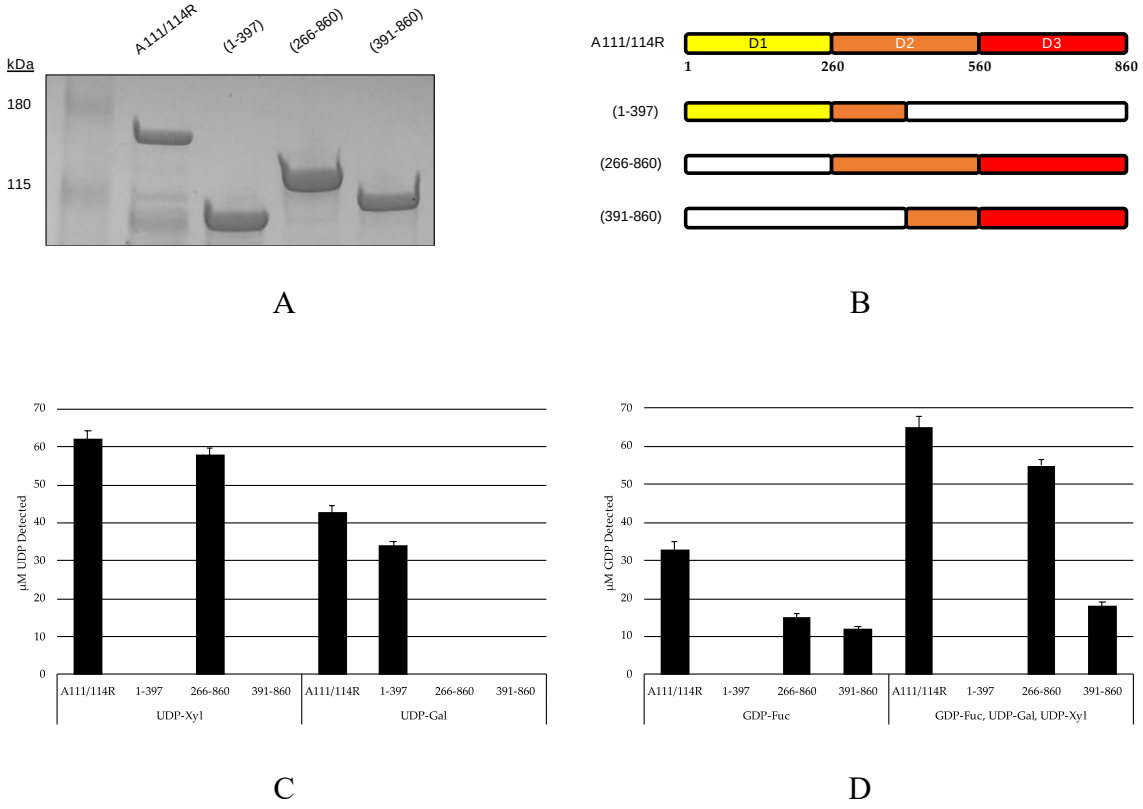

Figure 4.5. Hydrolysis of UDP- and GDP-sugars by A111/114R and truncated constructs. (a) SDS-PAGE analysis of the expressed proteins: full length recombinant MBP-A111/114R (144 kDa), and truncated variants MBP-A111/114R (1–397) (87 kDa), MBP-A111/114R (266–860) (110 kDa), and MBP-A111/114R (391–860) (96 kDa) were eluted from an amylose column and resolved by SDS/PAGE with Coomassie blue staining. BenchMark™ pre-stained protein ladder and recombinant proteins were

separated on a 4–20% tris-glycine gel. (b) Cartoon renderings of the full-length A111/114R protein and truncated versions are color coordinated by three putative domains (D1, D2, D3), each corresponding to a different nucleotide sugar donor. White regions denote omitted sections of A111/114R. Shortened constructs are defined by their residues in the left column. (c and d) Representative data from hydrolysis assays shown as a ratio of the UDP (c) or GDP (d) measured from reactions containing the indicated GTases relative to the negative controls where no enzyme was added. The release of UDP and GDP was detected by the UDP-Glo™ assay and GDP-Glo™ assay, respectively. GDP-hydrolysis from GDP-Fuc was significantly elevated in the presence of UDP-Gal and UDP-Xyl with A111/114R (266–860) and A111/114R (391–860). Data are representative from three replicates, and error bars represent standard deviation.

Analysis of UDP/GDP-hydrolysis by the A111/114R constructs reported in Figure 4.5b were especially revealing in regard to A111/114R domain assignments. Indeed, starting with the A111/114R (1–397) construct, the UDP molecules were detected only when UDP-Gal was used (Figure 4.5c). Given A111/114R-D2 is half omitted in A111/114R (1–397), hydrolysis of UDP-Gal implies A111/114R-D1 is a galactosyltransferase. In agreement, no UDP was produced from UDP-Gal containing reactions involving A111/114R constructs devoid of the first domain, A111/114R (266–860) or A111/114R (391–860). This finding is in agreement with the bioinformatic data.

Hydrolysis of UDP-Xyl by A111/114R (266–860) suggests either A111/114R-D2 or A111/114R-D3 has xylosyltransferase activity. However, UDP-Xyl was not detected in reactions involving A111/114R (1–397) or A111/114R (391–860), both of which are

devoid of a complete D2 (Figure 4.5b). This suggests that A111/114R-D2 (260–560) harbors the xylosyltransferase activity. GDP-Fuc hydrolysis was detected in reactions involving A111/114R (266–860) and A111/114R (391–860) exclusively (Figure 4.5d). These results suggest A111/114R-D3 is a fucosyltransferase. Notably, GDP-Fuc reactions supplemented with UDP-Gal and UDP-Xyl showed elevated levels of liberated GDP, especially in the presence of A111/114R (266–860). This could be a result of improved protein folding allowed by the extension of residues and co-presence of the nucleotide sugars.

Finally, to evaluate the residues of A111/114R involved in hydrolytic activity, we constructed Ala mutants by site-directed mutagenesis (SDM) (GenScript) to target amino acids from each domain predicted to be involved in nucleotide–sugar or metal–ion binding. Three mutants were expressed, each containing two Ala substitutions inside separate domains. A111/114R-D1, -D2, and -D3 SDM constructs contained Ala mutants of Asp-85 and Gln-194, Asp-351 and Asp-353, and Arg-693 and Lys-759, respectively. In the presence of UDP-Xyl, UDP-Gal, GDP-Fuc, and Glc, the SDM of D1, D2, and D3 resulted in a significant reduction in UDP-sugar hydrolysis, lowering the activity by 95%, 90%, and 80%, respectively (Figure S3 [234]). Likewise, in the presence of the same nucleotide–sugars, GDP-Fuc hydrolysis was reduced by 90%, 70%, and 96%, respectively. This dramatic reduction in detectable nucleotide–sugar hydrolysis supports the idea that these residues are critical for enzyme activity, and that A111/114R functions best when all three domains are active.

4.2.C. Conclusions

The giant chloroviruses continue to challenge our understanding of canonical metabolic pathways in host–virus interplay. The identification of the atypical N-glycan structure attached to the PBCV-1 Vp54 has led to the characterization of virus-encoded enzymes involved in glycosylation independent of host-derived ER and Golgi GTases. The N-glycan's core pentasaccharide conserved among the chloroviruses is especially noteworthy and prompted us to investigate candidate virus-encoded GTases involved in the assemblage of part or all the conserved core oligosaccharide. Results in this study establish that PBCV-1 encoded protein A111/114R has three GT domains of approximately 300 aa each. Evidence from a combination of amino acid alignments, three dimensional renderings, and GT assays indicate that the N-terminal (1 to 260 aa), central (261 to 559 aa), and the C-terminal (560 to 860 aa) regions resemble a galactosyltransferase, a xylosyltransferase, and a fucosyltransferase, respectively. The three-dimensional protein models built by Phyre2 are predictions and were used only for the purpose of identifying potential individual domains. As with all methods for protein modeling prediction, caution should be exercised when evaluating structural elements of new enzymes that lack homology to currently deposited structures in the Protein Data Bank archives. In fact, since the percent identity between the various domains and the structures on which they were modeled was low, it was not possible to model the various domains accurately. Eventually individual protein domains must be solved by biochemical and crystallographic methods to fully reveal their catalytic mechanism.

Preliminary evidence suggests that A111/114R-D2 has xylosyltransferase activity and, if biochemically confirmed, presents a new structural class of xylosyltransferases

based on its limited resemblance to known xylosyltransferases. Moreover, A111/114R would be the second three-domain protein encoded by PBCV-1 that is involved in glycan synthesis. Indeed, recent studies showed that PBCV-1 protein A064R (638 aa) has three functional domains; the first two are GTases (β-L-rhamnosyltransferase and α-L-rhamnosyltransferase, respectively) and the third is a MT that methylates O-2 of the terminal α-L-Rha residue [247].

It is known that many of the chlorovirus genes encode enzymes involved in various aspects of carbohydrate metabolism. However, it remains unclear how the virus-encoded proteins are involved in the synthesis and/or assembly of the Vp54 glycan. For example, are the sugars added to Vp54 sequentially or are they synthesized independently of Vp54, possibly on a lipid carrier, and then attached to the protein en bloc? A slight variation of these two possibilities is to synthesize a core glycan(s) independently of the protein and attach it to Vp54. Additional experiments will be required to address this issue.

Importantly, the results described herein provide support that the synthesis of the PBCV-1 core glycan structure, or at least part of it, is accomplished with a multidomain enzyme encoded by the virus itself. This finding is in line with the finding that another GT of PBCV-1, the protein A064R, is able to elongate the viral glycan with two units of Rha and to methylate the ultimate unit at O-2. Taking these finding together, the dogma that all viruses use host enzymes to glycosylate their proteins is further subverted.

CHAPTER V

CONCLUSIONS AND FUTURE PROSPECTS

5.1. OVERVIEW

All living cells in all domains of life are covered in glycans, typically as glycoproteins and glycolipids. Glycans play diverse roles, including critical functions in the areas of cell signaling, molecular recognition, immunity, and inflammation. Here, we investigated the potential involvement of chlorovirus-encoded putative GTases and MTases in glycosylation of the viral MCP. First, we aimed to generate site-directed virus mutants by targeting associated viral genes. However, at present a significant barrier to genetic transformation in chloroviruses is the recalcitrance of its host green alga making reverse genetic manipulation of chlorovirus genomes not possible. To address these limitations, we tested popular transformation methods using cell wall-degrading enzymes, electroporation, SiC whiskers, cell-penetrating peptides, and *Agrobacterium* to generate GT-gene mutations in the chlorovirus CA-4B. We successfully delivered preassembled Cas9 protein-sgRNA ribonucleoproteins (RNPs) to macerozyme-treated NC64A cells that resulted in a frameshift mutation in the CA-4B-encoded gene *034r*, a homolog of PBCV-1 GT gene *a064r*. Unable to duplicate these results, we shifted our focus to characterize PBCV-1-encoded proteins involved in glycan synthesis.

We aimed to characterize chlorovirus-encoded proteins related to the synthesis of the unusual glycans attached to the MCP of PBCV-1. Our goal was to biochemically corroborate the involvement of virus-encoded GTases with the monosaccharides that comprise the atypical viral N-glycans. Specifically, we tested the functional properties of

undetermined virus-encoded proteins that are predicted to be involved in the synthesis of its N-glycan(s) as disclosed by bioinformatic analyses. We showed that A064R has three functional domains: domain 1 is a β-L-rhamnosyltransferase; domain 2 is an α-L-rhamnosyltransferase, and domain 3 is a MT that methylates the C-2 hydroxyl group of the terminal α-L-Rha unit. We also established that methylation of the C-3 hydroxyl group of the terminal α-L-Rha is achieved by another virus-encoded protein, A061L which acts on the substrate, only if the methyl group on O-2 is already installed.

Moreover, genetic and structural analyses indicated the protein coded by PBCV-1 gene *a111/114r*, conserved in all chloroviruses, is a likely a GT with three putative domains: domain 1 is a galactosyltransferase, domain 2 is a xylosyltransferase, and domain 3 is a fucosyltransferase. Hydrolytic assays supported these predictions suggesting that A111/114R is likely responsible for the attachment of three of the five conserved residues of the core region of this complex glycan. These findings provide additional support that the chloroviruses do not use the canonical host ER-Golgi glycosylation pathway to glycosylate their glycoproteins; instead, they perform glycosylation independent of cellular organelles using virus-encoded enzymes.

The unusual chlorovirus N-glycan structure(s) is synthesized by enzymes encoded by the virus and does not involve the host glycosylation machinery. That is, unlike most other viruses, the prototype chlorovirus PBCV-1 encodes most, if not all, of the machinery required to glycosylate its MCP in the cytoplasm rather than in the ER and Golgi that is typical for most glycoprotein-containing viruses that use host enzymes to synthesize their glycans. This finding is particularly relevant as it challenges the dogma that all viruses use host enzymes to glycosylate their proteins. If our predictions about

chloroviruses are correct, investigations of these organisms will provide the field of glycobiology with new discoveries and may ultimately answer the central question about what benefits these viruses have by maintaining their own glycosylation machinery through evolution. These findings could suggest an alternative virus-encoded glycosylation pathway that perhaps pre-dates the evolution of the canonical Golgi-ER glycosylation pathway.

Historically, GTases from eukaryotic organisms have been difficult to study biochemically and structurally because they are membrane associated, notoriously unstable, often contain flexible loops and domains and are commonly associated with other interacting proteins. In contrast, as described here, GTases encoded by a group of viruses, called chloroviruses, lack transmembrane domains because they function in the cytoplasm of their host algae. Consequently, their recombinant proteins are soluble and easy to work with biochemically. Our work has assisted in identifying and characterizing many of these novel chlorovirus-encoded GTases as well as virus encoded MTases (and possibly other modifications) that add methyl groups to specific sites on sugar molecules. Chlorovirus-encoded GTases and MTases contribute towards the expanding repertoire of enzymes for producing glycoconjugates that can be used in glycotechnological applications, e.g., for creating adjuvants for vaccines, creating novel biosurfactants and novel antibiotics. Finally, we suspect that some of the other giant viruses that are being discovered also probably encode unique GTases. If this proves true, giant viruses will become an extremely valuable source for new GTases.

5.2. REFERENCES

1. Guiry, M.D., How Many Species of Algae Are There? *J Phycol*, **2012**. *48*(5): p. 1057-63.

2. Dunigan, D.D., Fitzgerald, L.A., and Van Etten, J.L., Phycodnaviruses: A Peek at Genetic Diversity. *Virus Res*, **2006**. *117*(1): p. 119-32.

3. Van Etten, J.L., Graves, M.V., Muller, D.G., Boland, W., and Delaroque, N., Phycodnaviridae--Large DNA Algal Viruses. *Arch Virol*, **2002**. *147*(8): p. 1479-516.

4. Derelle, E., Yau, S., Moreau, H., and Grimsley, N.H., Prasinovirus Attack of Ostreococcus Is Furtive by Day but Savage by Night. *J Virol*, **2018**. *92*(4).

5. Filee, J., Genomic Comparison of Closely Related Giant Viruses Supports an Accordion-Like Model of Evolution. *Front Microbiol*, **2015**. *6*: p. 593.

6. Filee, J., Giant Viruses and Their Mobile Genetic Elements: The Molecular Symbiosis Hypothesis. *Curr Opin Virol*, **2018**. *33*: p. 81-88.

7. Lefeuvre, P., Martin, D.P., Elena, S.F., Shepherd, D.N., Roumagnac, P., and Varsani, A., Evolution and Ecology of Plant Viruses. *Nat Rev Microbiol*, **2019**. *17*(10): p. 632-644.

8. Allen, M.J., Schroeder, D.C., Holden, M.T., and Wilson, W.H., Evolutionary History of the Coccolithoviridae. *Mol Biol Evol*, **2006**. *23*(1): p. 86-92.

9. Fuhrman, J.A., Marine Viruses and Their Biogeochemical and Ecological Effects. *Nature*, **1999**. *399*(6736): p. 541-8.

10. Suttle, C.A., Marine Viruses--Major Players in the Global Ecosystem. *Nat Rev Microbiol*, **2007**. *5*(10): p. 801-12.

11. Wommack, K.E. and Colwell, R.R., Virioplankton: Viruses in Aquatic Ecosystems. *Microbiol Mol Biol Rev*, **2000**. *64*(1): p. 69-114.

12. Norrby, E., Nobel Prizes and the Emerging Virus Concept. *Arch Virol*, **2008**. *153*(6): p. 1109-23.

13. Claverie, J.M., Viruses Take Center Stage in Cellular Evolution. *Genome Biol*, **2006**. *7*(6): p. 110.

14. Claverie, J.M. and Abergel, C., Mimivirus: The Emerging Paradox of Quasi-Autonomous Viruses. *Trends Genet*, **2010**. *26*(10): p. 431-7.

15. La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.M., and Raoult, D., A Giant Virus in Amoebae. *Science*, **2003**. *299*(5615): p. 2033.

16. Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J.M., The 1.2-Megabase Genome Sequence of Mimivirus. *Science*, **2004**. *306*(5700): p. 1344-50.

17. Van Etten, J.L., Lane, L.C., and Dunigan, D.D., DNA Viruses: The Really Big Ones (Giruses). *Annu Rev Microbiol*, **2010**. *64*: p. 83-99.

18. Philippe, N., Legendre, M., Doutre, G., Coute, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J.M., and Abergel, C., Pandoraviruses: Amoeba Viruses with Genomes up to 2.5 Mb Reaching That of Parasitic Eukaryotes. *Science*, **2013**. *341*(6143): p. 281-6.

19. Koonin, E.V. and Wolf, Y.I., Genomics of Bacteria and Archaea: The Emerging Dynamic View of the Prokaryotic World. *Nucleic Acids Res*, **2008**. *36*(21): p. 6688-719.

20. Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., Lescot, M., Poirot, O., Bertaux, L., Bruley, C., Coute, Y., Rivkina, E., Abergel, C., and Claverie, J.M., Thirty-Thousand-Year-Old Distant Relative of Giant Icosahedral DNA Viruses with a Pandoravirus Morphology. *Proc Natl Acad Sci U S A*, **2014**. *111*(11): p. 4274-9.

21. Colson, P., Levasseur, A., La Scola, B., Sharma, V., Nasir, A., Pontarotti, P., Caetano-Anolles, G., and Raoult, D., Ancestrality and Mosaicism of Giant Viruses Supporting the Definition of the Fourth Truc of Microbes. *Front Microbiol*, **2018**. *9*: p. 2668.

22. Koonin, E.V. and Yutin, N., Multiple Evolutionary Origins of Giant Viruses. *F1000Res*, **2018**. *7*.

23. Koonin, E.V. and Yutin, N., Evolution of the Large Nucleocytoplasmic DNA Viruses of Eukaryotes and Convergent Origins of Viral Gigantism. *Adv Virus Res*, **2019**. *103*: p. 167-202.

24. Yutin, N. and Koonin, E.V., Pandoraviruses Are Highly Derived Phycodnaviruses. *Biol Direct*, **2013**. *8*: p. 25.

25. Iyer, L.M., Aravind, L., and Koonin, E.V., Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. *J Virol*, **2001**. *75*(23): p. 11720-34.

26. Iyer, L.M., Balaji, S., Koonin, E.V., and Aravind, L., Evolutionary Genomics of Nucleo-Cytoplasmic Large DNA Viruses. *Virus Res*, **2006**. *117*(1): p. 156-84.

27. Koonin, E.V. and Yutin, N., Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses. *Intervirology*, **2010**. *53*(5): p. 284-92.

28. Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D.K., Cheng, X.W., Federici, B.A., Van Etten, J.L., Koonin, E.V., La Scola, B., and Raoult, D., "Megavirales", a Proposed New Order for Eukaryotic Nucleocytoplasmic Large DNA Viruses. *Arch Virol*, **2013**. *158*(12): p. 2517-21.

29. Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, R.D., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.F., Dougherty, B.A., Bott, K.F., Hu, P.C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchison, C.A., 3rd, and Venter, J.C., The Minimal Gene Complement of Mycoplasma Genitalium. *Science*, **1995**. *270*(5235): p. 397-403.

30. La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E., and Raoult, D., The Virophage as a Unique Parasite of the Giant Mimivirus. *Nature*, **2008**. *455*(7209): p. 100-104.

31. Katzourakis, A. and Aswad, A., The Origins of Giant Viruses, Virophages and Their Relatives in Host Genomes. *BMC Biol*, **2014**. *12*: p. 51.

32. Van Etten, J.L., Lane, L.C., and Meints, R.H., Viruses and Viruslike Particles of Eukaryotic Algae. *Microbiol Rev*, **1991**. *55*(4): p. 586-620.

33. Manton, I., Fine-Structural Observations on Six Species of Chrysochromulina from Wild Danish Marine Nanoplankton, Including a Description of C. Campanulifera Sp. Nov. And a Preliminary Summary of the Nanoplankton as a Whole. **1974**.

34. Mayer, J. and Taylor, F., A Virus Which Lyses the Marine Nanoflagellate Micromonas Pusilla. *Nature*, **1979**. *281*(5729): p. 299-301.

35. Cottrell, M. and Suttle, C., Wide-Spread and Clonal Variation in Viruses Which Causes Lysis of a Cosmopolitan, Eukaryote Marine Phytoplankter, Micromonas Pudlla. *Mar Ecol Prog Ser*, **1991**. *78*: p. l-9.

36. Bergh, O., Borsheim, K.Y., Bratbak, G., and Heldal, M., High Abundance of Viruses Found in Aquatic Environments. *Nature*, **1989**. *340*(6233): p. 467-8.

37. Van Etten, J.L., Unusual Life Style of Giant Chlorella Viruses. *Annu Rev Genet*, **2003**. *37*: p. 153-95.

38. KAWAKAMI, H. and KAWAKAMI, N., Behavior of a Virus in a Symbiotic System, Paramecium Bursaria—Zoochlorella. *The Journal of Protozoology*, **1978**. *25*(2): p. 217-225.

39. Meints, R.H., Van Etten, J.L., Kuczmarski, D., Lee, K., and Ang, B., Viral Infection of the Symbiotic Chlorella-Like Alga Present in Hydra Viridis. *Virology*, **1981**. *113*(2): p. 698-703.

40. Van Etten, J.L., Meints, R.H., Burbank, D.E., Kuczmarski, D., Cuppels, D.A., and Lane, L.C., Isolation and Characterization of a Virus from the Intracellular Green Alga Symbiotic with Hydra Viridis. *Virology*, **1981**. *113*(2): p. 704-11.

41. Van Etten, J.L., Meints, R.H., Kuczmarski, D., Burbank, D.E., and Lee, K., Viruses of Symbiotic Chlorella-Like Algae Isolated from Paramecium Bursaria and Hydra Viridis. *Proc Natl Acad Sci U S A*, **1982**. *79*(12): p. 3867-71.

42. J.L., V.E., Burbank, D.E., Kuczmarski, D., and Meints, R.H., Virus Infection of Culturable Chlorella-Like Algae and Dlevelopment of a Plaque Assay. *Science*, **1983**. *219*(4587): p. 994-6.

43. Short, S.M., The Ecology of Viruses That Infect Eukaryotic Algae. *Environ Microbiol*, **2012**. *14*(9): p. 2253-71.

44. Quispe, C.F., Sonderman, O., Seng, A., Rasmussen, B., Weber, G., Mueller, C., Dunigan, D.D., and Van Etten, J.L., Three-Year Survey of Abundance, Prevalence and Genetic Diversity of Chlorovirus Populations in a Small Urban Lake. *Arch Virol*, **2016**. *161*(7): p. 1839-47.

45. Flaviani, F., Schroeder, D.C., Lebret, K., Balestreri, C., Highfield, A.C., Schroeder, J.L., Thorpe, S.E., Moore, K., Pasckiewicz, K., Pfaff, M.C., and Rybicki, E.P., Distinct Oceanic Microbiomes from Viruses to Protists Located near the Antarctic Circumpolar Current. *Front Microbiol*, **2018**. *9*: p. 1474.

46. Kodama, Y., Suzuki, H., Dohra, H., Sugii, M., Kitazume, T., Yamaguchi, K., Shigenobu, S., and Fujishima, M., Comparison of Gene Expression of Paramecium Bursaria with and without Chlorella Variabilis Symbionts. *BMC Genomics*, **2014**. *15*: p. 183.

47. Bubeck, J.A. and Pfitzner, A.J.P., Isolation and Characterization of a New Type of Chlorovirus That Infects an Endosymbiotic Chlorella Strain of the Heliozoon Acanthocystis Turfacea. *Journal of General Virology*, **2005**. *86*(10): p. 2871-2877.

48. He, M., Wang, J., Fan, X., Liu, X., Shi, W., Huang, N., Zhao, F., and Miao, M., Genetic Basis for the Establishment of Endosymbiosis in Paramecium. *ISME J*, **2019**. *13*(5): p. 1360-1369.

49. Dean, A.D., Minter, E.J., Sorensen, M.E., Lowe, C.D., Cameron, D.D., Brockhurst, M.A., and Jamie Wood, A., Host Control and Nutrient Trading in a Photosynthetic Symbiosis. *J Theor Biol*, **2016**. *405*: p. 82-93.

50. Kodama, Y. and Fujishima, M., Timing of Perialgal Vacuole Membrane Differentiation from Digestive Vacuole Membrane in Infection of Symbiotic Algae Chlorella Vulgaris of the Ciliate Paramecium Bursaria. *Protist*, **2009**. *160*(1): p. 65-74.

51. Quispe, C.F., Esmael, A., Sonderman, O., McQuinn, M., Agarkova, I., Battah, M., Duncan, G.A., Dunigan, D.D., Smith, T.P., and De Castro, C., Characterization of a New Chlorovirus Type with Permissive and Non-Permissive Features on Phylogenetically Related Algal Strains. *Virology*, **2017**. *500*: p. 103-113.

52. Jeanniard, A., Dunigan, D.D., Gurnon, J.R., Agarkova, I.V., Kang, M., Vitek, J., Duncan, G., McClung, O.W., Larsen, M., and Claverie, J.-M., Towards Defining the Chloroviruses: A Genomic Journey through a Genus of Large DNA Viruses. *BMC genomics*, **2013**. *14*(1): p. 158.

53. Frickel, J., Theodosiou, L., and Becks, L., Rapid Evolution of Hosts Begets Species Diversity at the Cost of Intraspecific Diversity. *Proc Natl Acad Sci U S A*, **2017**. *114*(42): p. 11193-11198.

54. DeLong, J.P., Al-Ameeli, Z., Duncan, G., Van Etten, J.L., and Dunigan, D.D., Predators Catalyze an Increase in Chloroviruses by Foraging on the Symbiotic Hosts of Zoochlorellae. *Proc Natl Acad Sci U S A*, **2016**. *113*(48): p. 13780-13784.

55. DeLong, J.P., Al-Ameeli, Z., Lyon, S., Van Etten, J.L., and Dunigan, D.D., Size-Dependent Catalysis of Chlorovirus Population Growth by a Messy Feeding Predator. *Microb Ecol*, **2018**. *75*(4): p. 847-853.

56. Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Hartigan, J., and Van Etten, J.L., Sequence and Annotation of the 314-Kb Mt325 and the 321-Kb Fr483 Viruses That Infect Chlorella Pbi. *Virology*, **2007**. *358*(2): p. 459-471.

57. Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Nierman, W.C., and Van Etten, J.L., Sequence and Annotation of the 369-Kb Ny-2a and the 345-Kb Ar158 Viruses That Infect Chlorella Nc64a. *Virology*, **2007**. *358*(2): p. 472-484.

58. Fitzgerald, L.A., Graves, M.V., Li, X., Hartigan, J., Pfitzner, A.J., Hoffart, E., and Van Etten, J.L., Sequence and Annotation of the 288-Kb Atcv-1 Virus That Infects an Endosymbiotic Chlorella Strain of the Heliozoon Acanthocystis Turfacea. *Virology*, **2007**. *362*(2): p. 350-361.

59. Dunigan, D.D., Cerny, R.L., Bauman, A.T., Roach, J.C., Lane, L.C., Agarkova, I.V., Wulser, K., Yanai-Balser, G.M., Gurnon, J.R., and Vitek, J.C., Paramecium Bursaria Chlorella Virus 1 Proteome Reveals Novel Architectural and Regulatory Features of a Giant Virus. *Journal of virology*, **2012**. *86*(16): p. 8821-8834.

60. Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G.F., Rock, D.L., and Van Etten, J.L., Analysis of 74 Kb of DNA Located at the Right End of the 330-Kb Chlorella Virus Pbcv-1 Genome. *Virology*, **1997**. *237*(2): p. 360-77.

61. Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D.D., Grigoriev, I.V., Claverie, J.M., and Van Etten, J.L., The Chlorella Variabilis Nc64a Genome Reveals Adaptation to Photosymbiosis, Coevolution with Viruses, and Cryptic Sex. *Plant Cell*, **2010**. *22*(9): p. 2943-55.

62. Yan, X., Olson, N.H., Van Etten, J.L., Bergoin, M., Rossmann, M.G., and Baker, T.S., Structure and Assembly of Large Lipid-Containing Dsdna Viruses. *Nature structural biology*, **2000**. *7*(2): p. 101-103.

63. Cherrier, M.V., Kostyuchenko, V.A., Xiao, C., Bowman, V.D., Battisti, A.J., Yan, X., Chipman, P.R., Baker, T.S., Van Etten, J.L., and Rossmann, M.G., An Icosahedral Algal Virus Has a Complex Unique Vertex Decorated by a Spike. *Proc Natl Acad Sci U S A*, **2009**. *106*(27): p. 11085-11089.

64. Zhang, X., Xiang, Y., Dunigan, D.D., Klose, T., Chipman, P.R., Van Etten, J.L., and Rossmann, M.G., Three-Dimensional Structure and Function of the <Em>Paramecium Bursaria</Em> Chlorella Virus Capsid. *Proceedings of the National Academy of Sciences*, **2011**. *108*(36): p. 14837-14842.

65. Fang, Q., Zhu, D., Agarkova, I., Adhikari, J., Klose, T., Liu, Y., Chen, Z., Sun, Y., Gross, M.L., Van Etten, J.L., Zhang, X., and Rossmann, M.G., Near-Atomic Structure of a Giant Virus. *Nature Communications*, **2019**. *10*(1): p. 388.

66. Becker, B., Lesemann, D.E., and Reisser, W., Ultrastructural Studies on a Chlorella Virus from Germany. *Archives of Virology*, **1993**. *130*(1): p. 145-155.

67. Zhu, D., Wang, X., Fang, Q., Van Etten, J.L., Rossmann, M.G., Rao, Z., and Zhang, X., Pushing the Resolution Limit by Correcting the Ewald Sphere Effect in Single-Particle Cryo-Em Reconstructions. *Nature Communications*, **2018**. *9*(1): p. 1552.

68. Zhang, Y., Strasser, P., Grabherr, R., and Van Etten, J.L., Hairpin Loop Structure at the Termini of the Chlorella Virus Pbcv-1 Genome. *Virology*, **1994**. *202*(2): p. 1079-82.

69. Strasser, P., Zhang, Y.P., Rohozinski, J., and Van Etten, J.L., The Termini of the Chlorella Virus Pbcv-1 Genome Are Identical 2.2-Kbp Inverted Repeats. *Virology*, **1991**. *180*(2): p. 763-9.

70. Yamada, T., Onimatsu, H., and Van Etten, J., *Advances in Virus Research*. 2006, Academic.

71. Blanc, G., Mozar, M., Agarkova, I.V., Gurnon, J.R., Yanai-Balser, G., Rowe, J.M., Xia, Y., Riethoven, J.J., Dunigan, D.D., and Van Etten, J.L., Deep Rna Sequencing Reveals Hidden Features and Dynamics of Early Gene Transcription in Paramecium Bursaria Chlorella Virus 1. *PLoS One*, **2014**. *9*(3): p. e90989.

72. Nelson, M., Burbank, D.E., and Van Etten, J.L., Chlorella Viruses Encode Multiple DNA Methyltransferases. *Biological chemistry*, **1998**. *379*(4-5): p. 423-428.

73. Agarkova, I.V., Dunigan, D.D., and Van Etten, J.L., Virion-Associated Restriction Endonucleases of Chloroviruses. *J Virol*, **2006**. *80*(16): p. 8114-23.

74. Chan, S.H., Zhu, Z., Dunigan, D.D., Van Etten, J.L., and Xu, S.Y., Cloning of Nt.Cviqii Nicking Endonuclease and Its Cognate Methyltransferase: M.Cviqii Methylates Ag Sequences. *Protein Expr Purif*, **2006**. *49*(1): p. 138-50.

75. Nelson, M., Zhang, Y., and Van Etten, J.L., DNA Methyltransferases and DNA Site-Specific Endonucleases Encoded by Chlorella Viruses. *Exs*, **1993**. *64*: p. 186-211.

76. Zhang, X., Xiang, Y., Dunigan, D.D., Klose, T., Chipman, P.R., Van Etten, J.L., and Rossmann, M.G., Three-Dimensional Structure and Function of the Paramecium Bursaria Chlorella Virus Capsid. *Proceedings of the National Academy of Sciences*, **2011**. *108*(36): p. 14837-14842.

77. Onimatsu, H., Sugimoto, I., Fujie, M., Usami, S., and Yamada, T., Vp130, a Chloroviral Surface Protein That Interacts with the Host Chlorella Cell Wall. *Virology*, **2004**. *319*(1): p. 71-80.

78. Onimatsu, H., Suganuma, K., Uenoyama, S., and Yamada, T., C-Terminal Repetitive Motifs in Vp130 Present at the Unique Vertex of the Chlorovirus Capsid Are Essential for Binding to the Host Chlorella Cell Wall. *Virology*, **2006**. *353*(2): p. 433-42.

79. Meints, R.H., Lee, K., Burbank, D.E., and Van Etten, J.L., Infection of a Chlorella-Like Alga with the Virus, Pbcv-1: Ultrastructural Studies. *Virology*, **1984**. *138*(2): p. 341-6.

80. Agarkova, I., Hertel, B., Zhang, X., Lane, L., Tchourbanov, A., Dunigan, D.D., Thiel, G., Rossmann, M.G., and Van Etten, J.L., Dynamic Attachment of Chlorovirus Pbcv-1 to Chlorella Variabilis. *Virology*, **2014**. *466-467*: p. 95-102.

81. Meints, R.H., Burbank, D.E., Van Etten, J.L., and Lamport, D.T., Properties of the Chlorella Receptor for the Virus Pbcv-1. *Virology*, **1988**. *164*(1): p. 15-21.

82. Kang, M., Dunigan, D.D., and JL, V.A.N.E., Chlorovirus: A Genus of Phycodnaviridae That Infects Certain Chlorella-Like Green Algae. *Mol Plant Pathol*, **2005**. *6*(3): p. 213-24.

83. Milrot, E., Shimoni, E., Dadosh, T., Rechav, K., Unger, T., Van Etten, J.L., and Minsky, A., Structural Studies Demonstrating a Bacteriophage-Like Replication Cycle of the Eukaryote-Infecting Paramecium Bursaria Chlorella Virus-1. *PLoS Pathog*, **2017**. *13*(8): p. e1006562.

84. Sun, L., Adams, B., Gurnon, J.R., Ye, Y., and Van Etten, J.L., Characterization of Two Chitinase Genes and One Chitosanase Gene Encoded by Chlorella Virus Pbcv-1. *Virology*, **1999**. *263*(2): p. 376-87.

85. Sun, L., Gurnon, J.R., Adams, B.J., Graves, M.V., and Van Etten, J.L., Characterization of a Beta-1,3-Glucanase Encoded by Chlorella Virus Pbcv-1. *Virology*, **2000**. *276*(1): p. 27-36.

86. Suda, K., Tanji, Y., Hori, K., and Unno, H., Evidence for a Novel Chlorella Virus-Encoded Alginate Lyase. *FEMS Microbiol Lett*, **1999**. *180*(1): p. 45-53.

87. Sugimoto, I., Onimatsu, H., Fujie, M., Usami, S., and Yamada, T., Val-1, a Novel Polysaccharide Lyase Encoded by Chlorovirus Cvk2. *FEBS Lett*, **2004**. *559*(1-3): p. 51-6.

88. Ogura, K., Yamasaki, M., Yamada, T., Mikami, B., Hashimoto, W., and Murata, K., Crystal Structure of Family 14 Polysaccharide Lyase with Ph-Dependent Modes of Action. *J Biol Chem*, **2009**. *284*(51): p. 35572-9.

89. Romani, G., Piotrowski, A., Hillmer, S., Gurnon, J., Van Etten, J.L., Moroni, A., Thiel, G., and Hertel, B., A Virus-Encoded Potassium Ion Channel Is a Structural Protein in the Chlorovirus Paramecium Bursaria Chlorella Virus 1 Virion. *J Gen Virol*, **2013**. *94*(Pt 11): p. 2549-2556.

90. Thiel, G., Moroni, A., Dunigan, D., and Van Etten, J.L., Initial Events Associated with Virus Pbcv-1 Infection of Chlorella Nc64a. *Prog Bot*, **2010**. *71*(3): p. 169-183.

91. Qian, C., Wang, X., Manzur, K., Sachchidanand, Farooq, A., Zeng, L., Wang, R., and Zhou, M.M., Structural Insights of the Specificity and Catalysis of a Viral Histone H3 Lysine 27 Methyltransferase. *J Mol Biol*, **2006**. *359*(1): p. 86-96.

92. Rowe, J.M., Jeanniard, A., Gurnon, J.R., Xia, Y., Dunigan, D.D., Van Etten, J.L., and Blanc, G., Global Analysis of Chlorella Variabilis Nc64a Mrna Profiles During the Early Phase of Paramecium Bursaria Chlorella Virus-1 Infection. *PLoS One*, **2014**. *9*(3): p. e90988.

93. Milrot, E., Mutsafi, Y., Fridmann-Sirkis, Y., Shimoni, E., Rechav, K., Gurnon, J.R., Van Etten, J.L., and Minsky, A., Virus-Host Interactions: Insights from the Replication Cycle of the Large Paramecium Bursaria Chlorella Virus. *Cell Microbiol*, **2016**. *18*(1): p. 3-16.

94. Van Etten, J.L., Burbank, D.E., Xia, Y., and Meints, R.H., Growth Cycle of a Virus, Pbcv-1, That Infects Chlorella-Like Algae. *Virology*, **1983**. *126*(1): p. 117-25.

95. Markine-Goriaynoff, N., Gillet, L., Van Etten, J.L., Korres, H., Verma, N., and Vanderplasschen, A., Glycosyltransferases Encoded by Viruses. *J Gen Virol*, **2004**. *85*(Pt 10): p. 2741-2754.

96. Van Etten, J.L., Gurnon, J.R., Yanai-Balser, G.M., Dunigan, D.D., and Graves, M.V., Chlorella Viruses Encode Most, If Not All, of the Machinery to Glycosylate Their Glycoproteins Independent of the Endoplasmic Reticulum and Golgi. *Biochim Biophys Acta*, **2010**. *1800*(2): p. 152-9.

97. Van Etten, J.L., Agarkova, I., Dunigan, D.D., Tonetti, M., De Castro, C., and Duncan, G.A., Chloroviruses Have a Sweet Tooth. *Viruses*, **2017**. *9*(4).

98. Landstein, D., Graves, M.V., Burbank, D.E., DeAngelis, P., and Van Etten, J.L., Chlorella Virus Pbcv-1 Encodes Functional Glutamine: Fructose-6-Phosphate Amidotransferase and Udp-Glucose Dehydrogenase Enzymes. *Virology*, **1998**. *250*(2): p. 388-96.

99. DeAngelis, P.L., Jing, W., Graves, M.V., Burbank, D.E., and Van Etten, J.L., Hyaluronan Synthase of Chlorella Virus Pbcv-1. *Science*, **1997**. *278*(5344): p. 1800-1803.

100. DeAngelis, P.L., Hyaluronan Synthases: Fascinating Glycosyltransferases from Vertebrates, Bacterial Pathogens, and Algal Viruses. *Cell Mol Life Sci*, **1999**. *56*(7-8): p. 670-82.

101. DeAngelis, P.L., Evolution of Glycosaminoglycans and Their Glycosyltransferases: Implications for the Extracellular Matrices of Animals and the Capsules of Pathogenic Bacteria. *Anat Rec*, **2002**. *268*(3): p. 317-26.

102. Graves, M.V., Burbank, D.E., Roth, R., Heuser, J., DeAngelis, P.L., and Van Etten, J.L., Hyaluronan Synthesis in Virus Pbcv-1-Infected Chlorella-Like Green Algae. *Virology*, **1999**. *257*(1): p. 15-23.

103. Kawasaki, T., Tanaka, M., Fujie, M., Usami, S., Sakai, K., and Yamada, T., Chitin Synthesis in Chlorovirus Cvk2-Infected Chlorella Cells. *Virology*, **2002**. *302*(1): p. 123-31.

104. Yamada, T. and Kawasaki, T., Microbial Synthesis of Hyaluronan and Chitin: New Approaches. *J Biosci Bioeng*, **2005**. *99*(6): p. 521-8.

105. Tonetti, M., Zanardi, D., Gurnon, J.R., Fruscione, F., Armirotti, A., Damonte, G., Sturla, L., De Flora, A., and Van Etten, J.L., Paramecium Bursaria Chlorella Virus 1 Encodes Two Enzymes Involved in the Biosynthesis of Gdp-L-Fucose and Gdp-D-Rhamnose. *J Biol Chem*, **2003**. *278*(24): p. 21559-65.

106. Fruscione, F., Sturla, L., Duncan, G., Van Etten, J.L., Valbuzzi, P., De Flora, A., Di Zanni, E., and Tonetti, M., Differential Role of Nadp+ and Nadph in the Activity and Structure of Gdp-D-Mannose 4,6-Dehydratase from Two Chlorella Viruses. *J Biol Chem*, **2008**. *283*(1): p. 184-93.

107. De Castro, C., Speciale, I., Duncan, G., Dunigan, D.D., Agarkova, I., Lanzetta, R., Sturiale, L., Palmigiano, A., Garozzo, D., Molinaro, A., Tonetti, M., and Van Etten, J.L., N-Linked Glycans of Chloroviruses Sharing a Core Architecture without Precedent. *Angew Chem Int Ed Engl*, **2016**. *55*(2): p. 654-8.

108. Parakkottil Chothi, M., Duncan, G.A., Armirotti, A., Abergel, C., Gurnon, J.R., Van Etten, J.L., Bernardi, C., Damonte, G., and Tonetti, M., Identification of an L-Rhamnose Synthetic Pathway in Two Nucleocytoplasmic Large DNA Viruses. *J Virol*, **2010**. *84*(17): p. 8829-38.

109. Vigerust, D.J. and Shepherd, V.L., Virus Glycosylation: Role in Virulence and Immune Interactions. *Trends Microbiol*, **2007**. *15*(5): p. 211-8.

110. Wang, I.N., Li, Y., Que, Q., Bhattacharya, M., Lane, L.C., Chaney, W.G., and Van Etten, J.L., Evidence for Virus-Encoded Glycosylation Specificity. *Proc Natl Acad Sci U S A*, **1993**. *90*(9): p. 3840-4.

111. Speciale, I., Duncan, G.A., Unione, L., Agarkova, I.V., Garozzo, D., Jimenez-Barbero, J., Lin, S., Lowary, T.L., Molinaro, A., Noel, E., Laugieri, M.E., Tonetti, M.G., Van Etten, J.L., and De Castro, C., The N-Glycan Structures of the Antigenic Variants of Chlorovirus Pbcv-1 Major Capsid Protein Help to Identify the Virus-Encoded Glycosyltransferases. *J Biol Chem*, **2019**. *294*(14): p. 5688-5699.

112. Etten, J.L., Burbank, D.E., Xia, Y., and Meints, R.H., Growth Cycle of a Virus, Pbcv-1, That Infectschlorella-Like Algae. *Virology*, **1983**. *126*.

113. Graves, M.V., Bernadt, C.T., Cerny, R., and Van Etten, J.L., Molecular and Genetic Evidence for a Virus-Encoded Glycosyltransferase Involved in Protein Glycosylation. *Virology*, **2001**. *285*(2): p. 332-45.

114. Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V., Van Etten, J.L., and Rossmann, M.G., The Structure and Evolution of the Major Capsid Protein of a Large, Lipid-Containing DNA Virus. *Proc Natl Acad Sci U S A*, **2002**. *99*(23): p. 14758-63.

115. De Castro, C., Molinaro, A., Piacente, F., Gurnon, J.R., Sturiale, L., Palmigiano, A., Lanzetta, R., Parrilli, M., Garozzo, D., Tonetti, M.G., and Van Etten, J.L., Structure of N-Linked Oligosaccharides Attached to Chlorovirus Pbcv-1 Major Capsid Protein Reveals Unusual Class of Complex N-Glycans. *Proc Natl Acad Sci U S A*, **2013**. *110*(34): p. 13956-60.

116. Schwarz, F. and Aebi, M., Mechanisms and Principles of N-Linked Protein Glycosylation. *Curr Opin Struct Biol*, **2011**. *21*(5): p. 576-82.

117. De Castro, C., Klose, T., Speciale, I., Lanzetta, R., Molinaro, A., Van Etten, J.L., and Rossmann, M.G., Structure of the Chlorovirus Pbcv-1 Major Capsid Glycoprotein Determined by Combining Crystallographic and Carbohydrate Molecular Modeling Approaches. *Proc Natl Acad Sci U S A*, **2018**. *115*(1): p. E44-E52.

118. Wieland, F., Heitzer, R., and Schaefer, W., Asparaginylglucose: Novel Type of Carbohydrate Linkage. *Proc Natl Acad Sci U S A*, **1983**. *80*(18): p. 5470-4.

119. Mengele, R. and Sumper, M., Drastic Differences in Glycosylation of Related S-Layer Glycoproteins from Moderate and Extreme Halophiles. *J Biol Chem*, **1992**. *267*(12): p. 8182-5.

120. Schreiner, R., Schnabel, E., and Wieland, F., Novel N-Glycosylation in Eukaryotes: Laminin Contains the Linkage Unit Beta-Glucosylasparagine. *J Cell Biol*, **1994**. *124*(6): p. 1071-81.

121. Gross, J., Grass, S., Davis, A.E., Gilmore-Erdmann, P., Townsend, R.R., and St Geme, J.W., 3rd, The Haemophilus Influenzae Hmw1 Adhesin Is a Glycoprotein with an Unusual N-Linked Carbohydrate Modification. *J Biol Chem*, **2008**. *283*(38): p. 26010-5.

122. Que, Q., Li, Y., Wang, I.N., Lane, L.C., Chaney, W.G., and Van Etten, J.L., Protein Glycosylation and Myristylation in Chlorella Virus Pbcv-1 and Its Antigenic Variants. *Virology*, **1994**. *203*(2): p. 320-7.

123. Chuchird, N., Nishida, K., Kawasaki, T., Fujie, M., Usami, S., and Yamada, T., A Variable Region on the Chlorovirus Cvk2 Genome Contains Five Copies of the Gene for Vp260, a Viral-Surface Glycoprotein. *Virology*, **2002**. *295*(2): p. 289-98.

124. Yanai-Balser, G.M., Duncan, G.A., Eudy, J.D., Wang, D., Li, X., Agarkova, I.V., Dunigan, D.D., and Van Etten, J.L., Microarray Analysis of Paramecium Bursaria Chlorella Virus 1 Transcription. *J Virol*, **2010**. *84*(1): p. 532-42.

125. Mitra, A. and Higgins, D.W., The Chlorella Virus Adenine Methyltransferase Gene Promoter Is a Strong Promoter in Plants. *Plant molecular biology*, **1994**. *26*(1): p. 85-93.

126. Mitra, A., Higgins, D.W., and Rohe, N.J., A Chlorella Virus Gene Promoter Functions as a Strong Promoter Both in Plants and Bacteria. *Biochemical and biophysical research communications*, **1994**. *204*(1): p. 187-194.

127. Nguyen, P.S., Falcone, D.L., and Graves, M.V., The A312l 5'-Utr of Chlorella Virus Pbcv-1 Is a Translational Enhancer in Arabidopsis Thaliana. *Virus Res*, **2009**. *140*(1-2): p. 138-46.

128. Plugge, B., Gazzarrini, S., Nelson, M., Cerana, R., Van Etten, J.L., Derst, C., DiFrancesco, D., Moroni, A., and Thiel, G., A Potassium Channel Protein Encoded by Chlorella Virus Pbcv-1. *Science*, **2000**. *287*(5458): p. 1641-4.

129. Lim, S.L., Chu, W.L., and Phang, S.M., Use of Chlorella Vulgaris for Bioremediation of Textile Wastewater. *Bioresour Technol*, **2010**. *101*(19): p. 7314-22.

130. Gonzalez, R., Garcia-Balboa, C., Rouco, M., Lopez-Rodas, V., and Costas, E., Adaptation of Microalgae to Lindane: A New Approach for Bioremediation. *Aquat Toxicol*, **2012**. *109*: p. 25-32.

131. Ng, I.S., Tan, S.I., Kao, P.H., Chang, Y.K., and Chang, J.S., Recent Developments on Genetic Engineering of Microalgae for Biofuels and Bio-Based Chemicals. *Biotechnol J*, **2017**. *12*(10).

132. Jeon, K., Suresh, A., and Kim, Y.-C., Highly Efficient Molecular Delivery into Chlamydomonas Reinhardtii by Electroporation. *Korean Journal of Chemical Engineering*, **2013**. *30*(8): p. 1626-1630.

133. Kumar, S.V., Misquitta, R.W., Reddy, V.S., Rao, B.J., and Rajam, M.V., Genetic Transformation of the Green Alga—Chlamydomonas Reinhardtii by Agrobacterium Tumefaciens. *Plant Science*, **2004**. *166*(3): p. 731-738.

134. Tan, C., Qin, S., Zhang, Q., Jiang, P., and Zhao, F., Establishment of a Micro-Particle Bombardment Transformation System for Dunaliella Salina. *The Journal of Microbiology*, **2005**. *43*(4): p. 361-365.

135. Kindle, K.L., High-Frequency Nuclear Transformation of Chlamydomonas Reinhardtii. *Proceedings of the National Academy of Sciences*, **1990**. *87*(3): p. 1228-1232.

136. Sanitha, M., Radha, S., Fatima, A.A., Devi, S.G., and Ramya, M., Agrobacterium-Mediated Transformation of Three Freshwater Microalgal Strains. *Pol J Microbiol*, **2014**. *63*(4): p. 387-382.

137. Suttangkakul, A., Sirikhachornkit, A., Juntawong, P., Puangtame, W., Chomtong, T., Srifa, S., Sathitnaitham, S., Dumrongthawatchai, W., Jariyachawalid, K., and Vuttipongchaikij, S., Evaluation of Strategies for Improving the Transgene Expression in an Oleaginous Microalga Scenedesmus Acutus. *BMC biotechnology*, **2019**. *19*(1): p. 4.

138. Gimpel, J.A., Henríquez, V., and Mayfield, S.P., In Metabolic Engineering of Eukaryotic Microalgae: Potential and Challenges Come with Great Diversity. *Frontiers in microbiology*, **2015**. *6*: p. 1376.

139. Northcote, D., Goulding, K., and Horne, R., The Chemical Composition and Structure of the Cell Wall of Chlorella Pyrenoidosa. *Biochemical Journal*, **1958**. *70*(3): p. 391.

140. Kim, D.-H., Kim, Y.T., Cho, J.J., Bae, J.-H., Hur, S.-B., Hwang, I., and Choi, T.-J., Stable Integration and Functional Expression of Flounder Growth Hormone Gene in Transformed Microalga, Chlorella Ellipsoidea. *Marine Biotechnology*, **2002**. *4*(1): p. 63-73.

141. San Cha, T., Yee, W., and Aziz, A., Assessment of Factors Affecting Agrobacterium-Mediated Genetic Transformation of the Unicellular Green Alga, Chlorella Vulgaris. *World Journal of Microbiology and Biotechnology*, **2012**. *28*(4): p. 1771-1779.

142. Lin, H.-D., Liu, B.-H., Kuo, T.-T., Tsai, H.-C., Feng, T.-Y., Huang, C.-C., and Chien, L.-F., Knockdown of Psbo Leads to Induction of Hyda and Production of Photobiological H2 in the Green Alga Chlorella Sp. Dt. *Bioresource technology*, **2013**. *143*: p. 154-162.

143. Lou, S., Wang, L., He, L., Wang, Z., Wang, G., and Lin, X., Production of Crocetin in Transgenic Chlorella Vulgaris Expressing Genes Crtrb and Zcd1. *Journal of Applied Phycology*, **2016**. *28*(3): p. 1657-1665.

144. Yang, B., Liu, J., Jiang, Y., and Chen, F., Chlorella Species as Hosts for Genetic Engineering and Expression of Heterologous Proteins: Progress, Challenge and Perspective. *Biotechnology journal*, **2016**. *11*(10): p. 1244-1261.

145. Kumar, M., Jeon, J., Choi, J., and Kim, S.-R., Rapid and Efficient Genetic Transformation of the Green Microalga Chlorella Vulgaris. *Journal of Applied Phycology*, **2018**. *30*(3): p. 1735-1745.

146. Liu, L., Wang, Y., Zhang, Y., Chen, X., Zhang, P., and Ma, S., Development of a New Method for Genetic Transformation of the Green Alga Chlorella Ellipsoidea. *Molecular biotechnology*, **2013**. *54*(2): p. 211-219.

147. Niu, Y.F., Zhang, M.H., Xie, W.H., Li, J., Gao, Y., Yang, W.D., Liu, J.S., and Li, H.Y., A New Inducible Expression System in a Transformed Green Alga, Chlorella Vulgaris. *Genet Mol Res*, **2011**. *10*(4): p. 3427-34.

148. Chien, L.-F., Kuo, T.-T., Liu, B.-H., Lin, H.-D., Feng, T.-Y., and Huang, C.-C., Solar-to-Bioh2 Production Enhanced by Homologous Overexpression of Hydrogenase in Green Alga Chlorella Sp. Dt. *international journal of hydrogen energy*, **2012**. *37*(23): p. 17738-17748.

149. Bai, L.-L., Yin, W.-B., Chen, Y.-H., Niu, L.-L., Sun, Y.-R., Zhao, S.-M., Yang, F.-Q., Wang, R.R.-C., Wu, Q., and Zhang, X.-Q., A New Strategy to Produce a Defensin: Stable Production of Mutated Np-1 in Nitrate Reductase-Deficient Chlorella Ellipsoidea. *PloS one*, **2013**. *8*(1).

150. Guo, S.-L., Zhao, X.-Q., Tang, Y., Wan, C., Alam, M.A., Ho, S.-H., Bai, F.-W., and Chang, J.-S., Establishment of an Efficient Genetic Transformation System in Scenedesmus Obliquus. *Journal of biotechnology*, **2013**. *163*(1): p. 61-68.

151. Kilian, O., Benemann, C.S., Niyogi, K.K., and Vick, B., High-Efficiency Homologous Recombination in the Oil-Producing Alga Nannochloropsis Sp. *Proceedings of the National Academy of Sciences*, **2011**. *108*(52): p. 21265-21269.

152. Muñoz, C.F., de Jaeger, L., Sturme, M.H., Lip, K.Y., Olijslager, J.W., Springer, J., Wolbert, E.J., Martens, D.E., Eggink, G., and Weusthuis, R.A., Improved DNA/Protein Delivery in Microalgae–a Simple and Reliable Method for the Prediction of Optimal Electroporation Settings. *Algal research*, **2018**. *33*: p. 448-455.

153. Guihéneuf, F., Khan, A., and Tran, L.-S.P., Genetic Engineering: A Promising Tool to Engender Physiological, Biochemical, and Molecular Stress Resilience in Green Microalgae. *Frontiers in plant science*, **2016**. *7*: p. 400.

154. Karas, B.J., Diner, R.E., Lefebvre, S.C., McQuaid, J., Phillips, A.P., Noddings, C.M., Brunson, J.K., Valas, R.E., Deerinck, T.J., and Jablanovic, J., Designer Diatom Episomes Delivered by Bacterial Conjugation. *Nature communications*, **2015**. *6*(1): p. 1-10.

155. Diner, R.E., Bielinski, V.A., Dupont, C.L., Allen, A.E., and Weyman, P.D., Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods. *Frontiers in bioengineering and biotechnology*, **2016**. *4*: p. 65.

156. Muñoz, C.F., Sturme, M.H., D'Adamo, S., Weusthuis, R.A., and Wijffels, R.H., Stable Transformation of the Green Algae Acutodesmus Obliquus and Neochloris Oleoabundans Based on E. Coli Conjugation. *Algal research*, **2019**. *39*: p. 101453.

157. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F., Genome Engineering Using the Crispr-Cas9 System. *Nature protocols*, **2013**. *8*(11): p. 2281.

158. Horvath, P. and Barrangou, R., Crispr/Cas, the Immune System of Bacteria and Archaea. *Science*, **2010**. *327*(5962): p. 167-170.

159. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., and Doench, J.G., Genome-Scale Crispr-Cas9 Knockout Screening in Human Cells. *Science*, **2014**. *343*(6166): p. 84-87.

160. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., and Marraffini, L.A., Multiplex Genome Engineering Using Crispr/Cas Systems. *Science*, **2013**. *339*(6121): p. 819-823.

161. Nymark, M., Sharma, A.K., Sparstad, T., Bones, A.M., and Winge, P., A Crispr/Cas9 System Adapted for Gene Editing in Marine Algae. *Scientific reports*, **2016**. *6*: p. 24951.

162. Patel, V.K., Soni, N., Prasad, V., Sapre, A., Dasgupta, S., and Bhadra, B., Crispr–Cas9 System for Genome Engineering of Photosynthetic Microalgae. *Molecular biotechnology*, **2019**: p. 1-21.

163. Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L.B., Coppersmith, J., Kwok, K., Peach, L., Orchard, E., Kalb, R., and Xu, W., Lipid Production in Nannochloropsis Gaditana Is Doubled by Decreasing Expression of a Single Transcriptional Regulator. *Nature biotechnology*, **2017**. *35*(7): p. 647.

164. Wang, Q., Lu, Y., Xin, Y., Wei, L., Huang, S., and Xu, J., Genome Editing of Model Oleaginous Microalgae Nannochloropsis Spp. By Crispr/Cas9. *The Plant Journal*, **2016**. *88*(6): p. 1071-1081.

165. Jinkerson, R.E. and Jonikas, M.C., Molecular Techniques to Interrogate and Edit the Chlamydomonas Nuclear Genome. *The Plant Journal*, **2015**. *82*(3): p. 393-412.

166. Baek, K., Kim, D.H., Jeong, J., Sim, S.J., Melis, A., Kim, J.-S., Jin, E., and Bae, S., DNA-Free Two-Gene Knockout in Chlamydomonas Reinhardtii Via Crispr-Cas9 Ribonucleoproteins. *Scientific Reports*, **2016**. *6*: p. 30620.

167. Jiang, W., Brueggeman, A.J., Horken, K.M., Plucinak, T.M., and Weeks, D.P., Successful Transient Expression of Cas9 and Single Guide Rna Genes in Chlamydomonas Reinhardtii. *Eukaryotic cell*, **2014**. *13*(11): p. 1465-1469.

168. Jiang, W.Z. and Weeks, D.P., A Gene-within-a-Gene Cas9/Sgrna Hybrid Construct Enables Gene Editing and Gene Replacement Strategies in Chlamydomonas Reinhardtii. *Algal research*, **2017**. *26*: p. 474-480.

169. Guzmán-Zapata, D., Sandoval-Vargas, J.M., Macedo-Osorio, K.S., Salgado-Manjarrez, E., Castrejón-Flores, J.L., Oliver-Salvador, M.d.C., Durán-Figueroa, N.V., Nogué, F., and Badillo-Corona, J.A., Efficient Editing of the Nuclear Apt Reporter Gene in Chlamydomonas Reinhardtii Via Expression of a Crispr-Cas9 Module. *International journal of molecular sciences*, **2019**. *20*(5): p. 1247.

170. Ferenczi, A., Pyott, D.E., Xipnitou, A., and Molnar, A., Efficient Targeted DNA Editing and Replacement in Chlamydomonas Reinhardtii Using Cpf1 Ribonucleoproteins and Single-Stranded DNA. *Proceedings of the National Academy of Sciences*, **2017**. *114*(51): p. 13567-13572.

171. Sharma, A.K., Nymark, M., Sparstad, T., Bones, A.M., and Winge, P., Transgene-Free Genome Editing in Marine Algae by Bacterial Conjugation–Comparison with Biolistic Crispr/Cas9 Transformation. *Scientific reports*, **2018**. *8*(1): p. 1-11.

172. Greiner, A., Kelterborn, S., Evers, H., Kreimer, G., Sizova, I., and Hegemann, P., Targeting of Photoreceptor Genes in Chlamydomonas Reinhardtii Via Zinc-Finger Nucleases and Crispr/Cas9. *The Plant Cell*, **2017**. *29*(10): p. 2498-2518.

173. Shin, S.-E., Lim, J.-M., Koh, H.G., Kim, E.K., Kang, N.K., Jeon, S., Kwon, S., Shin, W.-S., Lee, B., and Hwangbo, K., Crispr/Cas9-Induced Knockout and Knock-in Mutations in Chlamydomonas Reinhardtii. *Scientific reports*, **2016**. *6*(1): p. 1-15.

174. Yoshimitsu, Y., Abe, J., and Harayama, S., Cas9-Guide Rna Ribonucleoprotein-Induced Genome Editing in the Industrial Green Alga Coccomyxa Sp. Strain Kj. *Biotechnology for biofuels*, **2018**. *11*(1): p. 1-10.

175. Speciale, I., Laugieri, M.E., Noel, E., Lin, S., Lowary, T.L., Molinaro, A., Duncan, G.A., Agarkova, I.V., Garozzo, D., and Tonetti, M.G., Chlorovirus Pbcv-1 Protein A064r Has Three of the Transferase Activities Necessary to Synthesize Its Capsid Protein N-Linked Glycans. *Proceedings of the National Academy of Sciences*, **2020**. *117*(46): p. 28735-28742.

176. Zorin, B., Hegemann, P., and Sizova, I., Nuclear-Gene Targeting by Using Single-Stranded DNA Avoids Illegitimate DNA Integration in Chlamydomonas Reinhardtii. *Eukaryotic cell*, **2005**. *4*(7): p. 1264-1272.

177. Van Etten, J.L., Burbank, D.E., Xia, Y., and Meints, R.H., Growth Cycle of a Virus, Pbcv-1, That Infects Chlorella-Like Algae. *Virology*, **1983**. *126*(1): p. 117-125.

178. Van Etten, J.L., Meints, R.H., Burbank, D.E., Kuczmarski, D., Cuppels, D.A., and Lane, L.C., Isolation and Characterization of a Virus from the Intracellular Green Alga Symbiotic with Hydra Viridis. *Virology*, **1981**. *113*(2): p. 704-711.

179. Popper, Z.A. and Tuohy, M.G., Beyond the Green: Understanding the Evolutionary Puzzle of Plant and Algal Cell Walls. *Plant physiology*, **2010**. *153*(2): p. 373-383.

180. Yamada, T. and Sakaguchi, K., Comparative Studies Onchlorella Cell Walls: Induction of Protoplast Formation. *Archives of Microbiology*, **1982**. *132*(1): p. 10-13.

181. Gerken, H.G., Donohoe, B., and Knoshaug, E.P., Enzymatic Cell Wall Degradation of Chlorellavulgaris and Other Microalgae for Biofuels Production. *Planta*, **2013**. *237*(1): p. 239-253.

182. Takeda, H., Sugar Composition of the Cell Wall and the Taxonomy of Chlorella (Chlorophyceae) 1. *Journal of Phycology*, **1991**. *27*(2): p. 224-232.

183. Kapaun, E. and Reisser, W., A Chitin-Like Glycan in the Cell Wall of a Chlorella Sp.(Chlorococcales, Chlorophyceae). *Planta*, **1995**. *197*(4): p. 577-582.

184. Huss, V.A., Frank, C., Hartmann, E.C., Hirmer, M., Kloboucek, A., Seidel, B.M., Wenzeler, P., and Kessler, E., Biochemical Taxonomy and Molecular Phylogeny of the Genus Chlorella Sensu Lato (Chlorophyta). *Journal of Phycology*, **1999**. *35*(3): p. 587-598.

185. Kapaun, E., Loos, E., and Reisser, W., Cell Wall Composition of Virus-Sensitive Symbiotic Chlorella Species. *Phytochemistry*, **1992**. *31*(9): p. 3103-3104.

186. Meints, R.H., Burbank, D.E., Van Etten, J.L., and Lamport, D.T., Properties of the Chlorella Receptor for the Virus Pbcv-11. *Virology*, **1988**. *164*(1): p. 15-21.

187. Braun, E. and Aach, H., Enzymatic Degradation of the Cell Wall of Chlorella. *Planta*, **1975**. *126*(2): p. 181-185.

188. Honjoh, K.-i., Suga, K., Shinohara, F., Maruyama, I., Miyamoto, T., Hatano, S., and Iio, M., Preparation of Protoplasts from Chlorella Vulgaris K-73122 and Cell Wall Regeneration of Protoplasts from C. Vulgaris K-73122 and C-27. *JOURNAL-FACULTY OF AGRICULTURE KYUSHU UNIVERSITY*, **2003**. *47*(2): p. 257-266.

189. Kaeppler, H.F., Gu, W., Somers, D.A., Rines, H.W., and Cockburn, A.F., Silicon Carbide Fiber-Mediated DNA Delivery into Plant Cells. *Plant Cell Reports*, **1990**. *9*(8): p. 415-418.

190. Dunahay, T., Transformation of Chlamydomonas Reinhardtii with Silicon Carbide Whiskers. *Biotechniques*, **1993**. *15*(3): p. 452-5, 457.

191. Kaeppler, H., Somers, D., Rines, H., and Cockburn, A., Silicon Carbide Fiber-Mediated Stable Transformation of Plant Cells. *Theoretical and applied genetics*, **1992**. *84*(5-6): p. 560-566.

192. Kim, S.-K., *Handbook of Marine Microalgae: Biotechnology Advances*. 2015: Academic Press.

193. Maruyama, M., Horáková, I., Honda, H., Xing, X.-h., Shiragami, N., and Unno, H., Introduction of Foreign DNA into Chlorella Saccharophila by Electroporation. *Biotechnology techniques*, **1994**. *8*(11): p. 821-826.

194. Run, C., Fang, L., Fan, J., Fan, C., Luo, Y., Hu, Z., and Li, Y., Stable Nuclear Transformation of the Industrial Alga Chlorella Pyrenoidosa. *Algal Research*, **2016**. *17*: p. 196-201.

195. Higuchi, R., Song, C., Hoshina, R., and Suzaki, T., Endosymbiosis-Related Changes in Ultrastructure and Chemical Composition of Chlorella Variabilis (Archaeplastida, Chlorophyta) Cell Wall in Paramecium Bursaria (Ciliophora, Oligohymenophorea). *European Journal of Protistology*, **2018**. *66*: p. 149-155.

196. Madani, F., Lindberg, S., Langel, Ü., Futaki, S., and Gräslund, A., Mechanisms of Cellular Uptake of Cell-Penetrating Peptides. *Journal of biophysics*, **2011**. *2011*.

197. Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.-Y., and Liu, D.R., Cationic Lipid-Mediated Delivery of Proteins Enables Efficient Protein-Based Genome Editing in Vitro and in Vivo. *Nature biotechnology*, **2015**. *33*(1): p. 73-80.

198. Schwarze, S.R., Ho, A., Vocero-Akbani, A., and Dowdy, S.F., In Vivo Protein Transduction: Delivery of a Biologically Active Protein into the Mouse. *Science*, **1999**. *285*(5433): p. 1569-1572.

199. Liang, X., Potter, J., Kumar, S., Zou, Y., Quintanilla, R., Sridharan, M., Carte, J., Chen, W., Roark, N., and Ranganathan, S., Rapid and Highly Efficient Mammalian Cell Engineering Via Cas9 Protein Transfection. *Journal of biotechnology*, **2015**. *208*: p. 44-53.

200. Lakshmanan, M., Kodama, Y., Yoshizumi, T., Sudesh, K., and Numata, K., Rapid and Efficient Gene Delivery into Plant Cells Using Designed Peptide Carriers. *Biomacromolecules*, **2013**. *14*(1): p. 10-16.

201. Numata, K., Mieszawska-Czajkowska, A.J., Kvenvold, L.A., and Kaplan, D.L., Silk-Based Nanocomplexes with Tumor-Homing Peptides for Tumor-Specific Gene Delivery. *Macromolecular bioscience*, **2012**. *12*(1): p. 75-82.

202. Leonetti, J.P., Degols, G., and Lebleu, B., Biological Activity of Oligonucleotide-Poly (L-Lysine) Conjugates: Mechanism of Cell Uptake. *Bioconjugate chemistry*, **1990**. *1*(2): p. 149-153.

203. Chen, Q., Zhang, L., Stass, S., and Mixson, A., Co-Polymer of Histidine and Lysine Markedly Enhances Transfection Efficiency of Liposomes. *Gene Therapy*, **2000**. *7*(19): p. 1698-1705.

204. Lakshmanan, M., Yoshizumi, T., Sudesh, K., Kodama, Y., and Numata, K., Double-Stranded DNA Introduction into Intact Plants Using Peptide–DNA Complexes. *Plant Biotechnology*, **2015**: p. 14.1210 b.

205. Pratheesh, P., Vineetha, M., and Kurup, G.M., An Efficient Protocol for the Agrobacterium-Mediated Genetic Transformation of Microalga Chlamydomonas Reinhardtii. *Molecular biotechnology*, **2014**. *56*(6): p. 507-515.

206. Speciale, I., Agarkova, I., Duncan, G.A., Van Etten, J.L., and De Castro, C., Structure of the N-Glycans from the Chlorovirus Ne-Jv-1. *Antonie van Leeuwenhoek*, **2017**. *110*(11): p. 1391-1399.

207. Piacente, F., Bernardi, C., Marin, M., Blanc, G., Abergel, C., and Tonetti, M.G., Characterization of a Udp-N-Acetylglucosamine Biosynthetic Pathway Encoded by the Giant DNA Virus Mimivirus. *Glycobiology*, **2014**. *24*(1): p. 51-61.

208. Huerta-Cepas, J., Serra, F., and Bork, P., Ete 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular biology and evolution*, **2016**. *33*(6): p. 1635-1638.

209. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O., New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of Phyml 3.0. *Systematic biology*, **2010**. *59*(3): p. 307-321.

210. Letunic, I. and Bork, P., Interactive Tree of Life (Itol) V4: Recent Updates and New Developments. *Nucleic acids research*, **2019**. *47*(W1): p. W256-W259.

211. Consortium, U., Uniprot: A Worldwide Hub of Protein Knowledge. *Nucleic acids research*, **2019**. *47*(D1): p. D506-D515.

212. Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V., A Completely Reimplemented Mpi Bioinformatics Toolkit with a New Hhpred Server at Its Core. *Journal of molecular biology*, **2018**. *430*(15): p. 2237-2243.

213. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J., The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nature protocols*, **2015**. *10*(6): p. 845-858.

214. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J., Patchdock and Symmdock: Servers for Rigid and Symmetric Docking. *Nucleic acids research*, **2005**. *33*(suppl_2): p. W363-W367.

215. Agrawal, P., Mishra, G., and Raghava, G.P., Sambinder: A Web Server for Predicting S-Adenosyl-L-Methionine Binding Residues of a Protein from Its Amino Acid Sequence. *Frontiers in pharmacology*, **2020**. *10*: p. 1690.

216. Neuhaus, J.-M., Sticher, L., Meins, F., and Boller, T., A Short C-Terminal Sequence Is Necessary and Sufficient for the Targeting of Chitinases to the Plant Vacuole. *Proceedings of the National Academy of Sciences*, **1991**. *88*(22): p. 10362-10366.

217. Zhang, Y., Xiang, Y., Van Etten, J.L., and Rossmann, M.G., Structure and Function of a Chlorella Virus-Encoded Glycosyltransferase. *Structure*, **2007**. *15*(9): p. 1031-1039.

218. Bock, K. and Pedersen, C., Carbon-13 Nuclear Magnetic Resonance Spectroscopy of Monosaccharides. *Advances in carbohydrate chemistry and biochemistry*, **1983**. *41*: p. 27-66.

219. Akey, D.L., Li, S., Konwerski, J.R., Confer, L.A., Bernard, S.M., Anzai, Y., Kato, F., Sherman, D.H., and Smith, J.L., A New Structural Form in the Sam/Metal-

Dependent O-Methyltransferase Family: Myce from the Mycinamicin Biosynthetic Pathway. *Journal of molecular biology*, **2011**. *413*(2): p. 438-450.

220. Liscombe, D.K., Louie, G.V., and Noel, J.P., Architectures, Mechanisms and Molecular Evolution of Natural Product Methyltransferases. *Natural product reports*, **2012**. *29*(10): p. 1238-1250.

221. García, I.G., Stevenson, C.E., Usón, I., Meyers, C.L.F., Walsh, C.T., and Lawson, D.M., The Crystal Structure of the Novobiocin Biosynthetic Enzyme Novp: The First Representative Structure for the Tylf O-Methyltransferase Superfamily. *Journal of molecular biology*, **2010**. *395*(2): p. 390-407.

222. Rigg, G.P. and Barrett, B., The Localization of Kpsc, S and T, and Kfia, C and D Proteins Involved in the Biosynthesis of the Escherichia Coli K5 Capsular Polysaccharide: Evidence for a Membrane-Bound Complex. *Microbiology*, **1998**. *144*(10): p. 2905-2914.

223. Clarke, B.R., Ovchinnikova, O.G., Sweeney, R.P., Kamski-Hennekam, E.R., Gitalis, R., Mallette, E., Kelly, S.D., Lowary, T.L., Kimber, M.S., and Whitfield, C., A Bifunctional O-Antigen Polymerase Structure Reveals a New Glycosyltransferase Family. *Nature chemical biology*, **2020**. *16*(4): p. 450-457.

224. Greenfield, L.K., Richards, M.R., Li, J., Wakarchuk, W.W., Lowary, T.L., and Whitfield, C., Biosynthesis of the Polymannose Lipopolysaccharide O-Antigens from Escherichia Coli Serotypes O8 and O9a Requires a Unique Combination of Single-and Multiple-Active Site Mannosyltransferases. *Journal of Biological Chemistry*, **2012**. *287*(42): p. 35078-35091.

225. Greenfield, L.K., Richards, M.R., Vinogradov, E., Wakarchuk, W.W., Lowary, T.L., and Whitfield, C., Domain Organization of the Polymerizing Mannosyltransferases Involved in Synthesis of the Escherichia Coli O8 and O9a Lipopolysaccharide O-Antigens. *Journal of Biological Chemistry*, **2012**. *287*(45): p. 38135-38149.

226. Williams, D.M., Ovchinnikova, O.G., Koizumi, A., Mainprize, I.L., Kimber, M.S., Lowary, T.L., and Whitfield, C., Single Polysaccharide Assembly Protein That Integrates Polymerization, Termination, and Chain-Length Quality Control. *Proceedings of the National Academy of Sciences*, **2017**. *114*(7): p. E1215-E1223.

227. Jing, W. and DeAngelis, P.L., Dissection of the Two Transferase Activities of the Pasteurella Multocida Hyaluronan Synthase: Two Active Sites Exist in One Polypeptide. *Glycobiology*, **2000**. *10*(9): p. 883-889.

228. Sobhany, M., Kakuta, Y., Sugiura, N., Kimata, K., and Negishi, M., The Chondroitin Polymerase K4cp and the Molecular Mechanism of Selective Bindings of Donor Substrates to Two Active Sites. *Journal of Biological Chemistry*, **2008**. *283*(47): p. 32328-32333.

229. Chavaroche, A.A., van den Broek, L.A., Boeriu, C., and Eggink, G., Synthesis of Heparosan Oligosaccharides by Pasteurella Multocida Pmhs2 Single-Action Transferases. *Applied microbiology and biotechnology*, **2012**. *95*(5): p. 1199-1210.

230. Zhang, H., Zhou, M., Yang, T., Haslam, S.M., Dell, A., and Wu, H., New Helical Binding Domain Mediates a Glycosyltransferase Activity of a Bifunctional Protein. *Journal of Biological Chemistry*, **2016**. *291*(42): p. 22106-22117.

231. Abdel-Mawgoud, A.M., Lépine, F., and Déziel, E., Rhamnolipids: Diversity of Structures, Microbial Origins and Roles. *Applied microbiology and biotechnology*, **2010**. *86*(5): p. 1323-1336.

232. Stipcevic, T., Piljac, A., and Piljac, G., Enhanced Healing of Full-Thickness Burn Wounds Using Di-Rhamnolipid. *Burns*, **2006**. *32*(1): p. 24-34.

233. Landstein, D., Burbank, D.E., Nietfeldt, J.W., and Van Etten, J.L., Large Deletions in Antigenic Variants of the Chlorella Virus Pbcv-11. *Virology*, **1995**. *214*(2): p. 413-420.

234. Noel, E., Notaro, A., Speciale, I., Duncan, G.A., De Castro, C., and Van Etten, J.L., Chlorovirus Pbcv-1 Multidomain Protein A111/114r Has Three Glycosyltransferase Functions Involved in the Synthesis of Atypical N-Glycans. *Viruses*, **2021**. *13*(1): p. 87.

235. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E., The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nature Protocols*, **2015**. *10*(6): p. 845-858.

236. Söding, J., Biegert, A., and Lupas, A.N., The Hhpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic acids research*, **2005**. *33*(suppl_2): p. W244-W248.

237. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J., Gapped Blast and Psi-Blast: A New Generation of Protein Database Search Programs. *Nucleic acids research*, **1997**. *25*(17): p. 3389-3402.

238. Culbertson, A.T., Ehrlich, J.J., Choe, J.-Y., Honzatko, R.B., and Zabotina, O.A., Structure of Xyloglucan Xylosyltransferase 1 Reveals Simple Steric Rules That Define Biological Patterns of Xyloglucan Polymers. *Proceedings of the National Academy of Sciences*, **2018**. *115*(23): p. 6064-6069.

239. Haltiwanger, R.S., Yu, H., Takeuchi, M., LeBarron, J., Kantharia, J., London, E., Bakker, H., Li, H., and Takeuchi, H., Regulation of Notch Signaling by O-Glucosylation: Notch-Modifying Xylosyltransferase-Substrate Complexes Support an Sni-Like Retaining Mechanism. *The FASEB Journal*, **2016**. *30*(1_supplement): p. 624.3-624.3.

240. Gibbons, B.J., Roach, P.J., and Hurley, T.D., Crystal Structure of the Autocatalytic Initiator of Glycogen Biosynthesis, Glycogenin. *Journal of molecular biology*, **2002**. *319*(2): p. 463-477.

241. Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C., Crystal Structure of the Retaining Galactosyltransferase Lgtc from Neisseria Meningitidis in Complex with Donor and Acceptor Sugar Analogs. *Nature structural biology*, **2001**. *8*(2): p. 166-175.

242. Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B., A Classification of Nucleotide-Diphospho-Sugar Glycosyltransferases Based on Amino Acid Sequence Similarities. *Biochemical Journal*, **1997**. *326*(3): p. 929-939.

243. Jiang, Y.-L., Jin, H., Yang, H.-B., Zhao, R.-L., Wang, S., Chen, Y., and Zhou, C.-Z., Defining the Enzymatic Pathway for Polymorphic O-Glycosylation of the Pneumococcal Serine-Rich Repeat Protein Psrp. *Journal of Biological Chemistry*, **2017**. *292*(15): p. 6213-6224.

244. Clarke, B.R., Ovchinnikova, O.G., Sweeney, R.P., Kamski-Hennekam, E.R., Gitalis, R., Mallette, E., Kelly, S.D., Lowary, T.L., Kimber, M.S., and Whitfield, C., A Bifunctional O-Antigen Polymerase Structure Reveals a New Glycosyltransferase Family. *Nature Chemical Biology*, **2020**: p. 1-8.

245. Ninomiya, T., Sugiura, N., Tawada, A., Sugimoto, K., Watanabe, H., and Kimata, K., Molecular Cloning and Characterization of Chondroitin Polymerase from Escherichia Coli Strain K4. *Journal of Biological Chemistry*, **2002**. *277*(24): p. 21567-21575.

246. Sun, H.-Y., Lin, S.-W., Ko, T.-P., Pan, J.-F., Liu, C.-L., Lin, C.-N., Wang, A.H.-J., and Lin, C.-H., Structure and Mechanism of Helicobacter Pylori Fucosyltransferase a Basis for Lipopolysaccharide Variation and Inhibitor Design. *Journal of Biological Chemistry*, **2007**. *282*(13): p. 9973-9982.

247. Speciale, I., Laugieri, M.E., Noel, E., Lin, S., Lowary, T.L., Molinaro, A., Duncan, G.A., Agarkova, I.V., Garozzo, D., and Tonetti, M.G., Chlorovirus Pbcv-1 Protein A064r Has Three of the Transferase Activities Necessary to Synthesize Its Capsid Protein N-Linked Glycans. *Proceedings of the National Academy of Sciences*, **2020**.