IDENTIFICATION AND BINDING CHARACTERISATION OF PEPTIDE-BEARING PHAGES TOWARDS RECOMBINANT PROTEASE (NS6) OF MURINE NOROVIRUS

NUR SAKINAH BINTI SOID

UNIVERSITI SAINS MALAYSIA

2019

IDENTIFICATION AND BINDING CHARACTERISATION OF PEPTIDE-BEARING PHAGES TOWARDS RECOMBINANT PROTEASE (NS6) OF MURINE NOROVIRUS

by

NUR SAKINAH BINTI SOID

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

March 2019

I dedicate this thesis to those who are supporting me mainly

my beloved father and mother

Soid bin Mansor

and

Shahrozat binti Abdul

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful.

Acknowledging the dedicated efforts of those who gave me support and made my postgraduate studies a great experience, is a great pleasure for me.

Alhamdulillah, all praises and grateful to Allah, God the Almighty for giving me strength, patience, inspiration and opportunity to complete my Msc project. A lot of experiences and knowledge were gained along the way.

Throughout the most exciting in learning the practical aspect of this course, I would like to thank those who support me along the way this project being done. My greatest gratitude goes to my family especially my parents, husband and siblings for their moral support. To my family, I love all of you.

I also would like to express my sincere appreciation to my supervisor, Dr Muhammad Amir bin Yunus for his knowledge, guidance and encouragement towards the completion of my thesis. Special thanks to my co-supervisor Dr Ida Shazrina and Dr Kumitaa for their endless supervision throughout my study. In addition, great thanks for my lab mates Nalini, Nurul, Nithya, Khirun, Laina, Adam, Azali and Din who are together with me in the hardness and happiness throughout this project. Not to be forgotten, to all IPPT staffs especially laboratory staffs, En Zul and Pn Hikmah whom their assistant, direction and knowledge have directly contributed towards completing my thesis project.

Apart from that, I dedicate my appreciation to my beloved housemate (Fatin, Aida and Athirah) for always being with me during ups and downs and a very special thanks to all my postgraduate friends, for their friendship and useful discussions over the last three years. Thanks a lot from the deepest of my heart for those who have contributed directly or indirectly for my project.

Last but not least, I would like to acknowledge Yayasan Khazanah who has supported me financially throughout my study year.

May Allah s.w.t bless all of you. Wassalam.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ΔC139A	NS6 active site mutant
BCA	Bicinchoninic Acid
BSA	Bovine serum albumin
CDC	Centre for Disease Control & Prevention
cDNA	complementary Deoxyribonucleic Acid
CIP	Calf alkaline phosphatase
DNA	Deoxyribonucleic acid
ECL	Enhanced Chemiluminescent
EDTA	Ehtylenediaminetetraacetic Acid
eIFs	Eukaryotic initiation factors
ELISA	Enzyme-linked Immnosorbent Assay
EtBr	Ethidium Bromide
FCV	Feline Calicivirus
FUT 2	α-(1,2)-fucosyltransferase
gRNA	genomic RNA
HBGAs	histo-blood group antigens
HCV	Hepatitis C virus
HPQ	Histidine-Proline-Glutamine
HRP	Horseradish Peroxidase
HuCv	Human calicivirus
HuNv	Human norovirus
IPTG	Isopropyl β-D-1-Thiogalactopyranoside

LB	Luria Bertani
MNV	Murine norovirus
NCBI	National Center for Biotechnology Information
NV	Norovirus
OD	Optical density
ORF	Open reading frame
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PEC	Porcine enteric calicivirus
PMO	phosphorodiamidate morpholino oligomer
PPMO	peptide conjugated phosphorodiamidate morpholino oligomers
PVDF	polyvinylidene fluoride
RAG 2	recombination activation gene 2
RAW264.7	murine macrophage cell line
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic acid
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sgRNA	subgenomic RNA
STAT-1	signal transducer and activation of transcription 1
TBS	tris buffered saline
TBST	tris buffered saline containing Tween 20
TMB	3,3’5,5’-Tetramethylbenzidine
UTR	Untranslated region

VF1	Virulence factor 1
VLP	Viral-like particle
VPg	Viral protein linked genome
WT	Wild type

LIST OF SYMBOLS

%	Percentage
>	more than
° C	degree Celsius
µl	Microliter
µM	Micromolar
h	Hour
mg	Milligram
mins	Minutes
mM	Millimolar
nm	Nanometer
nt	Nucleotides
pfu	plaque forming unit
secs	Seconds
v/v	volume per volume
w/v	weight per volume

IDENTIFIKASI DAN PENCIRIAN PENGIKATAN FAJ PEPTIDA KE ATAS REKOMBINAN PROTEASE (NS6) MURIN NOROVIRUS

ABSTRAK

Jangkitan norovirus dianggap sebagai penyebab utama kes penyakit gastroenteritis akut yang berlaku secara epidemik dan berkala di seluruh dunia. Sehingga kini, tiada langkah intervensi terapeutik berlesen sama ada dari segi vaksin atau ubat-ubatan yang tersedia untuk mengawal penularan patogen manusia ini. Menyasarkan protein bukan struktural virus seperti protease adalah relevan kerana bahagian yang terjaga di samping fungsinya yang penting dalam sistem replikasi virus. Dalam kajian ini, versi mutasi MNV NS6 yang aktif (ΔC139A NS6) telah diklon, diekspres dan ditulenkan sebelum digunakan dalam pendulangan bio / pemilihan yang menggunakan kit Ph.D-7^{TM} Phage Display Peptide Library (NEB). Selepas 6 pusingan pendulangan bio, beberapa peptida faj yang spesifik dengan ΔC139A NS6 telah berjaya dikenal pasti dan diasingkan. Peptida faj yang telah dikenalpasti kemudiannya telah disaring menggunakan pangkalan data peptida dalam talian dan hasilnya menunjukkan bahawa semua motif adalah unik dan tidak mempunyai persamaan dengan mana-mana peptida yang pernah diterbitkan. Spesifikasi ikatan beberapa klon peptida faj dengan jumlah pengulangan tertinggi telah diuji menggunakan ELISA. Keputusan ELISA menunjukkan klon yang membawa urutan peptida QTEKNPL mempunyai keupayaan mengikat yang tertinggi terhadap ΔC139A NS6 berbanding klon lain dan juga klon kawalan. Dalam analisa silico, struktur tersier ΔC139A NS6 dan interaksi peptida terpilih (ADARYKS, QTEKNPL dan NSKLVLG) terhadap reseptor ΔC139A NS6 telah diramal masing-masing menggunakan perisian I-TASSER dan CABS-dock. Analisis dok menunjukkan bahawa semua urutan peptida

yang terpilih berinteraksi dengan reseptor ΔC139A NS6 pada beberapa interaksi putatif. Menariknya, semua peptida yang dipilih mengikat reseptor ΔC139A NS6 di tapak yang tiada dalam laporan penyelidikan sebelum ini. Analisis bioinformatik ini memberikan beberapa maklumat penting untuk penyelidikan pada masa hadapan berkaitan interaksi reseptor ligan yang melibatkan peptida terpilih dan protein ΔC139A NS6. Maklumat ini berpotensi untuk dimanipulasi dalam pembangunan perencat peptida kecil terhadap protease NS6 serta untuk kegunaan diagnostik dengan tujuan utama bagi mengawal replikasi norovirus.

IDENTIFICATION AND BINDING CHARACTERISATION OF PEPTIDE-BEARING PHAGES TOWARDS RECOMBINANT PROTEASE (NS6) OF MURINE NOROVIRUS

ABSTRACT

Norovirus infections are considered the most common causes of epidemic and sporadic cases of acute gastroenteritis worldwide. To date, there are no licensed therapeutic intervention measures either in terms of vaccines or drugs available for these highly contagious human pathogen. Targeting non-structural viral protein such as protease is relevant due to their well conserved regions along with its critical functions in viral replication. In this study, the active site mutant version of MNV NS6 protease (ΔC139A NS6) was cloned, overexpressed and purified before being used as target in biopanning/selection employing the Ph. D-7^{TM} Phage Display Peptide Library kit (NEB) which was validated in streptavidin panning. After six rounds of biopanning, several peptide phages that are specific towards ΔC139A NS6 were isolated and identified successfully. Upon identification of peptide phage, the peptide sequences were screened using online peptide databases and results indicated that all the sequences were unique and did not possess common motif with any published peptides to date. Three peptide phage clones with highest number of hits were tested for their binding specificity towards purified C139A NS6 using ELISA. It was observed that clones carrying peptide sequence QTEKNPL had the highest binding affinity towards ΔC139A NS6 as compared to other clones and irrelevant phage control. Tertiary structure of ΔC139A NS6 and binding interaction of the selected peptides (ADARYKS, QTEKNPL and NSKLVLG) with ΔC139A NS6 receptor were predicted using I-TASSER and CABS-dock software, respectively. Docking analysis

showed that all the selected peptide sequences interact with ΔC139A NS6 receptor at several contacts on the receptor binding sites that had never been published. This bioinformatics analysis uncovers some important information for future investigation in ligand-receptor interaction involving selected peptides and ΔC139A NS6 protein which could potentially be manipulated in the development of small peptide inhibitor against NS6 protease as well as for the diagnostic purposes with ultimate purpose to control norovirus replication.

CHAPTER 1

INTRODUCTION

1.1 Research overview

Virus is a small, infectious particle that can reproduce only by infecting a living cell and using its machinery to produce more viruses. They comprise of nucleic acid that can be either DNA or RNA which are enclosed in a protein shell called a capsid. Some viruses have an internal or external layer of membrane known as envelope. Virus are very diverse since they vary in their shapes, structures, kinds of genomes and type of hosts that they infect. Generally, viruses work by invading the host cells to make more copies of themselves and throughout the process, they often cause destruction to the invaded cell (Koonin, Senkevich, & Dolja, 2006).

Noroviruses belong to the family of *Caliciviridae* are positive-sense, single-stranded RNA virus. Due to their stability, low infectious dose, large host reservoir, short term immunity, multiple transmission routes and large diversity between strains, human norovirus particularly has been established as one of the leading causes that account for gastroenteritis cases worldwide with several clinical manifestations of acute diarrheal and vomiting, accompanied by several signs/symptoms like abdominal cramps, myalgia, malaise, headache, nausea and low-grade fever. This virus is thought to have a considerable economic impact due to the number of cases reported annually in developed countries.

Despite complexity in the progress of the vaccines due to the high degree of genetic variation of this virus, there are few vaccines for human calicivirus that are under pre-clinical and clinical trial (Lucero, Vidal, & O'Ryan, 2017; Tan & Jiang, 2014). Furthermore, immunity to human calicivirus is also short-lived since body are

not able to maintain immunity for a long time once a person is infected with natural infection (Mattison, 2011). For these reasons, antiviral therapy that is a class of treatment used specifically for treating viral infection seems to be an attractive strategy to allow the control of calicivirus outbreaks. An antiviral function by inhibiting the development of virus specifically by targeting the critical function of viral proteins, or parts of proteins which ultimately will affect the viral life cycle.

Norovirus NS6 protease is an important enzyme that enables the activation of its other non-structural proteins. This protease acts as a key point for the activation of norovirus replication. The importance of the enzymatic activity of the NS6 protease through its main function in cleaving the chain of non-structural large polyproteins into several active proteins has made it as an attractive target for developing antiviral therapies. This could be achieved by inhibiting the function of this protease which may result in the inhibition of the cleavage process of non-structural proteins which subsequently disrupt the replication of the virus in infected host.

Even though norovirus causes a huge lost to the western country and not widely spread in Malaysia, but due to current setting where immigration and emigration take place every single day, it is fair to take the possibilities of transmission of this virus into consideration. In this study, we aim to identify a peptide-bearing phages with the ability to bind the target (NS6) with high affinity and specificity using the phage display technique. After the identification of such peptide sequences, several commercial synthetic peptides could be synthesized and use to assess their inhibitory effects and efficiency to act as antiviral therapy. The benefit of developing peptide-based antiviral drugs is that the peptides can be designed to mimic or interact with conserved surface proteins and in case of strains variety, the peptide sequence could be modified to preserve therapeutic efficiency. As compared to other approach, the

strength of peptide-based approach is in their good efficacy, safe, selectivity and predictable biocompatibility and biodegradability and are easy to being scaled-up.

1.2 Objectives of Study

1.2.1 General Objective

This study was conducted to identify the peptide-bearing phages and characterise their binding ability towards recombinant protease (NS6) of murine norovirus (MNV).

1.2.2 Specific Objectives

1. To clone, express and purify mutant NS6 protease.
2. To select peptide-bearing phage that binds to the recombinant mutant NS6 protease.
3. To determine the ability of the selected peptide-bearing phage to bind to the recombinant mutant NS6 protease.
4. To predict the tertiary structure of mutant NS6 protease and docking analysis of the selected peptide sequences on mutant NS6 protease receptor.

CHAPTER 2

LITERATURE REVIEW

2.1 Gastroenteritis

Gastroenteritis is a condition of an inflammation of gastrointestinal tract that includes stomach and intestinal mucosal surface (Schlossberg, 2015). It is also known as infectious diarrhea whereby it is associated with symptoms such as diarrhea, vomiting, malaise, muscle pain, fatigue, dehydration and high-grade fever (Glass, Parashar, & Estes, 2009; Karst, 2010). This condition can be caused by infectious agents such as viruses (rotavirus and norovirus), bacteria (*Escherichia coli* and *Campylobacter* species), parasites (*Giardia lamblia)*, fungus as well as non-infectious sources like medications and consumption of certain foods such as lactose and gluten (Ciccarelli, Stolfi, & Caramia, 2013; Helms & Quan, 2006; Szajewska & Dziechciarz, 2010). However, the most common cause of gastroenteritis is due to infection by viruses; for example norovirus which causes 685 million cases and 213,515 death globally in developed and developing country (Pires et al., 2015).

Virus that are commonly known to cause viral gastroenteritis are rotavirus, norovirus, adenovirus and astrovirus (Oude Munnink & van der Hoek, 2016). Rotavirus is the most known causes of gastroenteritis in children while norovirus has been identified as the major cause of epidemic gastroenteritis and common cause of diarrhea in adults (Glass et al., 2009; Hall et al., 2013). Initially in 1972, virus that was described as calicivirus was identified in the stool of patient with diarrhea during an outbreak of gastroenteritis in a school in Norwalk, Ohio (Kapikian, 2000; Kapikian et al., 1972). Currently, diarrheal disease is the top five causes of death worldwide with children are the most affected as compared to adult due to their lack of immunity development and relatively poor hygiene. In childhood gastroenteritis, norovirus have

been reported to be the second cause following rotavirus, but to date, the widespread use of vaccine against rotavirus has made norovirus to be more likely as the most common cause of gastroenteritis in childhood (Baehner, Bogaerts, & Goodwin, 2016; Karst, 2010)

2.2 Caliciviruses

The Caliciviridae family consists of small RNA viruses that are important in both medical and veterinary area (Royall & Locker, 2016). It belongs to Class IV of the Baltimore scheme where positive single-stranded RNA viruses whose mRNA is identical in base sequence to virion RNA can be accessed by host ribosomes to directly form proteins. Viruses with this type of RNA genome can be divided into two groups where both reproduce in the cytoplasm. The first group is polycistronic mRNA type viruses where mRNA from the genomic RNA is translated into a polyprotein. This product will then subsequently cleave to form single mature proteins. In contrast, the second group is where viruses with complex transcription use different mechanisms to produce proteins from the same strand of RNA. It can be either using subgenomic mRNAs, ribosomal frameshifting and proteolytic processing of polyprotein (Baltimore, 1971).

The name calicivirus derived from the Latin word *calyx* which means cup or chalice; and being referred to 32 cup-shaped depressions on the surface of the entire virus particle arranged in icosahedral symmetry (Green et al., 2000; Prasad, Matson, & Smith, 1994). They are typically single-stranded and non-segmented, positive-sense RNA genome with size ranging from 6.7–8.5 kb (Black, Burroughs, Harris, & Brown, 1978; Jiang, Wang, Wang, & Estes, 1993). The RNA genome of calicivirus were organised into two or three open reading frame (ORFs) flanked by two short

untranslated regions (Sosnovtsev et al., 2006). The genome structure in this family differs in the length of each ORF but they are all conserved. Major feature of Calicivirus is the absence of a methylated cap at the 5' terminus of the virion RNA. Instead, a small protein (VPg) with size around 15 kDa was shown to be covalently linked to both 5' end of the calicivirus genomic RNA (gRNA) and subgenomic RNA (sgRNA), with polyadenylated at the 3' end. This VPg protein drives the initiation of translation and define to be essential for the infectivity of the RNA (Black et al., 1978; Burroughs & Brown, 1978). In addition, 3' co-terminal sgRNA is also being transcribed during infection in all members of the Caliciviridae family (Royall & Locker, 2016).

Caliciviruses infection are major cause of disease in humans and many animals as it can naturally infect organism such as human, cattle, cats and pigs (Hansman, Jiang, & Green, 2010). Previously, this virus is not very well studied since there is no suitable animal model and lack of efficient cell culture system. However, advancement in modern genomic technologies has led to an increased in understandings of the virus family.

2.2.1 Classification and Taxonomy

Caliciviruses are classified into five established genera that are *Norovirus*, *Lagovirus*, *Vesivirus, Sapovirus* and *Nebovirus* together with numbers of additional genera that have been proposed (Kitchen, Shackelton, & Holmes, 2011). They are *Recovirus* (eg, Tulane virus) for a novel calicivirus detected in stool specimens from rhesus macaques, *Valovirus* that represent a novel group of swine caliciviruses known as the St-Valérien-like viruses, *Bavovirus* that are found in chicken, *Nacovirus* and *Bacovirus* that contain novel caliciviruses recovered from chickens and turkeys respectively, *Minovirus* that infect farmed fathead minnows (*Pimephales*

promelas), *Salovirus* that was identified in farmed Atlantic salmon (*Salmo salar*), *Sanovirus* was found in goose and also *Secalivirus* that was acquired from a sewage sample which consist of partial genome of a highly divergent calicivirus (Farkas, Sestak, Wei, & Jiang, 2008; L'Homme et al., 2009; Mikalsen et al., 2014; Mor et al., 2017; Ng et al., 2012; Wang, Wang, Dong, Zhang, & Zhang, 2017; Wolf et al., 2012).

Members of caliciviruses that causes acute gastroenteritis infections in human are norovirus and sapovirus while other members like lagovirus, vesivirus and becovirus/nebovirus are not pathogenic to humans (Bailey & Goodfellow, 2009). Initially, sapovirus was discovered in an outbreak of gastroenteritis in an orphanage in Sapporo, Japan (Chiba et al., 1979). This member of calicivirus is also known as "Sapporo-like viruses" includes Manchester Virus and porcine enteric calicivirus (PEC). The virus is transmitted through oral/fecal contact and most common symptoms are diarrhea and vomiting. The *Vesivirus* genus is circulating in swine, sea mammals, and felines as their natural hosts. This genus includes feline calicivirus and vesicular exanthema of swine virus; as well as San Miguel sea lion viruses that are found in marine and terrestrial mammals (Neill, Meyer, & Seal, 1995). Two main groups of Lagovirus are rabbit hemorrhagic disease virus that are highly fatal and European brown hare syndrome virus that was observed in the livers of infected hares in 1982 (Green et al., 2000). The genus *Nebovirus* contains two virus species that infect cattle are Newbury-1 virus and bovine enteric calicivirus NB (Nebraska). These viruses cause diarrhea and intestinal disease in calves (Oliver, Asobayire, Dastjerdi, & Bridger, 2006).

2.2.2 Morphology and Genome Organisation of Caliciviruses

In 1972, Albert Kapikian discovered a virion with 27 nm particle using immune-electron microscope (Kapikian, 2000). The icosahedral structure of virus is described

as the first calicivirus virion and was popularly known as Norwalk virus. For *Caliciviridae* family, the viral genome acts as an mRNA template that can be readily translated, following entry into the infected cell. Members of these family have structure of gRNA that are conserved, however they differ in the length of each open reading frame (ORF). The *Lagovirus, Sapovirus* and *Nebovirus* genera have two ORFs while three distinct ORFs were observed for the *Vesivirus* and *Norovirus* genera (Rohayem et al., 2010). Meanwhile, murine noroviruses was found to carry a fourth ORF (McFadden et al., 2011)

There are differences in term of the expression of structural proteins that are the major capsid protein (VP1) and minor capsid protein (VP2) among the genera. Within the *Vesivirus* and *Norovirus* genera, VP1 and VP2 are encoded by ORF2 and ORF3 from the sgRNA while in the *Sapovirus, Lagovirus,* and *Nebovirus* genera, both structural proteins are encoded by ORF1 and ORF2. The position of the VP1 coding region of the later genera is at the 3' end of ORF1. The VP1 termination sequence are typically overlap with VP2 start codons and they can be separated by a short stretch around 3–10 nucleotides. Thus, this condition allows the translation of VP2 to take place through reinitiation process even though they are originated from single strand of mRNA (Royall & Locker, 2016)

2.3 Norovirus

History of norovirus started in 1929, where common childhood illness that causes vomiting, diarrhea and fever was detected and named as "winter vomiting disease" by Dr John Zahorsky who is a paediatrician (Zahorsky, 1929). In 1968, a virus is suspected when students and teachers of an elementary school in Norwalk, Ohio, experiences an outbreak of winter vomiting disease with symptoms of acute gastroenteritis, including vomiting and diarrhea (Adler & Zickl, 1969). Several

experiments that was conducted to identify an etiology for this infectious form of acute gastroenteritis failed, until in 1972 where Dr. Albert Kapikian and his group successfully described the *Norovirus* prototype Norwalk virus found in stool samples using immune electron microscopy (Kapikian et al., 1972). Following the visualisation of the Norwalk virus particle, its genome was then cloned in 1990; led to the development of molecular tools to study noroviruses in more detail (Xi, Graham, Wang, & Estes, 1990)

Noroviruses infection has been identified in various host that includes humans, swine, cattle, sheep, mice, cats and dogs (de Graaf et al., 2017). Human caliciviruses can cause infection in a population even at low infectious dose (Teunis et al., 2008). It is also known as winter vomiting disease as the infection/epidemics illustrate that it occurred typically during cold weather (Robilotti, Deresinski, & Pinsky, 2015). Norovirus outbreak mostly occur in semi-closed setting/environment communities that favour person-to-person transmission such as cruise ships, nursing homes, day care centres, schools, restaurants, hospitals, disaster relief/evacuation site, army barracks and even airplanes (Glass et al., 2009; Monroe, Ando, & Glass, 2000; Patel, Hall, Vinjé, & Parashar, 2009; Robilotti et al., 2015; Widdowson, Monroe, & Glass, 2005). The uncontrollable outbreak of norovirus infections is most likely due to high infectivity of norovirus particles, persistence of the virus in the environment, prolonged shedding of virus from both symptomatic and asymptomatic individuals and also short-term immunity (Kaufman, Green, & Korba, 2014). Norovirus is one of the genus under the *Caliciviridae* family that is considered as the most common cause of foodborne disease outbreak by Centers for Disesase Control and Prevention (CDC). It has been estimated that human pathogen within this genus cause at least 95% of nonbacterial gastroenteritis outbreaks and over 50% of all outbreaks worldwide (Karst,

2010). Thus, human norovirus (HuNv) is considered as a leading cause of gastroenteritis worldwide contributed to a severe economic burden to the health organisations of both the developed and developing world. This genus is highly transmissible primarily through exposure to contaminated food or water sources, person-to-person contact, an increased density of people, aerosolized vomitus particles and fomites (Karst, 2010; Lopman et al., 2012). They have been classified under Category B biodefense agents due to their characteristics that are highly contagious, extremely stable in the environment, resistant to common disinfectants and associated with debilitating illness. These conditions provide fundamental towards a fast spreading and highly contagious disease (Karst, 2010).

It is reported that person of all age groups is vulnerable towards norovirus infection and secondary infection are also common (Estes, Prasad, & Atmar, 2006). Norovirus replicate in gastrointestinal tracts of their host that result in acute or mild gastroenteritis characterised by symptoms such as vomiting and diarrhea coupled with or without nausea and abdominal cramp. Other symptoms include low-grade fever and malaise. There is also condition where the infected person is asymptomatic (Karst, 2010). On the contrary, infection that occur in an immunocompromized mice (lacking components of the innate system) indicates that a broad tissue tropism was involved as MNV causes encephalitis, vasculitis, pneumonia and hepatitis (Karst, Wobus, Lay, Davidson, & Virgin, 2003).

Replication of norovirus RNA genome is dependent on virus-encoded RNA-dependent RNA polymerase, which has no proofreading activity that can correct the misincorporation of nucleotides during RNA synthesis. This condition results in high error-rate in genomic replication; causes the production of many progeny viruses with mutations. Therefore, norovirus is highly flexible and diversify as compared to other

RNA viruses (Estes et al., 2006; White, 2014). The *Norovirus* genus is classified based on phylogenetic clustering of the complete major capsid protein (VP1) amino acid sequence where members within a genogroup differ in their capsid genes by 45 -61%, while there are 14-44% and 0-14% different of members within genotype and strains within genotype respectively. These differences in term of intra-genus variation is high even compared to genera of other positive strand RNA virus families. High degree in genetic variation causes difficulty in developing protection against norovirus (Barclay et al., 2014; White, 2014; Zheng et al., 2006).

Currently there are 7 genogroups for norovirus and each of them is further characterised into few genotypes (Robilotti et al., 2015; Zheng et al., 2006). The norovirus strains that infect humans and causes acute gastroenteritis are found in 32 genotypes that are GI (n=9), GII (n=19) and GIV (n=1), whereas viruses in other genogroups infect other mammals. Noroviruses have also been isolated from other species including pigs (GII), bovine (GIII) and murine (GV) (Vinje, 2015). Recently proposed genogroups were GVI and GVII where norovirus was identified in domestic dogs with diarrhea (Mesquita, Barclay, Nascimento, & Vinjé, 2010). Among all the genogroups of norovirus, the GII.4 causes >80% of all human norovirus infections at any one time (Siebenga et al., 2009; White, 2014). This genotype virus is responsible for all six major norovirus pandemics of acute gastroenteritis in the last two decades (White, 2014).

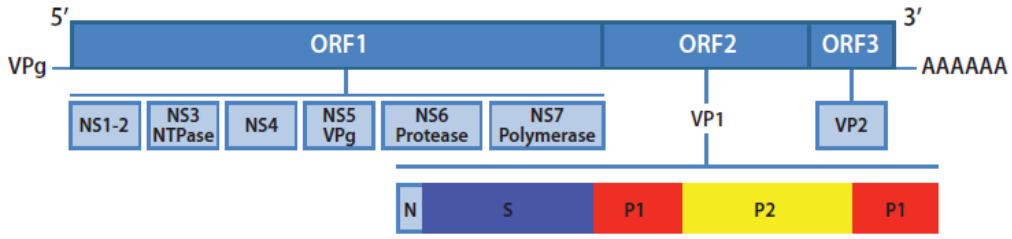
2.3.1 Norovirus Genome Structure

Norovirus virion is non-enveloped with icosahedral symmetry of proteinaceous capsid that packages the RNA genome. The capsid is made up of 180 copies of the ORF2 encoded protein, grouped as dimers in T=3 symmetry (Prasad, Hardy, Jiang, & Estes, 1996; Prasad et al., 1999). The linear single stranded positive sense RNA genome

with approximately 7.4–7.7 kb in size are typically organized into three open reading frames (ORF) (Figure 2.1A) (Jiang et al., 1993; Robilotti et al., 2015; Rocha-Pereira & Nascimento, 2012; Thorne & Goodfellow, 2014). Murine noroviruses possess a fourth ORF that is overlapping with ORF2 but in an alternate reading frame; and only some sapovirus members display an alternative open reading frame at the equivalent position (Clarke & Lambden, 2000; Thackray et al., 2007). All six of the nonstructural proteins of the norovirus genome are encoded in a single ORF (ORF1) at the 5' end region while the 3' end region is transcribed into a subgenomic information. The subgenomic region encodes the major and minor structural proteins in two separate ORFs that are ORF2 and ORF3 respectively (Karst, 2010).

A study conducted by Prasad et al (1999) using Norwalk VLP revealed that major capsid protein (VP1) which determines the antigenicity of the virus are divided into three parts that are N-terminal, shell domain (S) located at the base of the capsid and a protruding (P) domain that is further subdivided into the P1 and P2 regions (Figure 2.1A and 2.1B). The P domain has ability to change shape that can result in its mutation due to its condition that is linked to the S domain by a flexible hinge region. The S domain provides internal core of the capsid while P domain probably contain both immune and cellular recognition sites. The interaction of inner S domain subunits with their neighbouring S domains form a continuous internal shell structure to the capsid. Meanwhile, the cup-like structure was formed when P domain emerge from the S domain surface (Hardy, 2005; Prasad et al., 1999). Mutational analysis on VLP results described that even though shell domain consist of components that are sufficient to initiate the assembly of capsid, the entire P domain is also important to enhance the stability and regulate the size of the capsid structure (Bertolotti-Ciarlet, White, Chen, Prasad, & Estes, 2002).

A)



B)

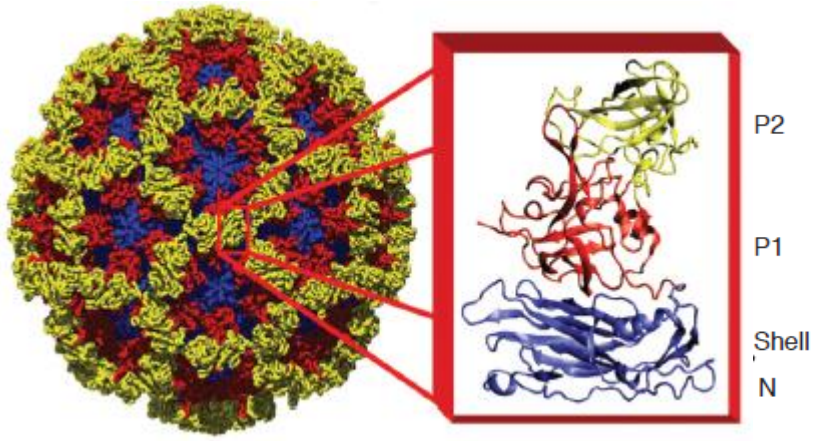


Figure 2.1 A) Schematic representation of genome organization of Norovirus. The non-structural protein is encoded in ORF1 while ORF2 and ORF3 encode for major and minor capsid protein respectively. There is presence of ORF4 that overlap with ORF2; only in murine norovirus. The VP1 (major capsid protein) is further characterised into domain. The small N-terminal domain (light blue) faces the interior of the particle. The shell domain (S) is colored blue. The P1 subdomains are colored red. The P2 domain (yellow) is an insertion in the P2 domain. The hypervariable region of VP1 is found in the P2 domain. B) The schematic structure of capsid protein from electron microscope based on description for ribbon in (A). The figure was adapted from (Lopman, 2015)

Norovirus RNA genome is covalently linked with a viral protein (VPg) at their 5’ ends and are polyadenlyated at their 3’ ends (Karst, 2010). Several conserved secondary structures that was important for the replication of a MNV was identified through bioinformatic analysis. These structures include two or more 5’ terminal stem-loops, a 3’ terminal hairpin and a stem-loop just upstream of the ORF1/2 junction in the antigenomic strand proposed to be a component of the subgenomic promoter (Simmonds et al., 2008).

2.3.1(a) Norovirus Non-Structural Proteins

Norovirus positive sense ssRNA genome is typically arranged into three ORFs and the genome layout is almost identical even though they are from different genogroups of norovirus. The genomic and sgRNA of norovirus are uncapped and they are instead linked to a short viral protein (VPg) which act as a cap substitute at 5' end followed by a short untranslated region (UTR) ranges from 3 and 5 nucleotides for human and murine norovirus respectively (Gutiérrez-Escolano, Brito, del Angel, & Jiang, 2000). The secondary structures of these UTRs play important roles in viral protein translation (McFadden et al., 2011).

The ORF1, encodes a large polyprotein that comprises of all six mature non-structural proteins (NS1-NS7), which is autocatalytically cleaved to produce individual non-structural proteins (Thorne & Goodfellow, 2014). The non-structural proteins were cleaved from polyprotein by the viral protease during and post-translation; whereby the viral protease itself is catalytically an active component of the polyprotein (Hardy, 2005). These non-structural proteins are important for the establishment of viral replication complexes and genome replication. ORF2 encodes the major capsid protein (VP1) in an alternate reading frame from ORF1 while ORF3 encodes for a minor structural protein (VP2). There are about 20 nucleotides overlapping between ORF1 and ORF2. Nevertheless, ORF2 is not co-translated with ORF1, but it is translated from a sgRNA covalently linked VPg-capped and polyadenylated that is produced during viral replication. ORF3 encodes protein in different reading frame than ORF2 (Thorne & Goodfellow, 2014). The ORF3 is proceed with 3' UTR (forms secondary structures which function during initiation of viral protein translation) as well as the genome is polyadenylated at the 3' end (McFadden et al., 2011; Rohayem, Robel, Jager, Scheffler, & Rudolph, 2006).

ORF1 that encode viral polyprotein which comprises of at least 6 individual protein subunits is the first protein translated following infection of a cell (Hardy, 2005). The first non-structural protein in norovirus is NS1/2 or known as N-term, p48 or '2B-like' protein due to characteristic of this protein that have limited homology to picornavirus 2B protein. This protein carries similar function as picornavirus 2B that engage in rearrangement of membrane which lead to alteration of the membrane permeability (Fernandez-Vega et al., 2004). In addition, NS1/2 also interacts with a host protein necessitate in regulating vesicle transport and can inhibit cellular protein trafficking of cell surface expression (Ettayebi & Hardy, 2003). Other than that, expression of this 45 kDa protein using Norwalk virus showed that this protein is associated with Golgi disassembly suggesting their potential role in replication complex formation (Ettayebi & Hardy, 2003; Fernandez-Vega et al., 2004). According to Sosnovtsev et al. (2006) there is a presence of caspase cleavage site within the NS1/2 of MNV, thus suggesting some cleavage during infection that probably associates virus replication to induction of program cell death.

The second protein constituent of norovirus viral polyprotein is the nucleoside triphosphatase (NTPase), also referred to as p41. NS3 has been described to exhibit NTPase activity that act to bind and hydrolyse nucleotide triphosphate (Pfister & Wimmer, 2001). This protein shares sequence motifs with picornavirus protein 2C. Both proteins do not have helicase activity but unlike 2C, p41 is able to hydrolyse all NTPs instead of ATP only. According to a study conducted by Hyde et al. (2009) in MNV infected cells, NTPase showed that it plays a critical role in the establishment of viral replication complexes.

The third protein component of the polyprotein is NS4, p22, or 3A-like protein as it is similar in gene order to the picornavirus 3A protein of picornavirus. VPg

recruitment to membranous replication complexes during replication is thought to be the function of NS4 (Wobus et al., 2004). Other than that, NS4 also may function in tissue culture adaptation of MNV since presence of sequence changes upon continuous passage of MNV-1 in RAW264.7 cells give rise to attenuated viruses (Bailey, Thackray, & Goodfellow, 2008).

The NS5 protein or VPg is covalently linked to the 5' ends of viral gRNA and sgRNA of caliciviruses in place of a typical 5' cap structure (Rohayem et al., 2006). It acts in a few roles in the viral life cycle but mainly it functions in translation initiation (Karst, 2010; Thorne & Goodfellow, 2014). During viral genome replication, the association of NS5 and RNA is thought to occur when NS5 is used as a peptide primer for genome replication by viral RNA-dependent RNA polymerase or RdRp (NS7). The uridylylation of VPg made it possible to act as primer during viral replication (Belliot, Sosnovtsev, Chang, McPhie, & Green, 2008; Rohayem et al., 2006). Biochemical analysis done *in vitro* performed by Rohayem et al (2006) confirmed the ability of norovirus NS7 to transfer nucleotide to NS5. Plus, there is study carried out by Guix et al (2007) stated that VPg that is covalently linked to 5' end of human Norwalk virus RNA is important for virus infectivity. It is proven when MNV RNA generated from cDNA clone that is capped *in vitro* were found to be infectious when transfected into cell (Yunus, Chung, Chaudhry, Bailey, & Goodfellow, 2010). Other than that, there is also an interaction between VPg and host translation initiation factors take place; probably to recruit them to the 5' end of the RNA genome for translation initiation of virus (Karst, 2010).

Norovirus ORF1 encodes for a large polyprotein that is post-translationally cleaved by the virus-encoded 3C-like protease (NS6) (Blakeney, Cahill, & Reilly, 2003; Sosnovtsev et al., 2006). Even though it is part of the ORF1 polyprotein, the

NS6 is able to be released first from the polyprotein by autocatalytic cleavage then followed by specific cleavage of other proteins from ORF1. This 19 kDa protein is thought to play a role in inhibition of cellular protein synthesis in infected cell. NS6 also able to cleave poly A-binding protein and host cell protein named as elF4G; both were necessary for mRNA translation of host cell (Kuyumcu-Martinez et al., 2004; Willcocks, Carter, & Roberts, 2004). This protein will be discussed in more detail in later Section 2.3.4.

The NS7 protein or can be referred as 3Dpol encodes for viral RNA-dependent RNA polymerase (RdRp) or also known as RNA replicase is an important enzyme for gRNA and sgRNA viral replicaton as it functions in catalysing the synthesis of RNA from RNA template. This protein that is positioned at C-terminus of norovirus non-structural polyprotein also possess a typical right-handed conformation (palm, fingers and thumb) shared by all positive-sense RNA viruses (Ferrer-Orta, Arias, Escarmís, & Verdaguer, 2006). There are four main mechanisms in which recombinant calicivirus RdRp have been demonstrated to initiate RNA synthesis. The mechanisms include a *de novo* initiation and primer-independent initiation, back-priming base initiation and a protein-primed initiation via VPg nucleotidylylation (Rohayem et al., 2006).

2.3.1(b) Norovirus Structural Proteins

Norovirus ORF2 and ORF3 encodes the major and minor structural protein respectively. Both structural proteins are expressed from the viral VPg-linked sgRNA that is 3' co-terminal with the gRNA and polyadenylated.

Norovirus ORF2 contain 58.9 kDa major structural protein VP1 or capsid. Independent expression of VP1 from other viral components results in formation of dimer that can lead to the self-assembly of VLPs (Jiang, Wang, Graham, & Estes,

1992). These structures are morphologically and antigenically exhibit the same properties of native norovirus virions and have been remarkably useful to study a variety of virus-host interactions in the absence of a cell culture system (Jiang et al., 2000). The P2 domain of VP1 consist of a hypervariable region that is suspected to have receptor binding and antigenic sites; consistent with its exposed location on the virion surface. Study conducted by Bailey et al (2008) showed that *in vivo* attenuation is observed when a single amino acid is changed in the P2 domain of MNV (Bailey et al., 2008; Wobus et al., 2004).

Minor structural protein or VP2 is encoded by norovirus ORF3 and it is present in only 1–2 copies per virion (Glass et al., 2000; Karst, 2010). Among the virus within *Caliciviridae* family, VP2 is small, basic, and quite divergent in both size and sequence. VP2 is thought to increases the expression level of VP1 and also stabilisation of capsid (Bertolotti-Ciarlet, Crawford, Hutson, & Estes, 2003). Its role in viral replication is currently undefined but there is evidence that VP2 is essential for production of infectious particle and for viral replication in feline calicivirus (Sosnovtsev, Belliot, Chang, Onwudiwe, & Green, 2005). Moreover, the basic charge of VP2 indicates that it may be responsible in encapsidation of the viral genome. The expression of VP1 is speculating to occur via VPg-directed translation initiation which is similar to expression of non-structural proteins from gRNA while VP2 proceeds via a translation termination re-initiation (TTR) mechanism (Karst, 2010).

Fourth ORF was found in MNV where it overlapped with VP1 coding region and encodes for virulence factor 1 (VF1) (McFadden et al., 2011).

2.3.2 Virus Life Cycle

Previously, knowledge on norovirus pathogenesis are limited and primarily based on physical, histological and biochemical studies conducted on infected human volunteer. Investigation and progression study on its biological aspect has been very slow due to lack of efficient cell culture system and absence of animal model system. However, in 2003, MNV has been discovered as biological model system to study norovirus since it was able to replicate efficiently in tissue culture and also due to establishment of reverse genetics system based on the first isolated strain from MNV (Chaudhry, Skinner, & Goodfellow, 2007; Karst et al., 2003; Ward et al., 2007; Wobus et al., 2004; Yunus et al., 2010). The discovery of excellent model system allowed investigation and analysis of the norovirus' infectivity and pathogenesis *in vivo*. Even though some aspects of MNV such as the disease manifestation does not apply to human norovirus, but this model is the only readily available system which could contribute to understand norovirus translation, replication and its pathogenesis.

2.3.2(a) Attachment and Entry

The life cycle of norovirus begins when there is interaction of virus with its receptor on the cell surface of the host which is known to involve carbohydrate structures such as histo blood group antigen (HBGAs) in human norovirus cases (Donaldson, Lindesmith, LoBue, & Baric, 2010). The susceptibility of individual towards norovirus infection is determined by the secretor status of the H antigen (Hutson, Airaud, LePendu, Estes, & Atmar, 2005). Study conducted by Thorven et al. (2005) showed that individuals who are non-secretors of H-type 1 were found to be resistant to norovirus infection. The absent of H type secretion is due to mutation in the α-(1,2)-fucosyltransferace (FUT2) gene that is involve in the production of H-type 1 antigen in saliva and mucosa. On the other hand, study performed on MNV indicated that the

attachment receptor that is bind in the strain dependent manner of the host cells is through ganglioside-linked terminal sialic acid moieties, glycolipids and glycoproteins (Taube et al., 2012; Taube et al., 2009). Once bound to its receptor, MNV enters cell by which the mechanism remains unclear. However, it is found out that the method of MNV entry into the host cell is dependent on dynamin and cholesterol instead of clathrin and caveolin like feline calicivirus (FCV) (Gerondopoulos, Jackson, Monaghan, Doyle, & Roberts, 2010; Stuart & Brown, 2006). This is because study conducted for MNV is by using RAW264.7 where it did not express caveolin and also there is no reduction in MNV infectivity when there is interfering in the clathrin pathway. Unlike FCV, flotillin depletion and endosomal acidification blockers also did not cause reduction in MNV infectivity. Cholesterol and dynamin pathways is thought to be the method of virus entry due to reduction in MNV infectivity post treatment with methyl-beta-cyclodextrin and dynamin inhibitor; even though the reduction is observed only when the cells is treated before and not after infection (Gerondopoulos et al., 2010).

2.3.2(b) Translation of Viral Proteins

After entry on the positive sense VPg-linked RNA genome into the cytoplasm of the permissive cell, it can immediately act as an mRNA template for initial round of viral RNA translation. Protein synthesis takes place upon recognition of viral RNA by cellular translation initiation factors first, before they are being translated into protein by cellular translation apparatus. Translation of viral gRNA typically use novel mechanism involving VPg to initiate translation and to increase coding capacity of their relatively short genome (Firth & Brierley, 2012; Thorne & Goodfellow, 2014). The VPg which is the viral protein genome linked is a non-structural protein that is covalently linked to the 5’ end of the gRNA and sgRNA where they mediate translation

process to produce viral protein. It also serves as a cap substitute to ensure that its RNA is translated instead of host cell mRNA which constitute of classical 5' cap structure (Bailey & Goodfellow, 2009; Thorne & Goodfellow, 2014). Recently, study conducted by Leen et al. (2013) provided information that virus-encoded VPg protein possessed a compact helical core flanked by intrinsically flexible disordered N- and C- terminus. These flexibilities might aid to the numerous roles carried by VPg in norovirus life cycle (Goodfellow, 2011).

The VPg protein also play a role in the interaction with the component of elF4F host cell translation initiation factor complex. The translation initiation machinery includes cap binding protein (elF4E), a scaffold protein (elF4G) that link the elF4F complex to the elF3 and RNA helicase (elF4A). The Vpg protein bind to the cap binding of elF4F component and this interaction result in recruitment of elF4G scaffold. Recruitment of RNA helicase component elF4A and elF3 complex is enabled following binding of elF4G and elF4E. After that, 43S ribosomal pre-initiation complex is recruited to the complex by elF3 prior to translation initiation (Chaudhry et al., 2006; Daughenbaugh, Fraser, Hershey, & Hardy, 2003; Goodfellow et al., 2005). The 43S pre-initiation complex is formed upon the interaction of the Met-tRNAiMet-elF2-GTP and 40S ribosomal subunit is stabilized by elF3 (Putics, Vashist, Bailey, & Goodfellow, 2010). Although Vpg is observed to bind with elF4E of both MNV and FCV, but its interaction is crucial only in FCV. This is because inhibition in the elF4E activity of FCV show that it severey affect the Vpg linked RNA of FCV but not the in vitro translation of MNV Vpg linked RNA (Chaudhry et al., 2006).

The viral genome translation process will result in the production of proteins where each carry an essential role for viral replication. The first open reading frame encode for non-structural protein that is translated in a form of large polyprotein. This

polyprotein is then co- and post-translationally cleaved by the viral encoded protease at five specific junctions; releasing six mature forms of non-structural proteins (Blakeney et al., 2003; Liu, Viljoen, Clarke, & Lambden, 1999). Research performed by Sosnovtsev et al. (2006, 2002) stated that these proteins able to carry out their function in their separate form as compared to FCV in which their protease and polymerase is in the fusion form named p76. On the other hand, ORF2 and ORF3 encode for VP1 (major) and VP2 (minor) capsid protein respectively. These proteins are part of sgRNA where in the infected cell, they are found to be in higher levels as compared to gRNA (Prasad, Rothnagel, Jiang, & Estes, 1994). Moreover, it is common for positive sense RNA viruses to produce sgRNA during the viral life cycle in order for them to control the viral protein expression. In addition to that, norovirus exhibit a polycistronic subgenomic that require a process to ensure that ribosomes are able to bind and translate the open reading frames that are located downstream. Hence, termination-reinitiation process is used with the aid of overlapping start and stop codon in between the ORF segment to allow expression of products from adjacent ORFs. In this process, the upstream ORF is translated by ribosome until it reaches stop codon at the end of its ORF. After that, the proportion of the ribosomal subunit that remain associated to the mRNA was reinitiated to translate downstream ORF at the start codon of the respective ORF (Napthine et al., 2009).

As mentioned previously, MNV exhibit the fourth ORF that is overlapped with VP1, known as virulence factor 1 (VF1). This VF1 is antagonist to the innate immune response since protein express from ORF4 slow down the upregulation gene activated by the innate pathway. In comparison to the position of ORF2, this sgRNA VF1 started 13 bases downstream and it is possibly initiated by leaky scanning and two slips in the ribosome. Reverse genetic study indicated that this protein is not required for

replication in tissue culture and indeed only function as a viral accessory protein (McFadden et al., 2011).

2.3.2(c) Genome Replication

The expression of viral non-structural protein lead to the formation of cytoplasmic membrane-bound replication complexes; thus allowing the replication process of viral genome (Bailey & Goodfellow, 2009). The replication complex is thought to act as the surface for the assembly of both viral and cellular replication factor where it also comprises of a few components that are essential for viral replication such as viral RdRp, viral RNA (single and double stranded intermediates), other viral enzymes as well as host cell factors. The replication for all positive strand RNA viruses requires genome to function as mRNA for production of viral protein needed for genome replication and as a template for negative strand RNA synthesis (Thorne & Goodfellow, 2014).

The RNA synthesis takes place following the interaction of RdRp with other viral and cellular factors which enable it to bind at the promoter sequence located 3' end of both strand of RNA. Since norovirus viral genome RNA is in the form of positive strand, it can directly be used as the template to generate negative strand RNA by *de novo* initiation. The produced negative sense RNA is then subsequently used for the production of higher levels of positive strand gRNA and sgRNA. The production of sgRNA is the unique characteristic possess by all caliciviruses during their replication in infected cells. The sgRNA is attach to Vpg at the 5'end and polyadenylated at the 3' end. They are expressed at the 3' co-terminal of the full-length gRNA. The sgRNA of caliciviruses consist of ORF2 and ORF3 (and ORF4 for MNV) which encode for viral structural proteins. There are two proposed mechanism for production of sgRNA where both use negative sense RNA intermediate and RdRp encoded by NS7 gene in

ORF1 (Karst, 2010; Putics et al., 2010; Royall & Locker, 2016). Plus, norovirus RdRp exhibit an active site residues that is structurally and functionally conserved as other positive strand RNA viruses RdRp (Högbom, Jäger, Robel, Unge, & Rohayem, 2009).

The first mechanism involves premature termination during the synthesis of negative-sense gRNA. The NS7 RdRp terminates during elongation upon reaching termination signal that is located upstream of VP1 coding region, generating negative sense of subgenomic strand. The negative strand is then used as a template for VPg-dependent RNA synthesis to produce positive sense sgRNA; arising from *de novo* initiation. The replication of sgRNA will be repeated and eventually packed using VP1 capsid protein to form new virus particle. The model also proposed that higher production of positive sense viral RNA as indicated by increasing level of VP1 would lead to the formation of multimeric complexes by VP1 in order to prevent its interaction with RdRp. These result in the capsid assembly and consequently the formation of new infectious viral particles (Putics et al., 2010; Thorne & Goodfellow, 2014).

The second mechanism for sgRNA synthesis known as internal initiation involved a presence of secondary structure upstream of ORF2 in the negative sense gRNA where it serves as the promoter to generate positive-sense sgRNA. The RdRp will initiate RNA synthesis following VPg-dependent fashion to produce new strand of sgRNA strand. Sufficient sgRNA is further encapsidated to produce new viral particle (Thorne & Goodfellow, 2014). Bioinformatic study conducted by Simmonds et al. (2008) in among all the members of *Caliciviridae* family found out the presence of highly conserved RNA stem loop structure located downstream of VP1 coding region on the negative sense gRNA. The structure that is positioned 6 nucleotide (nt) from the start of the sgRNA is confirmed for its accuracy by the genetic and biochemical