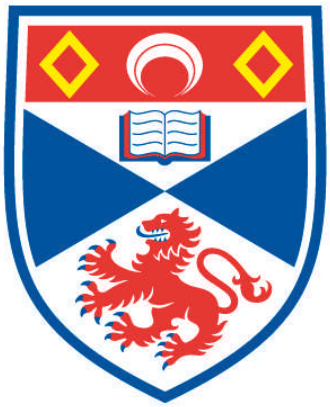
MOLECULAR BIOLOGY OF GIANT VIRUSES' DNA REPLICATION MACHINERY

Antonia Alexandra Evripioti

A Thesis Submitted for the Degree of MPhil
at the
University of St Andrews



2013

MPhil Research Project

Molecular Biology of Giant Viruses'

DNA Replication Machinery

Antonia Alexandra Evripioti

November 2011 - August 2013

Student ID: 040001067

Supervisor

Dr. Stuart MacNeill

Re-Submission due Date: 3rd August 2013

DECLARATIONS

1. Candidate's declarations:

I, **Antonia A. Evripioti**, hereby certify that this thesis, which is approximately **37,900** words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in **February 2010** and as a candidate for the degree of **M.Phil** in **November 2011**; the higher study for which this is a record was carried out in the University of St Andrews between **2011** and **2012**.

Date ..... Signature of candidate .....

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of **Master of Philosophy (MPhil)** in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date ..... Signature of supervisor .....

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

(i) Access to printed copy and electronic publication of thesis through the University of St Andrews.

Date .....

Signature of candidate ..... Signature of supervisor .....

ACKNOWLEDGEMENTS

I appreciatively acknowledge the people who helped me during this research project. Firstly, my heartfelt thanks go to my supervisor, Dr. S. MacNeill, for allowing me the opportunity to return and study at the University of St. Andrews while, at the same time, I am most grateful for his valuable support, guidance and interest. Secondly, my sincere thanks go to my tutor, Dr. R. Ramsay, and my internal examiner, Dr. M. Ryan, for their beneficial advice. I also wish to thank all my colleagues for their expertise. In particular, I would like to send my deep gratitude to Xavier Giroux, Jasmin Schnick and Agnieszka Swiatek, for their patience and kindness. Finally, my deepest thanks go to my parents and my boyfriend for their continuous encouragement, love and understanding.

ABSTRACT

Viruses are the most widespread and abundant entity on this planet, further constituting the largest part of the genosphere. The majority of these infectious agents are miniature, having been described as being smaller than the smallest bacteria. Even though they encode a limited number of viral proteins, they still obtain the bulk of the material they require for their replication and propagation from the infected host cell.

Recently, this traditional concept of viruses has been shaken up by the breakthrough finding of a new group of viruses, the Giant Viruses. They have been assigned this definition due to their amazingly and surprisingly large genomic size. The vast majority have their own replication machinery. They have been discovered in the sea, where they prefer to infect amoebas and other marine microorganisms. For the purpose of this study, we focused on three of these giant viruses; Mimivirus, Marseillevirus, and *Cafeteria roenbergensis virus* (CroV).

The aim of the study was to comprehend how these giant viruses replicate and propagate their genetic material through the generations, to have reached a point where their genome size is comparable to normal-sized bacteria. For this reason, an extensive biochemical analysis on the molecular biology of giant viruses' DNA replication machinery was performed, hoping to obtain new insights into the evolution and lifestyle of these unique viruses. We specifically focused on what we considered to be two of the most important DNA replication proteins; the *Proliferating Cell Nuclear Antigen* (PCNA) and *Flap structure-specific Endonuclease 1* (FEN1). Our goal was to determine their properties.

The protocols performed were a series of protein expression procedures, during which the particular synthetic genes were cloned in a selection of expression vectors and were then expressed in bacteria (i.e. *E.coli* host expression strains). Depending on the protein expression efficiencies, some trial protein purification procedures followed.

For the first few months of the project, however, it was impossible to obtain any conclusive results concerning the expression of the proteins. The synthetic genes were proving to be extremely difficult to express in vectors containing an expression tag. Only when we switched to un-tagged expression vectors, much later on in the project, did we start getting better and more promising results. This was a particularly useful outcome in itself, as it revealed that enhanced expression of the PCNA and FEN1 proteins preferentially occurs when no expression tags are present. Towards the end of the project, some protein purification trials were performed, but unfortunately they only resulted in an incredibly low protein purity level.

The discovery of these distinctive viruses has not only incited scientists to maybe rethink and change their view about the general nature of viruses, but it has also begun to alter and question the outlook regarding the history of life as a whole. As the investigation is still in its very early stages, there are many aspects concerning the giant viruses still to be discovered. This in the end could essentially teach us a great deal more than we ever hoped to expect, and therefore it is of great significance and importance to continue with this research.

ABBREVIATIONS

aa	Amino Acids
AEP	Archaeo-Eukaryotic Primase
AHT	AnHydroTetracycline
APMV	Acanthamoeba Polyphaga MimiVirus
ATP	Adenosine TriPhosphate
HAc	Acetic Acid
APS	Ammonium PeroxodiSulfate
ARS	Autonomously Replicating Sequences
BFM	Bacteria Freezing Medium
bp	Base Pairs
CDKs	Cyclin-Dependant Kinases
CroV	Cafeteria roenbergensis Virus
CO_2	Carbon Dioxide
^oC	Degree Celcius
DDKs	Dbf4-Dependant Kinases
DMSO	DiMethyl SulfOxide
DNK	DeoxyriboNucleoside Kinase
DTT	DiThioThreitol
E.coli	Escherichia coli
EDTA	-Diamine-Tetra-Acetic acid
ESI	ElectroSpray Ionisation
EtOH	Ethanol
FEN1	Flap structure-specific EndoNuclease 1
GFP	Green Fluorescent Protein
g	Grams
HCl	Hydrogen Chloride
HGT	Horizontal Gene Transfer
HIC	Hydrophobic Interaction Chromatography
HJR	Holiday Junction Resolvase
HRP	Horse Radish Peroxidase
I	Induced (protein expression)
IPTG	IsoPropyl β-D-1-ThioGalactopyranoside

Kb	Kilo bases
KDa	Kilo Dalton
LB	Luria-Bertani medium
MAR	Marseillevirus
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MCM	MiniChromosome Maintenance
ml	Milli liters
mM	Milli Molar
µl	Micro liters
µM	Micro Molar
mRNA	Messenger RiboNucleic Acid
MS	Mass Spectrometry
OD	Optical Density
ORC	Origin Recognition Complex
P	Pellet (non-soluble protein material)
PBS	Phosphate-Buffered Saline
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PD	Pull-Down (column purified protein sample)
P-groups	Phosphate-groups
PEG	PolyEthylene Glycol
PIP-box	PCNA-Interacting Peptide box
Pre-IC	Pre-Initiation Complex
Pre-RC	Pre-Replication Complex
RFC	Replication Factor C
RNA	RiboNucleic Acid
RPA	Replication Protein A
rpm	Revolutions per minute
RT	Room Temperature
NAD	Nicotinamide Adenine Dinucleotide
NCBI	National Center for Biotechnology Information
NCLDV	NucleoCytoplasmic Large DNA Viruses
NDK	Nucleoside-Diphosphate Kinase
NER	Nucleotide Excision Repair

2xSB	2x Sample Buffer
NaCl	Sodium Chloride
NaH_2PO_4	Sodium Dihydrogen Phosphate
S	Soluble (solubilised protein)
SDS-PAGE	Sodium Dodecyl Sulfate - PolyAcrylamide Gel Electrophoresis
SF	SuperFamily
SW1/SNF2	(SWItch/Sucrose NonFermentable)
Na_2HPO_4	Sodium Phosphate Dibasic
TAE	Tris-Acetate EDTA buffer
TEMED	TetraMethylEthyleneDiamine
TEV protease	Tobacco Etch Virus protease
TSB	TranSformation Buffer
U	Uninduced (no protein expression)
UV	UltraViolet
w/v	Weight/Volume

CONTENTS

LIST OF FIGURES

Chapter 1 - Introduction

Chapter 2 - Materials and Methods

Chapter 3 - Results

LIST OF TABLES

Chapter 1 - Introduction

Chapter 2 - Materials & Methods

-CHAPTER ONE-

Introduction

1.1 Viruses

Viruses were first discovered and differentiated from bacteria in the late 1800s when it was observed that they could not be isolated by filtration protocols due to their exceptionally minute sizes (*Minor PD, 2007*; *Van Etten JL, 2011*). A typical virus can be seen in *Figure 1.1*.

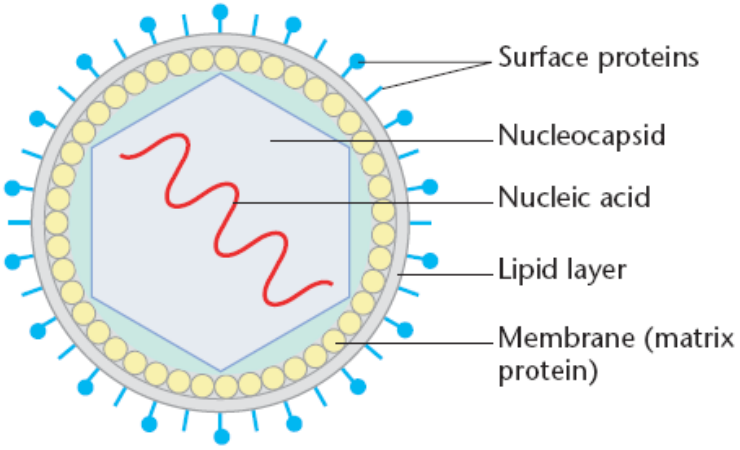


Figure 1.1: Example of a common virus structure and the parts it comprises of [*Minor PD, 2007*].

According to the traditional perception of viruses; these are tiny infectious cells forming a large and distinct group, defined by their ability to cause disease. They are obligate intracellular parasites that infect all kinds of organisms and by doing so they depend on these organisms' raw material, cell machinery and metabolism for their own replication.

Nevertheless, they do contain some of their own genetic material necessary for the production of progeny, some of which may or may not have been acquired from their hosts' during their evolution. [*McKenna R & Faulkner L, 2001; Minor PD, 2007*]. The process that viruses make use of for increasing their genetic diversity is known as horizontal gene transfer (HGT), during which they can basically incorporate host genes into their own genome; this method is of spectacular importance during their evolutionary reproduction, because by performing this procedure viruses can essentially select and only pick up the relevant host material required for their subsequent replication, thus providing them with a way by which they can productively evolve (*Domingo E, 2007*; *Koonin EV & Yutin N, 2010*).

Viruses distinctively replicate by forcing their infected host organism to manufacture more of the same complex viral components. They do this by inserting their own genetic information, which has been specifically packaged for this purpose, into the host cells. This process will cause the host system to divert its replication mechanism and biosynthesis machinery in a way that the virus requires, so instead of producing cellular proteins it will produce more viruses. Viruses pass their packaged genetic information into the hosts' cells by an accurate delivery system that ensures that the suitable cells and cell compartments are reached for the virus to multiply successfully. For this purpose, the parasitised cell must also be able to recognise and decode the viral genetic information, a fact that the virus itself has ensured to occur properly once inside the host cells. By using their host as a template for viral replication, which will take place in a shielded environment, and hijacking their normal cellular processes, viruses can also take advantage of the host cells' biology. This ensures that the viral material produced is assembled in an appropriate way

and emerges from the cell in the correct packaged form so as to infect other cells. These newly manufactured viruses will then escape the host and infect passively another one, hence continuing their replication cycle passing on their viral genome to subsequent generations, whilst at the same time they constantly mutate and acquire new genetic material. This whole process is done in a very profound, precise, but at the same time, subtle way. [*Cann AJ, 2001*; *Minor PD, 2007*; *Harper DR, 2012*]. A diagrammatical representation of the viral life cycle is illustrated in *Figure 1.2*.

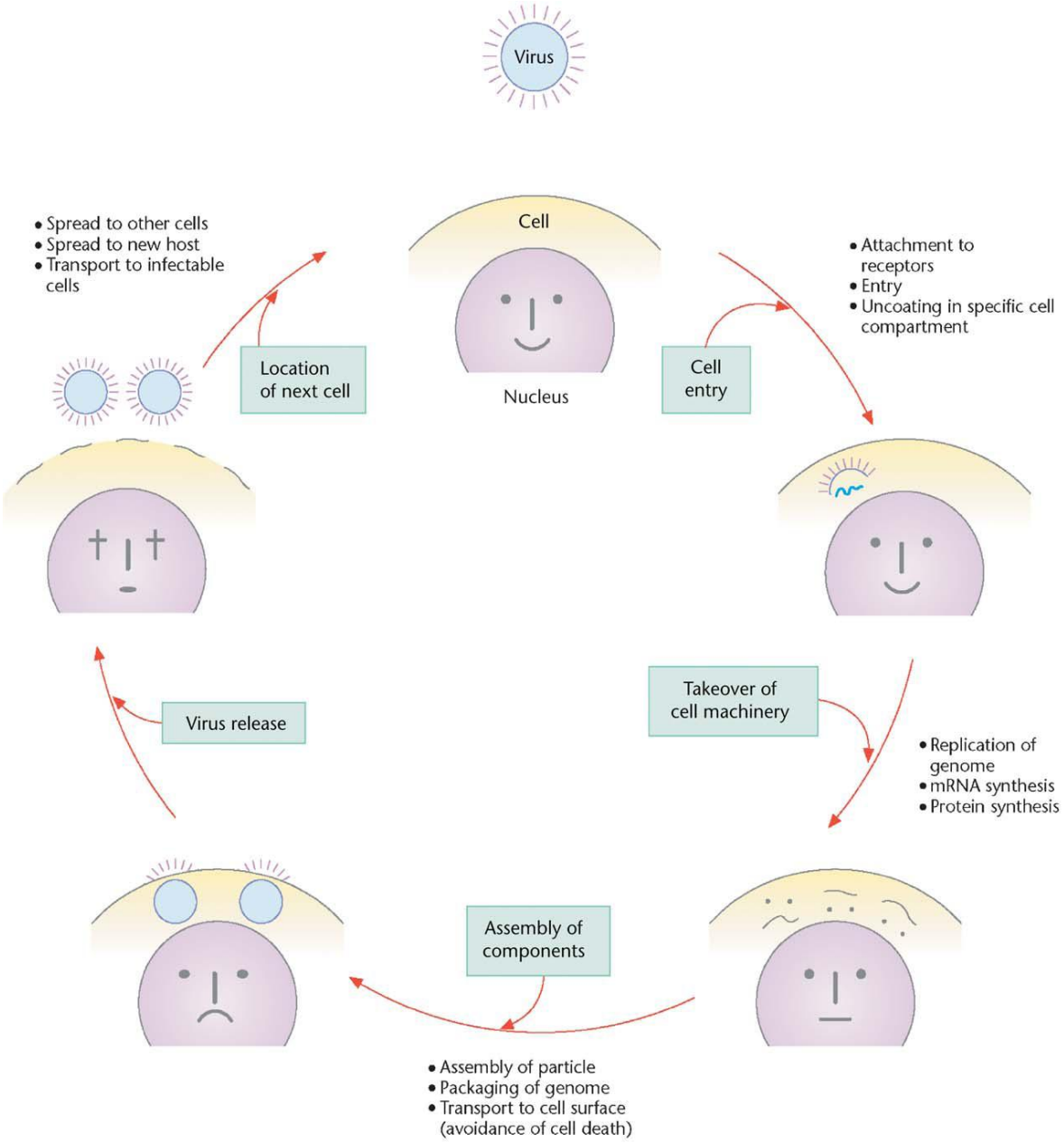


Figure 1.2: The life cycle of a virus [*Minor PD, 2007*].

In general, viruses are as diverse as the number of species that exist; this is the case due to genetic variation and evolution. Viruses differ significantly both in their magnitude and their genetic complexity, while they are also capable of using an incredible diversity of strategies so as to reproduce in the host cells. Therefore, they are commonly classified depending on their replication pathways, and the fact that there is a huge variety of such procedures suggests that viruses have an amazing evolutionary divergence having evolved through multiple origins during the course of history. [*McKenna R & Faulkner L, 2001*; *Iyer LM et al, 2006*; *Domingo E, 2007*].

The origin of viruses is a subject extensively questioned and greatly debated upon amongst experts (*Minor PD, 2007; Wessner DR, 2010*). Several models concerning this type of origin have been proposed over the years. Three main hypothesis have especially been circulated; 1. The progressive, or escape; 2. The regressive, or reduction; and 3. The predatory or co-evolutionary hypothesis. First of all, the progressive theory asserts that viruses arose from primitive genetic elements or forms of cells, which gained the ability to shift between other types of cells hence gradually acquiring genetic material from their host organisms. Secondly, the regressive theory assumes that viruses broke loose from the original complex cells, which they now infect. This suggests that viruses are actually remnants of the modern complex cellular organisms that gradually degenerated, losing their ability to synthesise important genetic material thus rendering themselves dependent on the host cells. [*Iyer LM et al, 2006*; *Moreira D & Brochier-Armanet C, 2008; Claverie JM & Abergel C, 2010*]. Lastly, the predatory or co-evolutionary theory simply states that viruses predate or

co-evolved simultaneously with their current cellular hosts, both of which had common replication origins promoting lateral gene transfer. This refers to an independent co-evolution of viruses and complex cell organisms, both of which exploited cells that developed concurrently. [*Domingo E, 2007; Filee J et al, 2008; Flugel RM, 2010; Sinkovics JG; 2011*]. Co-existence, nevertheless, of these two entities surely suggests that viruses have also evolved mechanisms to counteract the hosts' defensive responses (*Forterre P, 2010*).

Over the years, scientists all over the world have been struggling with the challenge of deriving a single phylogenetic tree that would relate all known viruses (*Domingo E, 2007*; *Minor PD, 2007*). However, their great diversity and abundance has designated them as polyphyletic, even though different categories of viruses appear to have significant similarities in their structure, organisation of their genome and replication strategies (*Moreira D & Brochier-Armanet C, 2008*).

An exciting new aspect is the fairly recent knowledge of the existence of viral 'hallmark genes', meaning the genes found within viral genomes to be central for virus replication that are in fact absent from cells (*Koonin EV & Yutin N, 2010*). This information has now led to the suggestion of an ancient 'virus world'. This proposal would further suggest that during evolution viruses have had a major part in the formation of all the other types of organisms, i.e. archaea, bacteria and eukarya. [*Domingo E, 2007*; *Van Etten JL, 2011*]. This is a topic of immense discussion.

A question that has arisen quite recently, specifically after the discovery of some members of the giant viruses group, is whether or not viruses should be included in the tree of life, together with archaea, bacteria and eukarya. Should viruses be considered as living organisms or not? This is another issue that has caused tremendous discussions over recent periods of time. [*Minor PD, 2007*; *Raoult D & Forterre P, 2008; Claverie JM & Abergel C, 2010; Ruiz-Saenz J & Rodas JP, 2010; Van Etten JL, 2011*]. The conventional definition of a 'living organism' is in simple terms an organism that can move, grow, reproduce and evolve, carrying out metabolic process and responding to external stimuli. But do viruses comply with these characteristics? The answer is not as straight forward as expected; they do and they do not. Even though we are aware that they reproduce, transfer between cells and evolve over time, they do not, however, perform metabolic processes. A living organism requires the generation of adenosine triphosphate (ATP), as well as the presence of ribosomes and other translational machinery, for the purpose of forming proteins. Thus, going by the textbook definition, viruses cannot independently form proteins from mRNA; they require a living host cell for their replication, thus leaving them completely bound up with the cellular processes of the cell. [*Moreira D & Brochier-Armanet C, 2008; Wessner DR, 2010*].

The discovery of the giant viruses is now here to stir up this conservative concept and give reason to argue that viruses may indeed be living organisms, which have their own right to be included in the tree of life (*Moreira D & Brochier-Armanet C, 2008; Yutin N & Koonin EV,*

2009; *Yutin N et al, 2009*; *Boyer M et al, 2010*; *Colson P et al, 2011*). Nonetheless, will it ever be entirely possible to tell where all the diverse species of viruses truly originate from?

1.2 Giant Viruses

Around the turn of the decade, a completely new group of viruses unexpectedly began to be discovered in aquatic habitats. The only reason that these viruses were not discovered until recently is that up till now researchers have focused all their time and energy on viruses that infect humans, animals and plants, as it has been crucial to understand all aspects of their replication strategies and hence pathogenicity (*Boyer M et al, 2009*). At the same time, because viruses have always been considered the smallest entities on earth, no scientist could even come close to imaging the existence of significantly bigger ones. This conservative notion changed with the breakthrough finding of novel viruses that were assigned into a separate group, namely the giant viruses group (often referred to as "giruses"), relating to their unique properties and remarkably unanticipated large size (i.e. some of them being as big as a normal-sized bacterium cell, thus possibly making their isolation by filtration protocols more plausible). [*Claverie JM & Abergel C, 2010; Fischer MG et al, 2010*; *Raoult D & Boyer M, 2010*; *Van Etten JL et al, 2010*; *Van Etten JL, 2011*].

These giant viruses were notably located in marine microorganisms, such as amoebas and microzooplankton. These microorganisms, which are thought of as wild phagocytes consuming everything that comes their way (from phytoplankton to heterotrophic bacteria

and viruses), can ingest quite large-sized particles rendering them a highly potential source of giant viruses, besides making them the perfect environment from which diverse forms of viruses may have emerged from throughout evolution. [*Boyer M et al, 2009; Fischer MG et al, 2010; Moliner C et al, 2010; Van Etten JL et al, 2010*; *Van Etten JL, 2011*]. A number of the amoebae discovered have currently the largest genome size estimated on Earth (*Raoult D & Boyer M, 2010*). New viruses belonging to the giant viruses group are being explored with increasing frequency ever since.

The giant viruses found in the aqueous microenvironments display a remarkable degree of 'biological sophistication', when compared to simpler cellular life forms. Even the viruses that are classified as being in the same sub-family can have divergent lifestyles, morphologies, and of course, they may differ vastly in their genetic complement. Specifically, having risen from various origins, they have an extremely complex repertoire of genetic material. [*Ogata H & Claverie JM, 2007*; *Van Etten JL et al, 2010*; *Van Etten JL, 2011*]. Nevertheless, both giant viruses and cellular life forms appear to have evolved by analogous mechanisms, including HGT and gene duplication events (*Suhre K, 2005*; *Monier A et al, 2007*; *Moreira D & Brochier-Armanet C, 2008; Boyer M et al, 2009; Filee J & Chandler M, 2010*).

It has been determined that amoebas play a major role in this diverse genomic repertoire, due to the fact that they allow themselves to be parasitised at the same time by various other organisms ─from *bacteria*, *archaea* and *eukarya* to viruses─ and, as such, they act as

the perfect environment in which intracellular bacteria and viruses can live and experience a sympatric lifestyle (*Figure 1.3*) (*Colson P & Raoult D, 2010*; *Moliner C et al, 2010*; *Raoult D, 2010*; *Raoult D & Boyer M, 2010*). This further accounts for the lateral gene transfer and gene exchange between the host and the parasites, but also between the different parasites themselves.

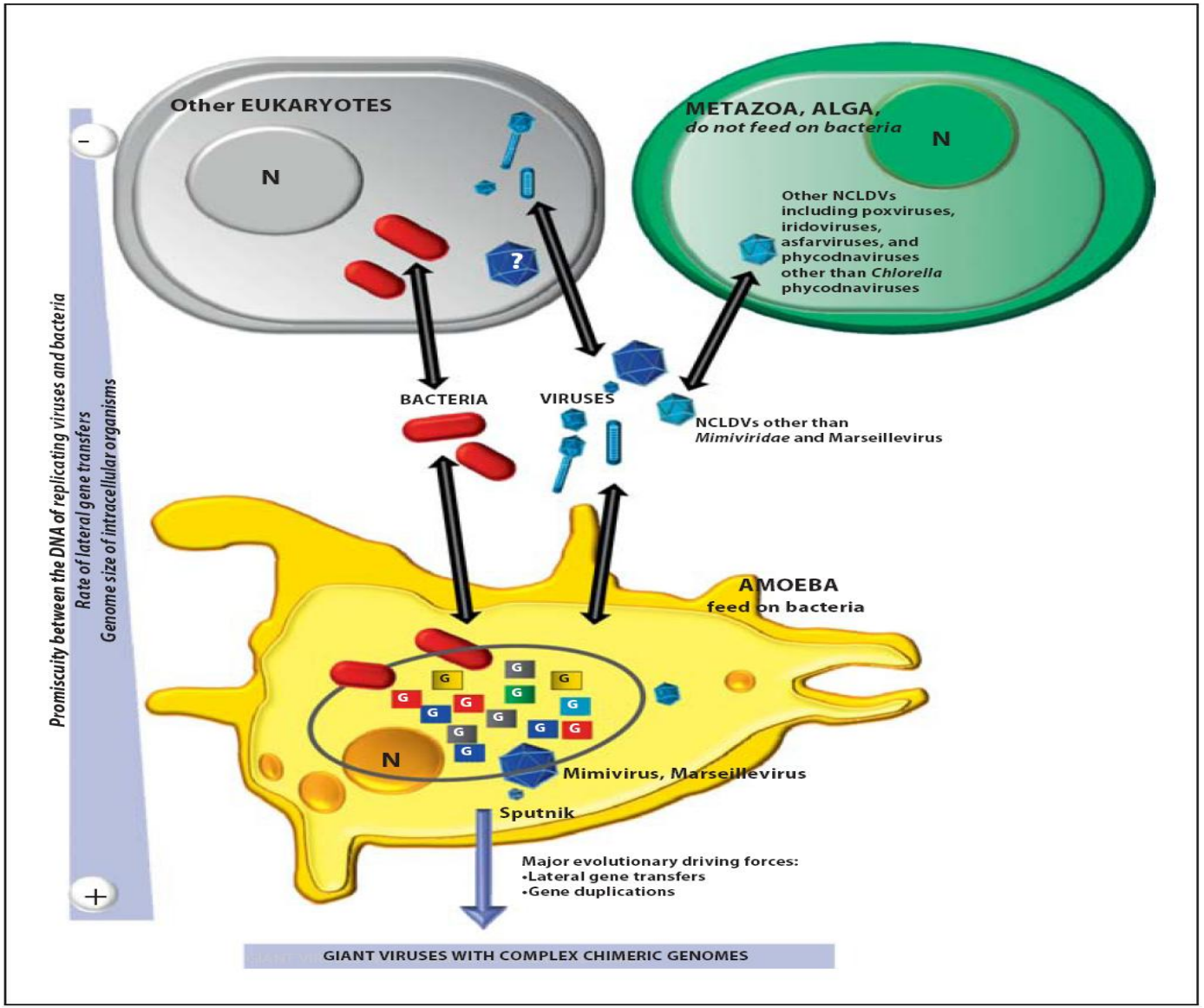


Figure 1.3: Intra-amoebal lifestyle as a source of complex chimeric gene contents. Colored boxes containing a G indicate genes from various origins (bacteria, viruses, eukaryotes) [*Colson P and Raoult D, 2010*].

The biology and evolutionary origin of giant viruses has, without a doubt, provoked ardent debate over time (*Forterre P, 2010*; *Van Etten JL, 2011*); with some scientists arguing that these viruses are simply "gene robbers", acquiring their genetic material from their parasitised hosts through the HGT pathway (*Moreira D & Brochier-Armanet C, 2008; Filee J & Chandler M, 2010; Fischer MG et al, 2010*), whereas others support the hypothesis that these viruses date back to the original emergence of eukaryotic cells and hence their genetic material is viral in origin (*Monier A et al, 2007*; *Mrazek J & Karlin S, 2007*; *Flugel RM, 2010*).

In recent years, and especially since the discovery of giant viruses, it has become apparent that protozoans possibly host the largest and most complex viruses, while the remaining giant viruses are most likely prevalent in the oceans (*Fischer MG et al, 2010*; *Raoult D & Boyer M, 2010; Van Etten JL et al, 2010*; *Van Etten JL, 2011*). However, so far only the giant viruses infecting the *Acanthoamoeba spp.* and the *Cafeteria spp.* have been to some extent characterised, even though it has become quite evident that some of the giant viruses are most likely pathogens of phytoplankton (*Claverie JM et al, 2009-a/b*). Therefore, maybe it is about time that scientists turned their attention to other forms of life present in our oceans; they may be in for more surprises.

Nowadays, a gradual rising number of giant viruses have been uncovered from diverse aquatic environments, implying that they may possibly comprise an ubiquitous and

quantitatively important part of marine viruses (*Moliner C et al, 2010*; *Van Etten JL et al, 2010*; *Van Etten JL, 2011*). As a consequence, it is worthy to note that marine giant viruses may have a considerable impact on the ocean ecosystems as a whole. Thus, understanding the biology behind these viruses may potentially be very significant for the management of future ecosystems.

1.2.1 Mimivirus (APMV)

Mimivirus (or *Acanthoamoeba polyphaga mimivirus*; Mimivirus is short for "mimicking microbe") was discovered in 2003 within the freshwater *Acanthoamoeba polyphaga spp*., from which it took its name. When it was first discovered it had been mistaken for a bacterial cell due to its unusually big size. [*Raoult D et al, 2004; Renesto P et al, 2006*; *Claverie JM & Abergel C, 2009*; *Fischer MG et al, 2010*; *Raoult D & Boyer M, 2010*; *Van Etten JL, 2011*]. It is the first member of the *Mimiviridae* family of viruses and it belongs to the nucleocytoplasmic large DNA viruses (NCLDV) group, which is a monophyletic group of double-stranded DNA viruses containing a conserved core set of replication proteins (viruses belonging to this group may have a common ancestor). These viruses have been named as such because, apart from having a typical stage during their replication cycle that takes place in the nucleus of the host, they also have a stage that occurs within the cytoplasm, hence separating their replication and expression activities from the host genome, whilst at the same time they contain within their own genome the appropriate genes to help them do so. [*Yutin N & Koonin EV, 2009*; *Colson P and Raoult D, 2010*; *Filee J & Chandler M, 2010*; *Koonin EV & Yutin N, 2010*; *Mutsafi Y et al, 2010*; *Colson P et al, 2011*].

Mimivirus is the second largest virus in size, but has the largest and rather complex genome, explaining why it was originally characterised as being a bacterial cell. It has a 1.2 Mb genome (more than double the size of any previously sequenced viral genome), which specifically possesses an estimated 981 protein coding genes, as determined by multiple sequencing analysis studies. [*Raoult D et al, 2004*; *Suhre K, 2005*; *Kuznetsov YG et al, 2010; Legendre M et al, 2011*]. From these protein coding genes, 21 of them encode homologs to proteins that are found to be highly conserved in the majority of NCLDVs, while some of its genes are unique amongst viruses (some of these are expressed in living organisms). Nonetheless, the majority of its genes have no cellular homologs and are thus presumed to be very ancient. [*Renesto P et al, 2006*; *Claverie JM & Abergel C, 2009*; *Yutin N & Koonin EV, 2009*; *Colson P and Raoult D, 2010*; *Colson P et al, 2011*]. A second group of the Mimivirus' genes is predicted to have arisen by gene duplication events, while the third group consists of genes that were horizontally transferred into the Mimivirus genome from eukaryotic and bacteria hosts (*Filee J et al, 2006; Moreira D & Brochier-Armanet C, 2008; Colson P and Raoult D, 2010*; *Filee J & Chandler M, 2010*; *Fischer MG et al, 2010*; *Raoult D & Boyer M, 2010*). Mimivirus is the first giant viruses described to contain genes involved in replication, transcription and translation, bringing it somewhat closer to the description of a typical living cell (*Claverie JM et al, 2006*). A summarised representation of the gene content of the Mimivirus genome can be seen in *Figure 1.4*. Not much is known about its replication cycle.

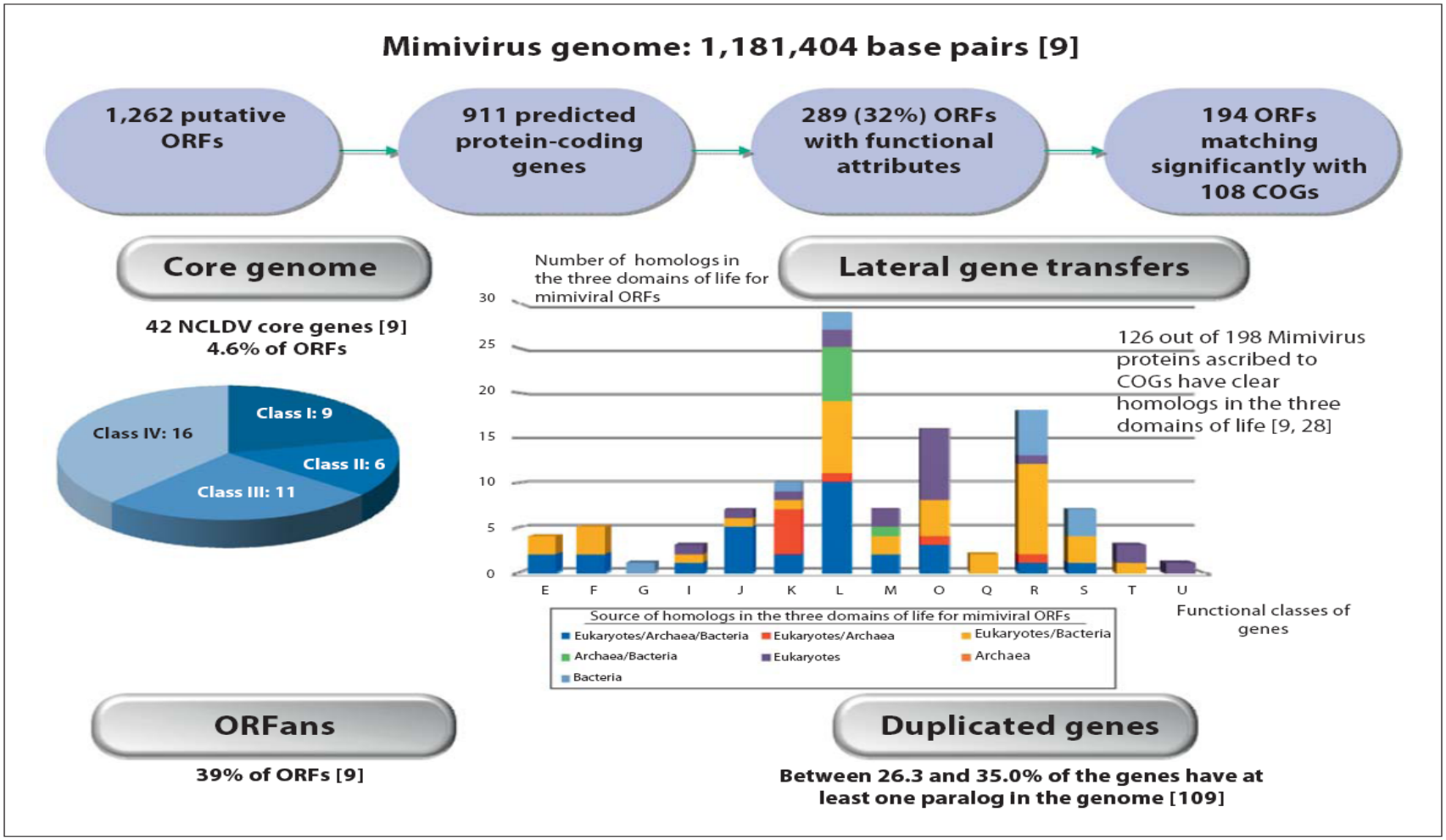


Figure 1.4: Schematic representation of the Mimivirus gene content. Abbreviations for the COG functional classes: E = amino acid transport and metabolism; F = nucleotide transport and metabolism; G = carbohydrate transport and metabolism; I = lipid transport and metabolism; J = translation; K = transcription; L = replication, recombination and repair; M = cell wall/membrane biogenesis; O = post-translational modification, protein turnover, chaperones; Q = secondary metabolite biosynthesis, transport and catabolism; R = general function prediction only; S = function unknown; T = signal transduction mechanisms; U = intracellular trafficking and secretion [*Colson P and Raoult D, 2010*].

Different phylogenetic studies conducted on this virus have suggested that it may possibly be one of the very early divergents of the NCLDV group that evolved through extensive integration of genes between widely diverse genomes, while other studies have hypothesised that it may be related to a type of DNA virus that emerged even before cellular organisms did, and therefore, played a key role in the development of all life on Earth (*Claverie JM et al, 2006; Iyer LM et al, 2006*; *Koonin EV & Yutin N, 2010*; *Colson P et al, 2011*). An alternative hypothesis, however, is that originally three distinct types of DNA viruses existed that were involved in generating the three known domains of life (*Filee J et al, 2008*; *Filee J and Chandler M, 2010*; *Van Etten JL, 2011*). Whatever the actual origin of this virus is, its remarkably large size definitely helps it establish a bridge between the viral and cellular worlds. The origin of viruses, or rather giant viruses, is a topic that requires immense investigation.

A number of recent studies, nevertheless, have provided exceptionally strong arguments in favour of a fourth domain of life containing the NCLDV group of viruses (*Claverie JM, 2006; Raoult D & Forterre P, 2008; Ruiz-Saenz J & Rodas JP, 2010; Nasir A et al, 2012*). This theory has basically been supported by various bioinformatic and phylogenetic analyses based on a variety of common sets of proteins involved in information storage and processing, including genome replication (*Colson P et al, 2011*). The proteins under investigation were found to be conserved in all three domains of life, as we know them today (i.e. *Archaea*, *Bacteria* and *Eukarya*), while at the same time they were discovered to be conserved also in viruses. Additionally, more detailed analysis of the phylogenetic trees,

presented in some of these investigations (e.g. paper published by *Boyer M et al, 2010*), revealed distinct *Archaea*, *Bacteria*, *Eukarya* and NCLDV monophylies, and so led the researchers to the conclusion of "the existence of a viral clade with ancestral DNA replication machinery branching separately from *Archaea*, *Bacteria* and *Eukarya*". On the contrary, other explorations (i.e. research conducted by *Williams TA et al, 2011*) argue against this proposition; they particularly state that information gene phylogenies do not in fact support a fourth domain of life containing the NCLDVs (*Moreira D & Brochier-Armanet C, 2008*). A third group of studies argues that the answer to this question is much more complex, and as such, cannot be solved by conducting only computation analysis of the giant viruses (*Yutin N & Koonin EV, 2009*; *Yutin N et al, 2009*). The question is where does the truth lie in reality, and the answer is that no one truly knows as yet.

As a final point, Mimivirus may be a causative agent of some forms of pneumonia. This matter arose from a single reported case when a laboratory technician, who had previously dealt with a living Mimivirus particle, unfortunately died (*Raoult D et al, 2006*). There is no evidence to prove, however, that this incident was directly linked to the Mimivirus particle, and no more cases have been accounted for in the following years. Nowadays, however, hospitalised pneumonia patients are screened for the presence of Mimivirus particles in their respiratory tracts (*La Scola B et al, 2005; Dare RK et al, 2008*).

1.2.2 Marseillevirus (MAR)

Marseillevirus was first isolated in 2009 from *Acanthoamoeba polyphaga spp.* It is the only member so far of a new family of viruses termed *Marseilleviridae* (closely related to the *Iridoviridae* and *Ascoviridae* sub-families), but it is a prototype of the NCLDV group of viruses. It has the fifth largest viral genome sequenced to date, encoding for a minimum of 49 proteins, as well as some mRNAs encompassing a 368 kb genome. The genetic material included within the core of this virus is rather varied compared to other viruses. Even though it contains typical NCLDV genes (some also found within the Mimivirus), nonetheless it also contains other genes that have apparently been obtained from some of its eukaryotic hosts, as well as their parasites or symbionts, these being viral, bacterial and/or archaeal, perhaps through the HGT mechanism (*Figure 1.5*). [*Boyer M et al, 2009; Raoult D & Boyer M, 2010; Deresinski S, 2010*]. As with all other giant viruses characterised, gene duplication events must have also occurred sometime during the Marseillevirus evolutionary history (*Colson P and Raoult D, 2010*; *Filee J & Chandler M, 2010*; *Van Etten JL, 2011*). Not much more is known about this giant virus or its replication cycle, but, similarly to Mimivirus, it is considered to contain genes involved in its replication, transcription and translation processes.

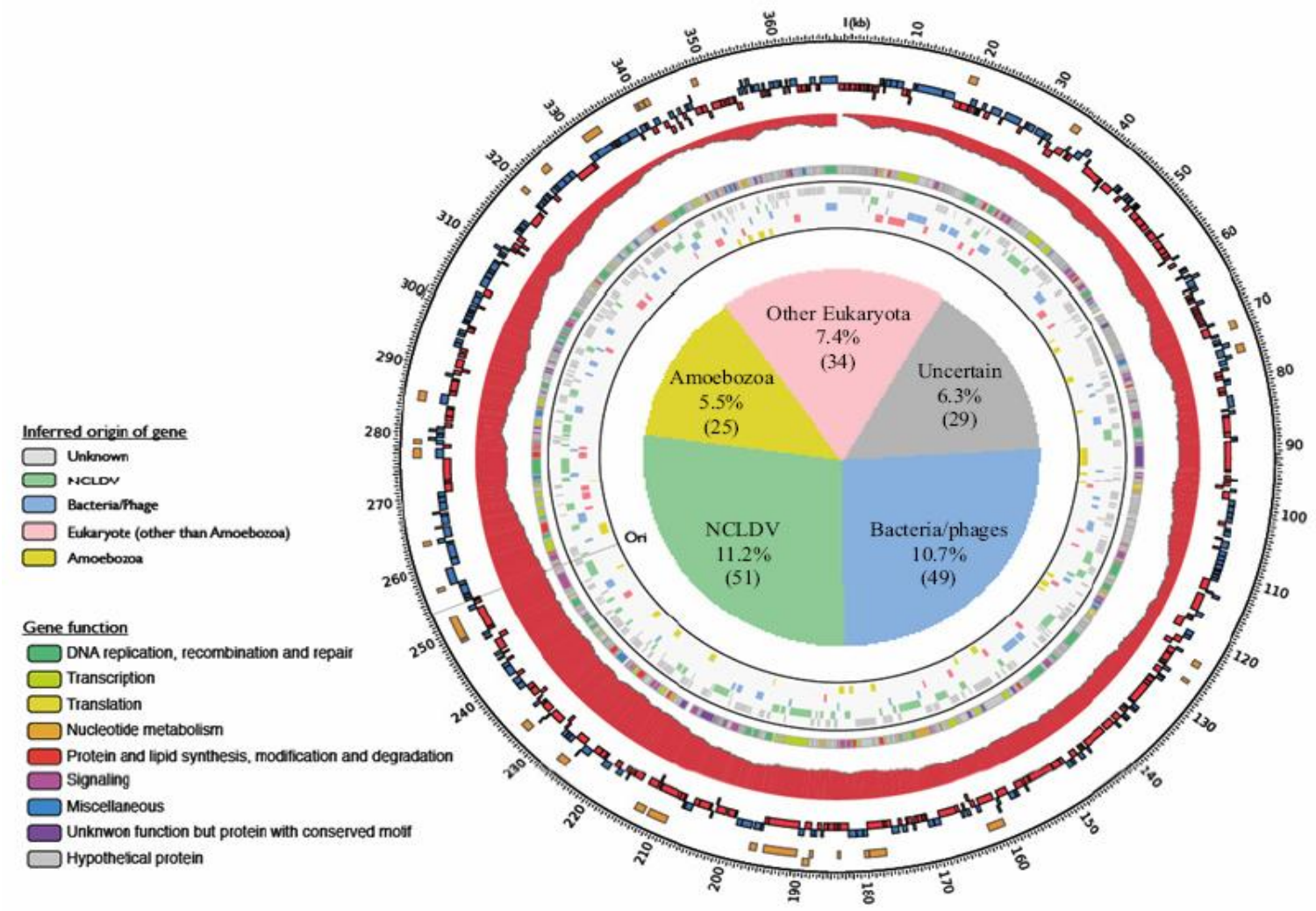


Figure 1.5: Map of the Marseillevirus chromosome, depicting the different variety of sources from which it has obtained its genetic material [*Boyer M et al, 2009*].

1.2.3 Cafeteria roenbergensis virus (CroV)

CroV was only quite recently isolated (2010), when it was discovered to infect the *Cafeteria roenbergensis* microorganism, a widespread marine microflagellate zooplankton grazer that is one of the oceans major and abundant predators. CroV is itself parasitised by a virophage named *Mavirus*. The microzooplankton *Cafeteria roenbergensis* is phylogenetically relatively distant from the amoeba hosts of the Mimivirus and the Marseillevirus. The CroV virus, however, is for some reason very closely related to the Mimivirus, although there is a huge phylogenetic distance and difference between their hosts and also less than a third of the CroV genes have been found to have homologs in the Mimivirus. Both these viruses have recently been classified within a new sub-family, the *Megaviridae*, belonging to the NCLDV group of viruses. It has the second largest genome, encompassing a 730 kb double-stranded DNA genome (618 kb represent the total protein coding genome) that includes 544 protein coding genes, many of which encode putative functions that are exceptionally unusual for a virus, maybe revealing the presence of a diverse coding potential (*Figure 1.6*). Nevertheless, it is the largest genome of any known marine virus and the most extraordinarily complex genome studied so far. As with most other giant viruses, the same is the case for CroV; a group of its genetic material has arisen due to gene duplication events, while another group is the result of HGT. [*Filee J & Chandler M, 2010*; *Fischer MG et al, 2010*; *Colson P et al, 2011*; *Van Etten JL, 2011*]. This giant virus is at its early stages of investigation, thus not much is known about its replication cycle.

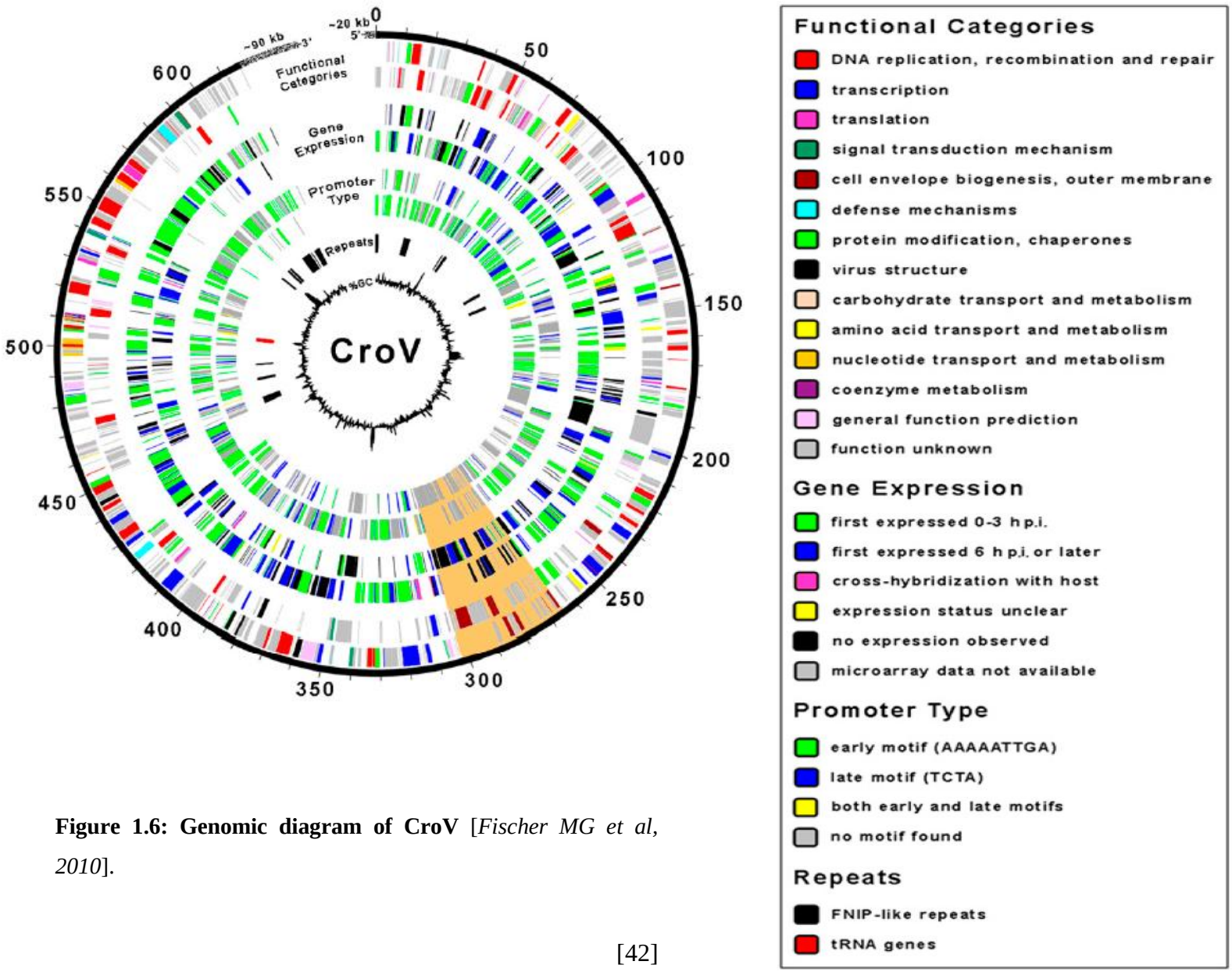


Figure 1.6: Genomic diagram of CroV [*Fischer MG et al, 2010*].

Some of the component factors included in the diverse coding potential of the CroV are DNA repair enzymes, DNA replication and promoter motifs, in addition to translation/ transcription factors. Furthermore, a 38 kb genomic region has been detected and has been characterised as being of 'a putative bacterial origin'. This coding sequence encoded several enzymes that were predicted to be involved in the carbohydrate metabolising pathway. This discovery is exceptionally interesting as nutrient recycling and carbon transfer, both in freshwater and marine environments, is majorly achieved by protistan gracers' predation (*Van Etten JL, 2011*). What is more, a significant number of genes have been shown to be expressed during CroV infection. All these facts indicate that CroV has a highly autonomous reproduction and propagation strategy during infection (*Fischer MG et al, 2010*).

1.3 Replication

DNA replication is regarded as the most fundamental process for all entities, as survival requires that replication of genomic material occurs in an extremely precise but also efficient way. For this purposes, it is a highly coordinated procedure engaging many proteins that work cooperatively towards correct DNA replication and hence accurate transmission of genetic information. All three known domains of life, *Archaea*, *Bacteria* and *Eukarya*, seem to replicate their genome in a similar way; in basic terms, they copy their genetic material and divide it into the next generation cells. [*Alberts BM, 1987*; *DePamphilis ML, 1993*; *Vas A & Leatherwood J, 2000*; *Bell SP & Dutta A, 2002*; *Mechali M, 2010*; *Kazlauskas D & Venclovas C, 2011*]. The textbook definition is: DNA replication is a

process during which the particular DNA molecule is duplicated in a semi-conservative way further resulting in the production of two identical DNA molecules that are eventually divided, hence forming new and identical progeny or daughter cells synthesised from that original cell (*Karp G, 2009*). As a result, each progeny cell contains an identical DNA molecule to that of the parental cell from where it came from. Consequently, *Archaea, Bacteria* and *Eukarya* grow and divide, whereas viruses do not follow this kind of process (*Cann AJ, 2003*).

1.3.1 Eukaryotic Replication

As previously mentioned, DNA replication is the accurate and timely duplication of the eukaryotic genome. For this process to be completed in a precise and successful way the cooperation of multiple factors and enzymes is required. These will further ensure that the genetic information will be maintained and stably passed down to the progeny cells each time the parental cells divide. This process is detrimental for the fate of the newly generated daughter cells. [*Bell SP & Dutta A, 2002; Kelly TJ & Spillman B, 2006*]

More specifically, in eukaryotes DNA synthesis is initiated by the orderly binding of initiator proteins in a series of steps to the origins of replication, found at multiple chromosomal sites (*Bell SP & Dutta A, 2002*). A multi-subunit protein called the origin recognition complex (ORC) initially recognises and binds specifically to autonomously replicating sequences (ARS) found within conserved replication initiation sites (origins)

(*Bell SP& Stillman B, 1992; Bryant JA et al, 2001; Shen Z & Prasanth SG, 2012*). This ORC complex particularly consists of six proteins, Orc1p-Orc6p, which are all essential for initiation and viability of DNA replication (*Stillman B, 1996; Kelly TJ & Spillman B, 2006; Bochman ML & Schwacha A, 2009 (Refer to Figure 1.7)*). In eukaryotes, the ORC complex forms the core of the origin complex to which other components are loaded onto the DNA replication fork in a step-wise manner (*Wang TA & Li JJ, 1995*).

The first step in the initiation of eukaryotic DNA replication is the assembly of the pre-replication complex (pre-RC), a multi-protein complex that controls where and when replication will initiate. The assembly of this complex specifically begins with ORC marking the origins of replication, loading onto them and thus recruiting two other factors, namely Cdc6 and Cdt1. The binding of both Cdc6 and Cdt1 proteins to the ORC complex is essential for the next step in DNA synthesis initiation, which is loading of the minichromosome maintenance (MCM) proteins (i.e. MCM2-7) onto chromatin (*Bryant JA et al, 2001*). As a result, the protein complex comprising of ORC, Cdc6/Cdt1 and MCMs forms the complete pre-RC, and this is established at the end of mitosis of the previous cell cycle after separation of sister chromatids (*Kelly TJ & Spillman B, 2006; Bochman ML & Schwacha A, 2009 (Refer to Figure 1.7); Shen Z & Prasanth SG, 2012*).

At the onset of S-phase, the pre-RC has to be converted into a pre-initiation complex (pre-IC) that leads to initiation of DNA synthesis by causing the initial denaturation of the double-helix and thus the formation of a replication 'bubble'. This activation of pre-RC is

accomplished by the action of S-phase specific cyclin-dependent kinases (CDKs) and Dbf4-dependent kinases (DDKs), which activate the firing of the replication origins allowing access and hence assembly of further replication factors (i.e. Dbp11, Sld3/Sld2, Mcm10, GINS complex, Cdc45, DNA polymerases etc.) to the exposed DNA template (*Gerbi S et al, 2002; Kelly TJ & Spillman B, 2006; Bochman ML & Schwacha A, 2009 (Refer to Figure 1.7); Shen Z & Prasanth SG, 2012*). The DNA double strand is specifically unwound by the action of a DNA helicase complex known as CMG (Cdc45-MCMs-GINS) complex (*Remus D & Diffley JFX, 2009; Zegerman P, 2013*), while this unwound DNA state is maintained by replication protein A (RPA) that binds single-stranded DNA and prevents it from winding back or forming other secondary structures (*Bambara RA et al, 1997; Hickey RJ et al, 2003; Chilkova O et al, 2007*).

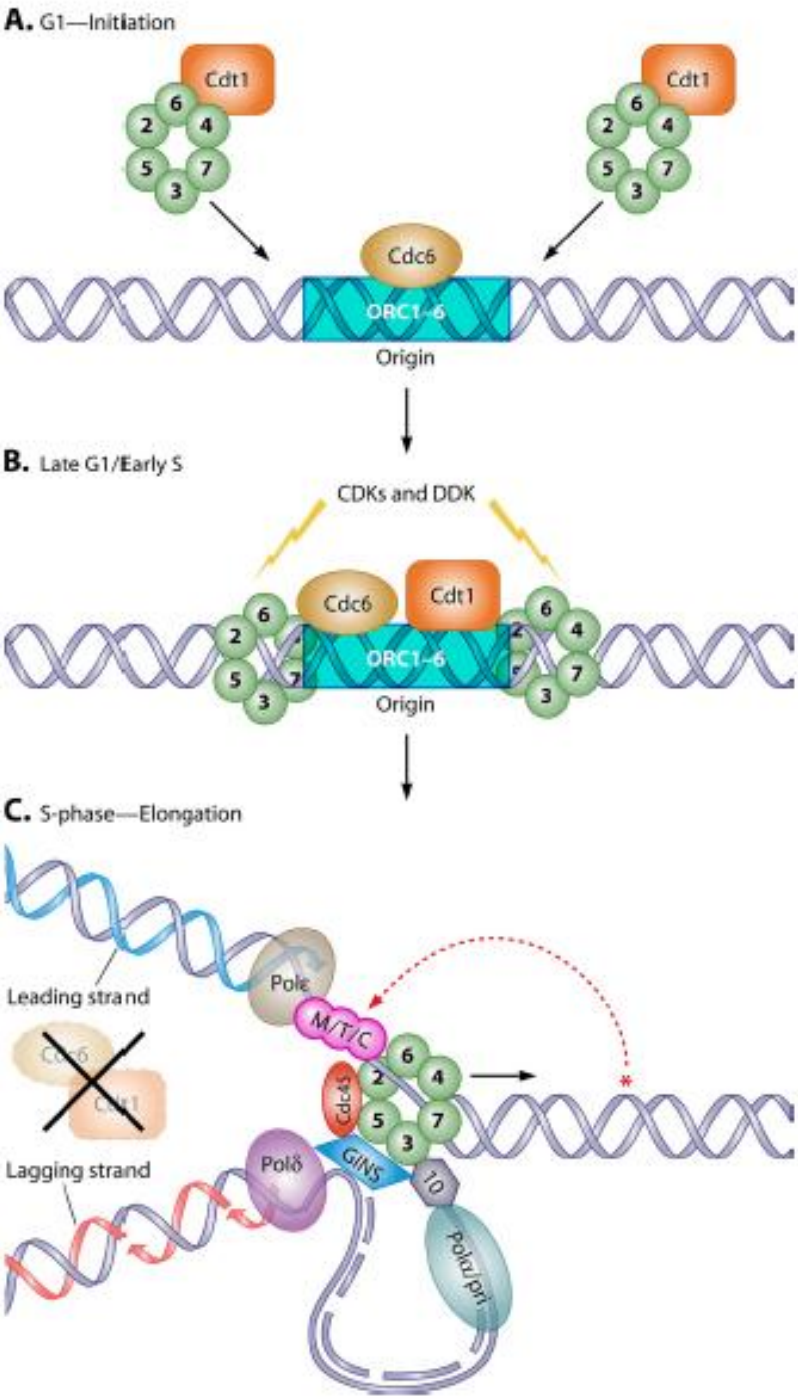


Figure 1.7: Initiation and early stages of eukaryotic DNA replication. "(A) During G1 phase, Cdc6 and Cdt1 recruit and load the MCMs complex (Mcm2-7) to origins of replication (marked by the binding of Orc1-6) to form a stable and inactive complex called the pre-RC. (B) In late G1/early S phase, the pre-RC is activated for DNA unwinding by the CDKs and DDKs, and is now named the pre-IC complex. This facilitates the loading of additional replication factors (e.g. Cdc45, Mcm10, GINS, polymerase α/primase and DNA polymerases δ and ε) and unwinding of the DNA at the origin. (C) During S phase, bidirectional DNA replication ensues." [*Bochman ML & Schwacha A, 2009*]

Once the parental DNA strands have been separated by the DNA helicase complex, as well as with the help of topoisomerases (Topoisomerase I) that remove DNA supercoils ahead of the replication fork, and the replication 'bubble' has been formed, then DNA synthesis can commence. However, only one of the original DNA strands is synthesised continuously, and this is termed the leading strand, while the other strand is produced in short discontinuous segments in the opposite orientation of the fork movement, and this is termed the lagging strand. In the latter case, the fragments generated are known as Okazaki fragments and these are specifically synthesised from a series of short RNA primers, which are eventually removed and the several DNA fragments produced are finally joined by DNA ligases (DNA ligase I) to complete lagging strand synthesis. In particular, lagging strand synthesis is initiated by multiple RNA primers that are generated by the intrinsic primase subunits of DNA polymerase α, while the polymerase subunit of the polymerase then adds a stretch of deoxyribonucleotides to the primer (*Bambara RA et al, 1997 (Refer to Figure 1.8); Takisawa H et al 2000; Kelly TJ & Spillman B, 2006*). The DNA polymerase α is also responsible for causing an initial priming event on the leading strand (*Mossi R et al, 2000; Hickey RJ et al, 2003; Chilkova O et al, 2007*).

Next step in eukaryotic DNA synthesis, for both the leading and the lagging strand, is the binding of the replication factor C (RFC) that will initiate polymerase switching in an ATP-dependent manner. This has two effects; first it causes the displacement of the DNA polymerase α-primase complex, and secondly it allows the assembly of the proliferating cell nuclear antigen (PCNA) that will form a sliding clamp structure and encircle the DNA

strand. The DNA polymerase δ (and ε) then comes along and interacts with the PCNA, ensuring the PCNA remains tightly bound to the DNA for processive synthesis (*Bambara RA et al, 1997 (Refer to Figure 1.7); Takisawa H et al 2000; Hickey RJ et al, 2003; Kelly TJ & Spillman B, 2006*). Experiments have shown that DNA polymerase δ is responsible for lagging strand synthesis, while DNA polymerase ε participates in leading strand synthesis (*Chilkova O et al, 2007; Pavlov YI & Shcherbakova PV, 2010*). This polymerase switching event occurs only once for the leading strand just after it has been primed, while for the lagging strand it happens during the synthesis of every Okazaki fragment (*Mossi R et al, 2000*).

In the final stages of the replication process, the initiator RNA primers are removed from the DNA strands by nucleases, commonly RNase H1 that has an endonucleolytic activity, while any remaining 5'-ribonucleotides are removed by the FEN1/RTH1 complex, containing both an endonucleolytic and exonucleolytic activity. The numerous single-stranded short DNA fragments generated from the lagging strand are finally ligated together to create a uniform DNA strand. The resulting daughter DNA strands intertwine to form a complete newly replicated double-helix (*Bambara RA et al, 1997 (Refer to Figure 1.8); Hickey RJ et al, 2003; Kelly TJ & Spillman B, 2006*).

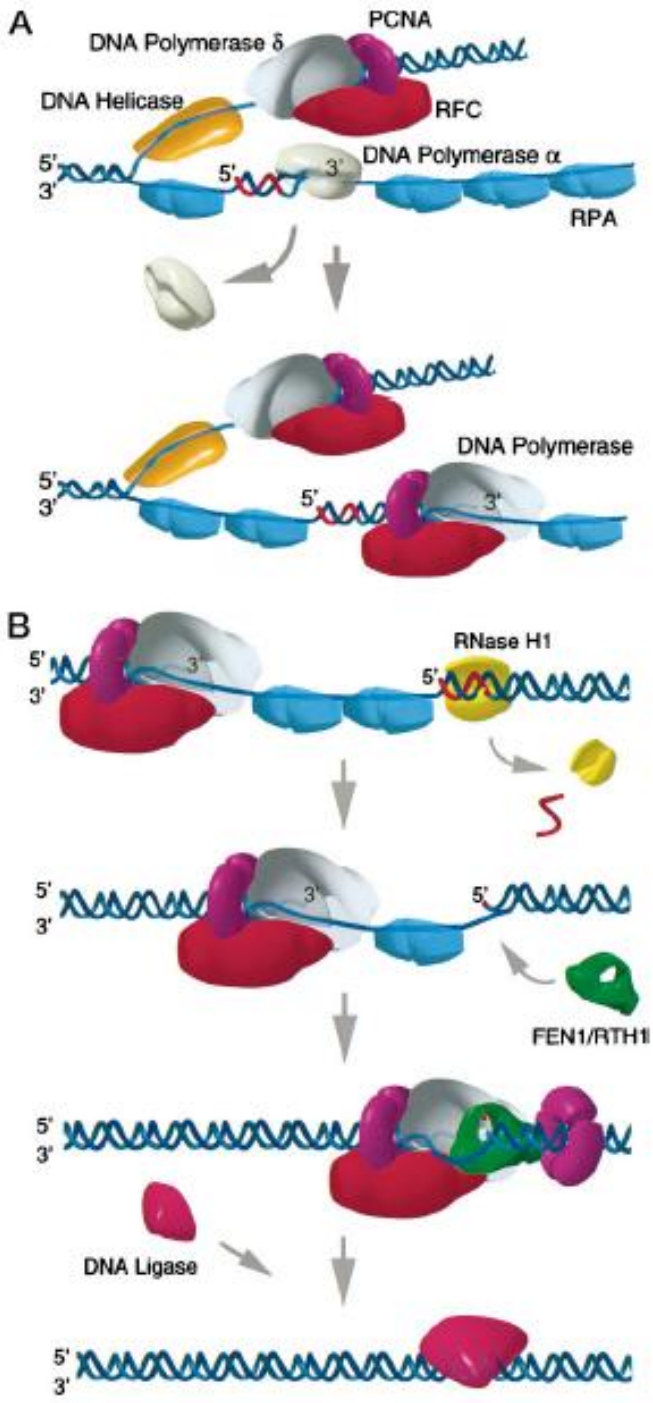


Figure 1.8: Mechanisms of DNA replication on the leading and lagging strands. "(A) As the DNA helicase promotes unwinding at the replication fork, DNA pol δ with RFC and PCNA synthesizes DNA on the leading strand. DNA pol α initiates synthesis on the lagging strand by generating an RNA primer (*red* segment) followed by a short segment of DNA. Then, RFC and PCNA load a second DNA polymerase (δ or ϵ) to continue synthesis of the Okazaki fragment. (B) As DNA pol δ approaches the downstream Okazaki fragment, cleavage by RNase H1 removes the initiator RNA primer leaving a single 5′-ribonucleotide. Then, FEN1/RTH1 removes the 5′-ribonucleotide. The resulting nick is sealed by DNA ligase." [*Bambara RA et al, 1997*]

In general, DNA replication is a very tightly controlled process for the reason that it should only occur once per cell-cycle, and therefore ensures that DNA is not re-replicated within

one cycle (*Wang TA & Li JJ, 1995; Bryant JA et al, 2001; Shen Z & Prasanth SG, 2012*). This restriction also ensures that DNA replication is completed only with minimal mistakes, with various damage and checkpoint controls having been evolved for the purpose of arresting or slowing down cell-cycle progression until the 'problem' encountered has been resolved by specific mechanisms (*Kelly TJ & Spillman B, 2006; Bochman ML & Schwacha A, 2009*).

1.3.2 Virus Replication

Viruses, as mentioned previously (*See Section 1.1; Viruses*), are considered to be obligate intracellular parasites that replicate by relying on their host to provide the majority of the material and machinery necessary for their reproduction. In other words, they manage in particular ways to force the infected host organism to produce more of the same viral components, hence more of the same viruses. [*Cann AJ, 2001*; *Minor PD, 2007*; *Harper DR, 2012*]. Therefore, their replication cycle differs from that of other living organisms, as it basically relies on the accurate assembly of already pre-formed viral components, whereas in the case of *archaea, bacteria* and *eukarya* these replicate by duplicating their genetic material and then dividing it into equal and identical parts (*Cann AJ, 2003*).

To do so, the viruses initially rely on their structural properties. The basic structure of a virus can be seen in *Figure 1.1*. In particular, during their replication cycle the viral component that plays the major role is the viral protein coat or envelope. The proteins of

the viral outer layer not only protect the viral genome from destruction, but they further enable the viruses to recognise, interact, and thus, infect the suitable host cells. Once this interaction has been established the viruses can then initiate a cycle that will eventually lead to the production of hundreds of identical viruses within a considerably short period of time. The steps that specifically follow after the viruses have identified, attached to and initiated infection of their hosts' cells are: a) penetration into the host cells that immediately causes an alteration in the host cells' cellular functions so as to support the time-dependent viral replication cycle; b) uncoating of the viral genetic material; c) replication and hence expression of the viral genome; d) assembly of the viral components produced and maturation of new generation virus particles; e) and finally, release of the mature virions from the parasitised host cell (*Cann AJ, 2003*; *Minor PD, 2007*; *Harper DR, 2011*). A summarised diagram of the typical virus replication cycle is illustrated in *Figure 1.2*.

1.3.3 Giant Virus Replication

The giant viruses under investigation all belong to the NCLDV group, as described previously (*See Section 1.2; Giant Viruses*). The NCLDV group of viruses is considered to be a monophyletic group, meaning that viruses belonging to this group are assumed to have emerged from a common ancestor (*Filee J et al, 2008; Colson P and Raoult D, 2010*). These viruses contain a linear, double-stranded DNA molecule and have a conserved core set of genes that are thought to play an important part in the viruses' metabolism, replication and propagation (*Filee J et al, 2006*; *Filee J & Chandler M, 2010*). As a result,

these viruses may not be so dependent on the host cells' material and machinery for their reproduction (*Yutin N & Koonin EV, 2009*; *Koonin EV & Yutin N, 2010*).

The main difference between other viruses and giant viruses is that the latter are essentially capable of replicating their genomic material entirely within the cytoplasm of the parasitised cell (as described by the name given to the NCLDV group), instead of only being able to do so in the nucleus (*Claverie JM et al, 2009-a/b*). Subsequently, they have two replication phases from which they can choose (i.e. one exclusively in the cytoplasm and one that may initiate in the nucleus before finalising the process in the cytoplasm), compared to only having one replication phase like normal viruses do (i.e. in the nucleus) (*Filee J et al, 2006; Yutin N & Koonin EV, 2009*; *Mutsafi Y et al, 2010*). This fact is possibly sufficient to disengage the giant viruses' replication and propagation activities from the hosts' genome.

In general terms, giant virus replication occurs in the following way (as has been already described for the Mimivirus): after endocytosis is initiated, the genomic material is released into the cytoplasm of the host cell causing an early initiation of transcription. This results in the production of mRNAs, which are transported to isolated sites within the cytoplasm. Next, DNA replication is commenced within specific replication factories generated for this particular purpose. These factories have been seen in areas of the host cytoplasm that are completely opposite to the areas where the mRNAs have accumulated, while the number of these factories generated depends on the starting number of infecting

virus particles. Therefore, when the individual replication factories expand due to extensive replication, they then fuse to form a single large factory taking up most of the hosts' cytoplasm. Eventually, the huge single factory will burst and release the newly generated virus particles. [*Katsafanas GC & Moss B, 2007; Suzan-Monti M et al, 2007; Mutsafi Y et al, 2010*; *Fischer MG, 2012*].

Even though an entirely cytoplasmic replication cycle has been suggested for some giant viruses (i.e. Poxvirus, Mimivirus), this does not imply in any way an exclusively nucleus-independent process. The entirely cytoplasmic replication cycle does indeed provide proof of the fact that these giant viruses encode an important number of proteins essential for their own DNA replication and transcription mechanisms, as well as mRNA synthesis. Nevertheless, during these processes there will still be a participation of host-encoded protein factors. That being either due to the fact that, even though these viruses have a huge genome, they are still not capable of encoding all the necessary machinery for their complete replication, or due to some host nuclear factors being passively leaked out of the nucleus during viral replication. In the first case, the host proteins are actively exported from the nucleus and imported into the cytoplasm for association with the virus-encoded proteins and participation in the viral replication processes. This active transportation will take place as long as the host-encoded proteins found in the nucleus are not anchored to a nuclear structure (i.e. nuclear membrane), in which case they probably become resistant to cytoplasmic delivery. [*Oh J & Broyles SS, 2005; Mutsafi Y et al, 2010*]. However, the exact

requirements that are fundamental for this export-import process of host nuclear protein factors from the nucleus into the cytoplasm are not yet fully understood.

1.4 Replication Components

For efficient and complete DNA replication, taking place either in *Archaea, Bacteria, Eukarya* or viruses, specific DNA replication proteins are required. The eukaryotic replication proteins have already been discussed in detail in *Section 1.3.1*. In most cases, all organisms necessitate the presence of the same or equivalent proteins for their reproduction. [*Leipe DD et al, 1999; Forterre P et al, 2000; Robinson NP & Bell SD, 2005; Barry ER & Bell SD, 2006; Aves SJ, 2009; Boyer M et al, 2010; Yutin N & Koonin EV, 2012*]. There are a great number of such diverse DNA replication proteins, but for the purpose of this project we will specifically focus our interest on the PCNA and FEN1 proteins.

1.4.1 Proliferating Cell Nuclear Antigen (PCNA), a sliding clamp protein

DNA replication is performed by a multicomponent complex of proteins known as DNA replicases. This complex typically contains a DNA sliding clamp, which is the central factor for DNA replication processes, the clamp loader, and of course, a DNA polymerase. The sliding clamp is a ring-shaped polymerase processivity factor, which is loaded onto the double-stranded DNA with the help of the clamp loader. The interaction between the sliding clamp (e.g. PCNA) and DNA is rendered stronger in the presence of positively charged residues located in the center of the circular PCNA molecule that create a suitable

DNA binding surface. [*McNally R et al, 2010*; *Fig. 1.7*]. By encircling the DNA molecule, the sliding clamp forms a platform and permits other replication factors, such as the DNA polymerases, to assemble and engage at the heart of the replication folk, hence allowing the initiation of the replication process (*Kirchmaier AL, 2011*). What is more, by tethering the various replication factors to the DNA molecule, it is able to augment their activity (*Warbrick E, 1998; Maga G & Hubscher U, 2003; Kazlauskas D & Venclovas C, 2011*).

In addition, a vast array of other factors involved in DNA processing, such as DNA modulating and damage by-pass/repair enzymes, cell cycle regulators, as well as other enzymes that play an active part in chromatin assembly, cohesion and remodeling, bind to and interact with the sliding clamp in a highly dynamic and coordinated fashion. In the case of PCNA's, such interactions usually involve the hydrophobic cleft of the PCNA interdomain connector loop and the PCNA-interacting peptide (PIP) box motif found on the interacting partner (*De Biasio A et al, 2012*; *Ulrich HD & Takahashi T, 2013*). This mechanism illustrates the crucial role of sliding clamps in controlling access to the DNA and to its machinery, by regulating and coordinating the function of a plethora of other enzymes (*Winter JA & Bunting KA, 2012*). This fine and complex interplay between sliding clamps and their effector proteins, at different stages of DNA replication and repair, has to be tightly controlled by a series of regulatory mechanisms; a) differential binding affinities of each protein to the sliding clamps leading to association or disassociation of one interacting partner by another, b) post-translational modifications [*Ulrich HD & Takahashi T, 2013*], c) accessory factors that modulate these interactions, d) appropriate destruction of

the complexes formed by recruiting other necessary proteins, when required. As such, sliding clamps play an essential role in maintaining the genome integrity and stability (*Kirchmaier AL, 2011*; *Mailand N et al, 2013*).

A schematic representation of a characteristic PCNA/DNA binding interaction is depicted in *Figure 1.9*.

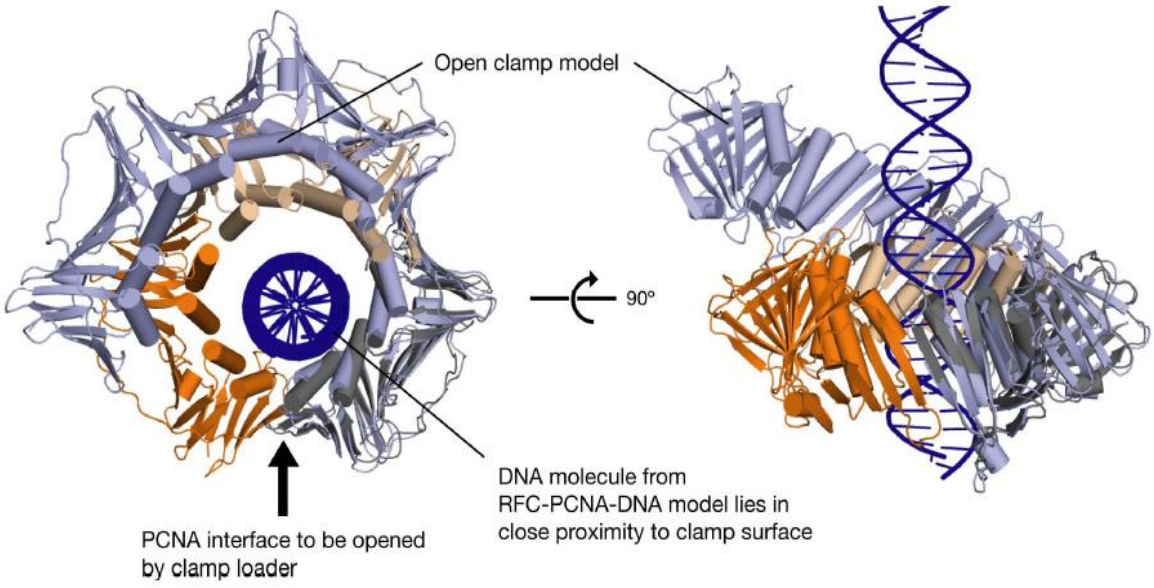


Figure 1.9: Standard PCNA/DNA interaction model [*McNally R et al, 2010*].

In general, in the case of the PCNA sliding clamp; During DNA replication and repair, the PCNA protein is loaded onto the DNA template by the RFC clamp-loading complex. Once loaded, PCNA initially interacts and enhances the activity of the DNA polymerase Pol δ and Pol ε. In addition to these enzymes, however, PCNA acts as a platform for the direct

binding of other proteins involved in DNA synthesis and repair, ranging from the Flap Endonuclease I (FEN1) and DNA ligase I to other DNA polymerases or DNA damage repair proteins (e.g. DNA Pol η). Other DNA processes require the assembly of different protein that will bind to and interact with the PCNA protein (*Majka J & Burgers PMJ, 2004*; *Kirchmaier AL, 2011*; *Mailand N et al, 2013*).

Computation analysis of the various components of the DNA replicase complex revealed that, in spite of DNA replication being a uniformal procedure between *archaea, bacteria, eukarya* and viruses, some of the components involved are not universally conserved. This was mostly the case for the DNA polymerase proteins, which were shown to have evolved independently from different ancestral proteins for *bacteria* and *archaea/eukarya* (*Kazlauskas D & Venclovas C, 2011*). In addition, the sliding clamp family is also extremely divergent in terms of their amino acid sequence, revealing no sequence homology between the diverse sliding clamps belong to the different organisms. The only fact that shows homology between all DNA processivity factors is their three dimensional ring-shaped structure (*O'Reilly DR et al, 1989; Bruck I & O'Donnell M, 2001*).

Though the bacteria DNA sliding clamp was discovered to be a homodimer, the archaeal and eukaryotic sliding clamps, known as Proliferating Cell Nuclear Antigen (PCNA), were mostly observed as homotrimers (*De Biasio A et al, 2012*). Some archaea may also have a heterotrimeric PCNA. In the case of viruses, they all require the presence of processivity factors for successful viral DNA replication, however only a number of them encode for

PCNA-like proteins and this fact is basically dependent on their genome size. However, the different viruses, in which PCNA-like proteins have been studied, reveal different forms of this protein (*Kazlauskas D & Venclovas C, 2011*).

Concerning the giant viruses; in the Mimivirus three varying PCNAs have been identified (these are referred to as MIMI_L108, MIMI_L823 and MIMI_R493; *obtained from Uniprot, and in accordance with work conducted by Raoult D et al, 2004*), but it is still not known if these altogether form a heterotrimer or if the individual proteins form homotrimers, homodimers, heterodimers or even monomers. In the case of the Marseillevirus the identified PCNA protein has been named MAR_ORF212, while for the CroV the identified PCNA protein is known as CroV_219 (*information obtained from Uniprot*). Unfortunately, not much more is known about these giant viruses and their DNA replicase complexes. Therefore, as this group of viruses appears to be of great significance due to their spectacular way of life, it is crucial to further investigate in depth exactly how they acquired such an enormity, and as such, precisely how they replicate their genomic material.

As a final note; computation analysis studies surprisingly revealed that as the size of the organisms' (including viruses) genome increases, then by rule they will encode for their own DNA replicase components (*Kazlauskas D & Venclovas C, 2011*). In the case of giant viruses this explains why most possibly they do not have to rely on the infected host to provide them with DNA replicase factors for their reproduction.

1.4.2 Flap Structure-specific Endonuclease 1 (FEN1) protein

All organisms, *Archaea, Bacteria, Eukarya* and viruses, require the action of a nuclease protein that will assist in DNA repair following DNA replication. Nucleases are key enzymes for controlling replication, taking part in repair processes, as well as multiple other metabolic pathways, and as such, maintaining the genome stability. One such enzyme is the FEN1 protein that plays a major role during DNA replication procedures, where it effectively remove the 5' overhanging ends from the Okazaki fragments generated during double-stranded DNA synthesis by the DNA polymerase enzyme, hence forming a substrate for the DNA ligase enzyme to assist Okazaki fragment maturation (*Warbrick E, 1998; Gomes XV & Burgers PMJ, 2000*; *Sakurai S et al, 2005*). Its activity in this type of process is based on the recognition of the bifurcated ends of the double-stranded DNA, and its specific action is to cut the phosphodiester bond at the 5' prime end (firstly, exonucleolytically removes the ribonucleotide, and secondly, endonycleolytically removes the entire primer) leaving a 3' hydroxyl end (*Kaiser MW et al, 1999*; *Rumbaugh JA et al, 1999*). To complete accurately its action, FEN1 interacts with particular proteins. A schematic diagram of the representative FEN1 activity is illustrated in *Figure 1.10*.

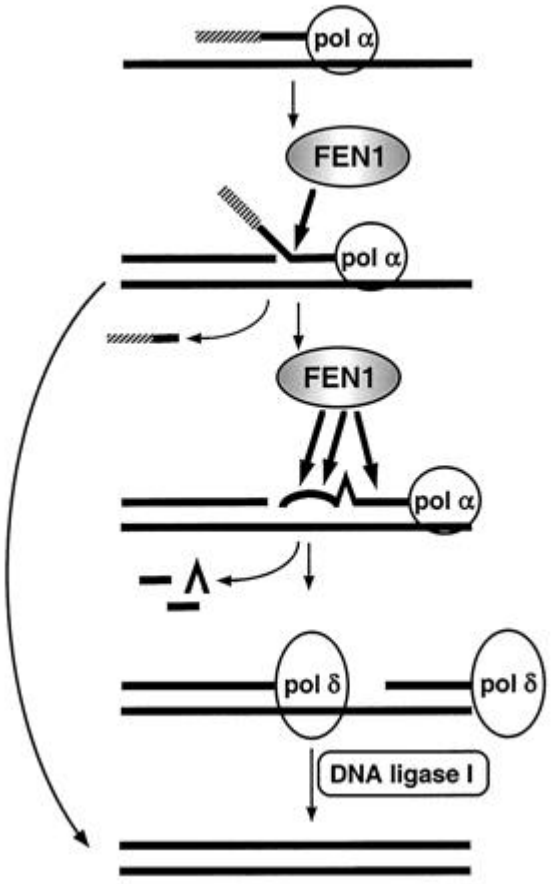


Figure 1.10: Model for repair of DNA strands and completion of Okazaki fragment maturation. "First, the DNA polymerase complex makes a RNA primer (*hatched line*) and begins DNA synthesis, and then FEN1 removes the initiator RNA, perhaps including some DNA as well, with or without extension from an upstream fragment. Meanwhile, DNA polymerase continues DNA synthesis and inserts a mismatch that slightly disrupts the DNA helix. This disruption promotes the removal of the mismatch by FEN1 in one cut or a series of endonucleolytic cuts, depending on the location of the mismatch**"**. Eventually, ligation of the DNA fragments is completed [*Rumbaugh JA et al, 1999*].

In the case of *Bacteria*, their DNA polymerase protein itself contains a 5' exonuclease domain. On the other hand, *Archaea* and *Eukarya* encode for a FEN1 protein. Both these nuclease enzymes are members of a family of 'structure-specific 5' exonucleases', and even though they perform fairly similar functions, they have a very limited similarity between their sequences. Despite this fact, they all have the same substrate specificity. [*Kaiser MW et al, 1999*; *Grabowski B & Kelman Z, 2003*]. Moreover, viruses are believed to encode the FEN1 type nuclease protein (or a FEN1-like protein), but not much is known about this protein in the diverse virus families and groups, as minimum investigation has been carried out over the years. As a consequence, very little is mentioned in the literature

regarding the nuclease activity present in the various giant virus families [*Da Silva M et al, 2006; Iyer LM et al, 2006; Senkevich TG et al, 2009; Yoshida T et al, 2011*]; so even though the Mimivirus, Marseillevirus and CroV have been determined to have a FEN1-like endonuclease protein (*according to the UniProt database; MIMI_L386, MAR_ORF365, CroV_037, respectively*), no more details have been acknowledged about this protein in the three viral organisms; the only slight exception being CroV for which the Flap (FEN1)-like endonuclease has been identified as a probable XPG nuclease (*Fischer MG et al, 2010*).

1.4.3 Additional DNA replication proteins encoded by the three Giant viruses of interest, Mimivirus, Marseillevirus and CroV

Most information available to date about the proteins encoded by giant viruses has been based on computational/bioinformatic studies and analysis of their viral genome sequences. In particular, various molecular sequence analyses of the Mimivirus, Marseillevirus and CroV genomes has revealed that these giant viruses encode homologous DNA replication proteins, a fact that appears to be dependent on their genome size. The key replication proteins found to be conserved in these viruses, actually form a core set of proteins found to be distributed in all giant NCLDVs (*Kazlauskas D & Venclovas C, 2011*). An early study performed by *Iyer LM et al (2006)* found a number of these conserved proteins involved in viral DNA replication. These include a shared Ser/Thr kinase, a D5R-like replicative primase/helicase, as well as other helicases, Topoisomerases, a DNA polymerase of the B family, a PCNA-like DNA clamp, RFC clamp loaders, ATP (Adenosine TriPhosphate)- and NAD (Nicotinamide Adenine

Dinucleotide)-dependent DNA ligases, Exonucleases, a Flap (FEN1)-like endonuclease, as well as additional endonucleases, and a RuvC-like Holiday junction resolvase (HJS).

The initial studies performed for the purpose of deciphering the genetic content of these giant viruses, together with subsequent studies, were conclusive. In particular, concerning the three most vital DNA replication proteins (i.e. DNA polymerases, PCNA and RCF proteins), the outcomes from these studies were as follows: All three giant viruses, Mimivirus, Marseillevirus and CroV, were revealed to encode for a Family B DNA polymerase, while only Crov further encodes for a Family X DNA polymerase. In addition, they all encode for PCNA sliding clamps, with Mimivirus actually encoding for three separate PCNA proteins. With the exception of Marseillevirus, both Mimivirus and CroV were found to encode for all five RFC subunits, hence probably have a fully functional RFC protein. Studies performed by *Kazlauskas D & Venclovas C* (*2011*) showed that only three out of the five subunits (i.e. RFC1, RFC3, and RFC5) of the RFC protein complex, belonging to the Mimivirus and CroV, contain a PIP-box for interaction with the PCNA protein, and the affinities of each subunit for this particular interaction vary probably due to differential evolution. In Mimivirus, specifically, the PIP-boxes belonging to each of the RFC1, RFC3 and RFC5 subunits have progressively 'weaker' PCNA-binding strengths. However, the actual PIP-box in RFC5 is mostly similar to the one identified in RFC1. For a summary of the DNA replication proteins discussed and the organisms they are encoded by see *Table 1.1*. [*Raoult D et al, 2004; Boyer M et al, 2009 & 2010; Colson P et al, 2010; Fischer MG et al, 2010; Kazlauskas D & Venclovas C, 2011*].

What is more, all viruses in question encode for various kinases. The most important ones were determined as being Thymidine and Serine/Threonine kinases. These, or equivalent enzymes, are encoded by all three giant viruses of interest; Marseillevirus and CroV encode for those exact enzymes, while in the case of Mimivirus the Thymidine kinase has been replaced by a Deoxynucleoside Kinase (DNK) and a Nucleoside Diphosphate Kinase (NDK). In regards to the central DNA primases, the Mimivirus, Marseillevirus and CroV primase proteins have been shown to be linked to and work in conjunction with the D5-like helicases. Interestingly, Marseillevirus additionally encodes for an AEP (Archaeo-Eukaryotic Primase)-type primase. For a summary see *Table 1.1*. [*Raoult D et al, 2004; Boyer M et al, 2009 & 2010; Colson P et al, 2010; Fischer MG et al, 2010; Kazlauskas D & Venclovas C, 2011*].

Other essential DNA replication proteins, such as DNA helicases, topoisomerases, ligases and nucleases, are abundant in all three giant viruses of interest. Amongst others, SF (SuperFamily)-type helicases are very common and present in all three viruses [i.e. SW1/SNF2 (SWItch/Sucrose NonFermentable) ATPase helicase encoded by Mimivirus and Marseillevirus, Types I and II encoded by Marseillevirus, and only Type II encoded by CroV], while only Mimivirus and Marseillevirus moreover encode for a D6R-type helicase. In terms of the topoisomerase proteins present in these viruses, both Mimivirus and CroV encode for Types IA, IB and IIA, while Marseillevirus only encodes for Type II. (*See Table 1.1)* [*Raoult D et al, 2004; Boyer M et al, 2009 & 2010; Colson P et al, 2010; Fischer MG et al, 2010; Kazlauskas D & Venclovas C, 2011*].

Regarding the repair of the DNA strands during replication, the most common types of DNA ligases present in giant viruses are the ATP-dependent DNA ligase and the NAD-dependent DNA ligase. Both Mimivirus and Crov have the later type, while Marseillevirus encodes for the former type. Furthermore, all three encode for 5'-3' exonucleases and RNases [i.e. Mimivirus and Marseillevirus were found to encode RNases Types III and HI, whereas CroV encodes for Types H and HI], while they have homologs of Flap (FEN1)-like endonucleases. Finally, all three viruses encode for a RuvC-like HJS. (*See Table 1.1)* [*Raoult D et al, 2004; Boyer M et al, 2009 & 2010; Colson P et al, 2010; Fischer MG et al, 2010; Kazlauskas D & Venclovas C, 2011*].

	Mimivirus	**Marseillevirus**	**CroV**
Kinases	- DNK, - NDK, - Ser/Thr	- Thymidine, - Ser/Thr	- Thymidine, - Ser/Thr
Helicases	- Type III (D5-type ATPase), - D6R-type, -SW1/SNF2	- SF-types I and II, - D6R-type, - SW1/SNF2	- SF-type II
Topoisomerases	Types IA, IB and IIA	Type II	Types IA, IB and IIA
Primases	D5-like primase/helicase	- D5-like primase/helicase, - AEP-type	D5-like primase/helicase
DNA Polymerases	Family B	Family B	- Family B, - Family X
RFC protein	Five RFC subunits	-	Five RFC subunits
PCNA protein	Three PCNAs	+	+
DNA ligases	NAD-dependent	ATP-dependent	NAD-dependent
RNases	Types III and HI	Types III and HI	Types H and HI
Exo-/ Endonucleases	- 5'-3' exonuclease, -Flap (FEN1)-like endonuclease	- 5'-3' exonuclease, - Flap (FEN1)-like endonuclease,	- 5'-3' exonuclease, - XPG (Flap-like) endonuclease,
Resolvases	RuvC-like HJS	RuvC-like HJS	RuvC-like HJS

Table 1.1: **Summary of the key DNA replication proteins encoded by Mimivirus, Marseillevirus and CroV.** (The symbol '+' indicates the presence of a protein, whereas '-' indicates its absence).

A number of the key DNA replication proteins belonging to the giant viruses, Mimivirus, Marseillevirus and CroV, as discussed in this section, were subsequently compared, in terms of their sequence similarity, to the equivalent eukaryotic proteins. In particular, the proteins selected for this investigation were the PCNA, RFC and FEN1 proteins, while the equivalent proteins from two eukaryotic organisms were chosen for comparison reasons; these organisms were homo sapiens and the most extensively studied yeast, *Saccharomyces cerevisiae*.

The three proteins of interest discovered in the various organisms selected for this study are demonstrated in *Table 1.2*. The details of each protein entry are in accordance with the information described on the Uniprot website (http://www.uniprot.org/), while at instances the NCBI protein database was also referred to (http://www.ncbi.nlm.nih.gov/protein).

	PCNA	**RFC**	**FEN1**
Homo sapiens	PCNA_HUMAN, 261aa, Accession No. P12004	RFC subunit 1: RFC1_HUMAN, 1148aa, Accession No. P35251, RFC subunit 2: RFC2_HUMAN, 354aa, Accession No. P35250, RFC subunit 3: RFC3_HUMAN, 356aa, Accession No. P40938, RFC subunit 4: RFC4_HUMAN, 363aa, Accession No. P35249, RFC subunit 5: RFC5_HUMAN, 340aa, Accession No. P40937	FEN1_ HUMAN, 380aa, Accession No. P39748

S. cerevisiae	PCNA_YEAST, 258aa, Accession No. P15873	RFC subunit 1: RFC1_YEAST, 861aa, Accession No. P38630, RFC subunit 2: RFC2_YEAST, 353aa, Accession No. P40348, RFC subunit 3: RFC3_YEAST, 340aa, Accession No. P38629, RFC subunit 4: RFC4_YEAST, 323aa, Accession No. P40339, RFC subunit 5: RFC5_YEAST, 354aa, Accession No. P38251	FEN1_ YEAST, 382aa, Accession No. P26793
Mimivirus	1. PCNA_ MIMIV, 464aa, Accession No. Q7T6Y0 2.YL108_ MIMIV, 273aa, Accession No. Q5UPJ0 (*NCBI Reference: YP_ 003986598.1*), 3. YL823_ MIMIV, 323aa, Accession No. Q5UQH4 (*NCBI Reference: YP_ 003987355.1*)	RFC large subunit: RFCL_MIMIV, 533aa, Accession No. Q5UQK9, RFC small subunit 1: RFCS1_MIMIV, 363aa, Accession No. Q5UQ72, RFC small subunit 2: RFCS2_MIMIV, 344aa, Accession No. Q5UP47, RFC small subunit 3: RFCS3_MIMIV, 319aa, Accession No. Q5UQ47, RFC small subunit 4: RFCS4_MIMIV, 370aa, Accession No. Q5UQE8. *According to *Kazlauskas D & Venclovas C, 2011* paper, the entry names of the small RFC subunits on Uniprot for Mimivirus have been recorded incorrectly. Here are presented the corrected versions.	Putative endonuclease YL386_ MIMIV, 473aa, Accession No. Q5UQW7
Marseillevirus	D2XAL4_ 9VIRU, 298aa, Accession No. D2XAL4	-	Flap-specific endonuclease D2XB04_ 9VIRU, 362aa, Accession No. D2XB04

CroV	E3T4Y9_9VIRU, 278aa, Accession No. E3T4Y9	1. E3T5A4_9VIRU, 429aa, Accession No. E3T5A4 (*not fully characterised protein*), Putative RFCs (*NCBI*): 2. Ref: YP_003970094.1, 334aa 3. Ref: YP_003969962.1, 316aa	Putative DNA endo-nuclease E3T4F7_ 9VIRU, 321aa, Accession No. E3T4F7

Table 1.2: **The PCNA, RFC and FEN1 proteins as characterised in a selection of eukaryotic and viral organisms.** Information acquired from the Uniprot database, as well as on some occasions from the NCBI website.

The established protein sequences for each protein mentioned in *Table 1.2* were obtained from Uniprot (or NCBI when required), and these were utilised in three individual protein sequence alignments. Multiple sequence alignments were specifically performed separately for the PCNA, RFC and FEN1 proteins, using either the Clustal Omega (for small scale alignments) or the MUSCLE (for big scale alignments, as it is pronounced to achieve 'better average accuracy' compared to other tools) protein alignment software on the EMBL-EBI website (http://www.ebi.ac.uk/Tools/msa/clustalo/ and http://www.ebi.ac.uk/Tools/msa/muscle/, respectively). The corresponding results for the PCNA, RFC and FEN1 proteins were finally visualised using the Jalview 2.8 software program and are illustrated in *Figures 1.11*, *1.12* and *1.13*, respectively.

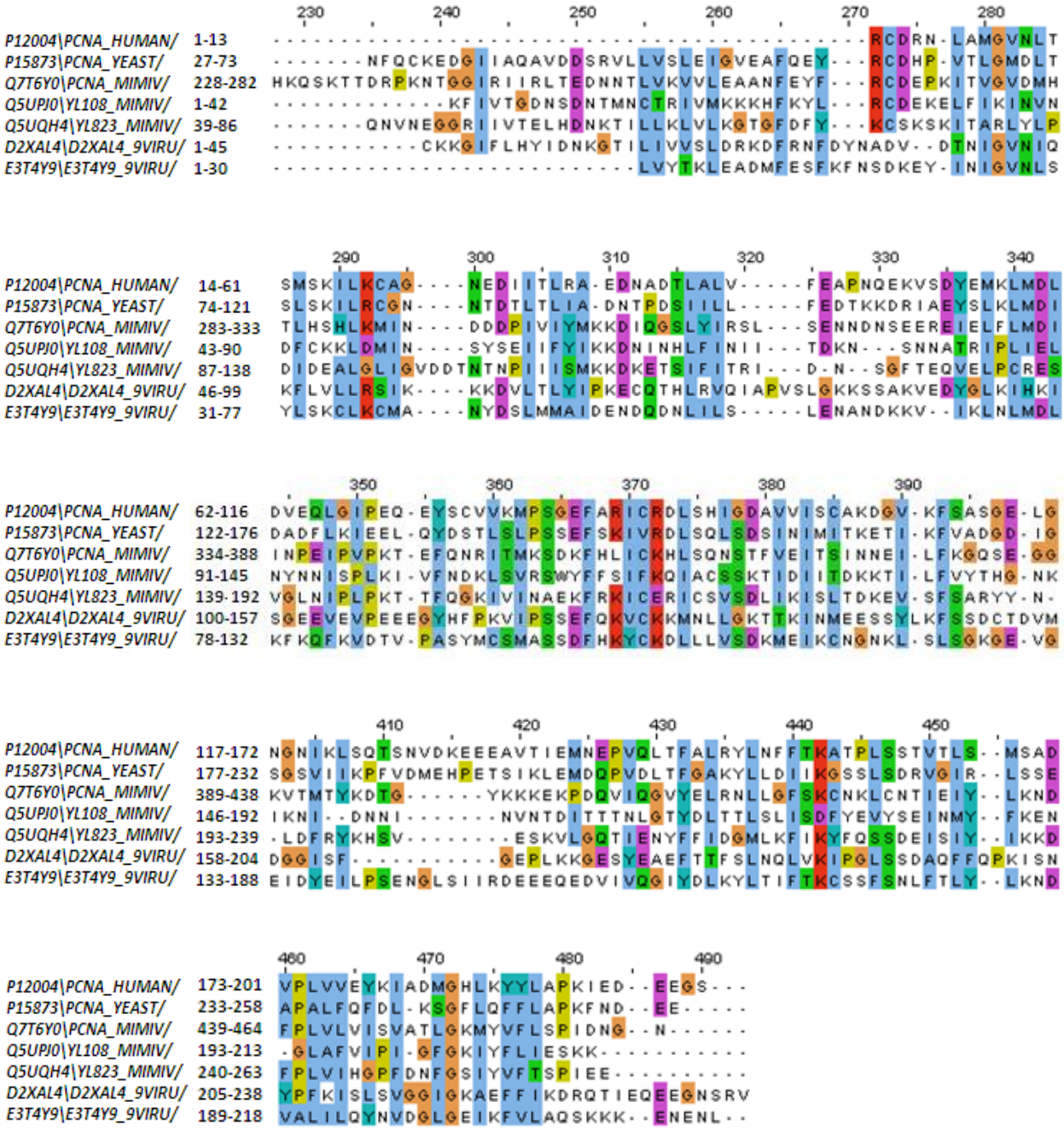
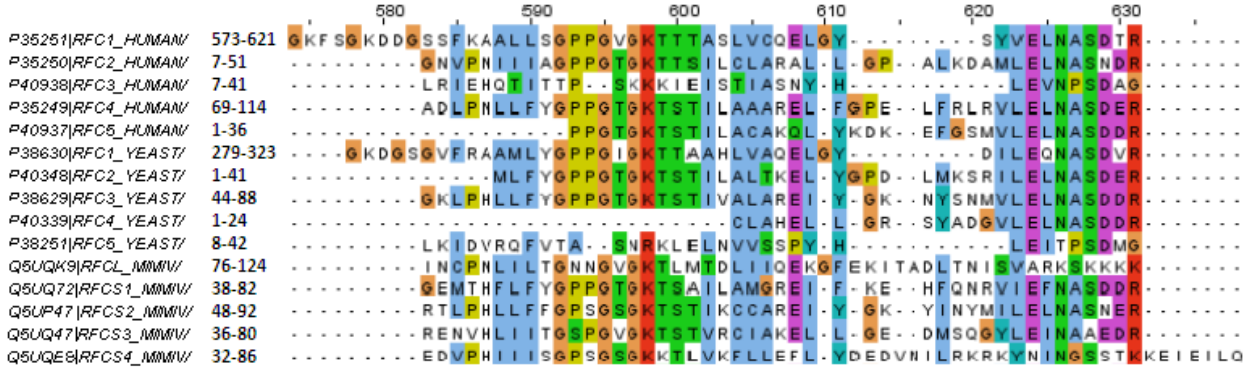
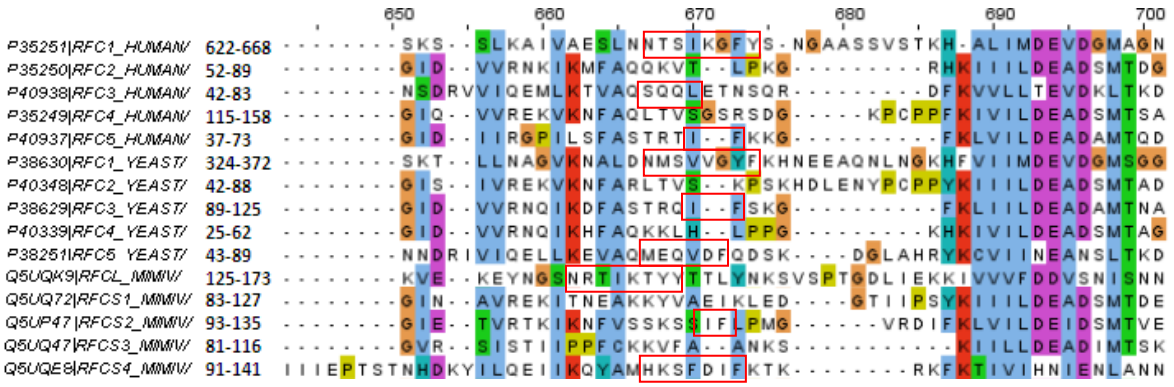
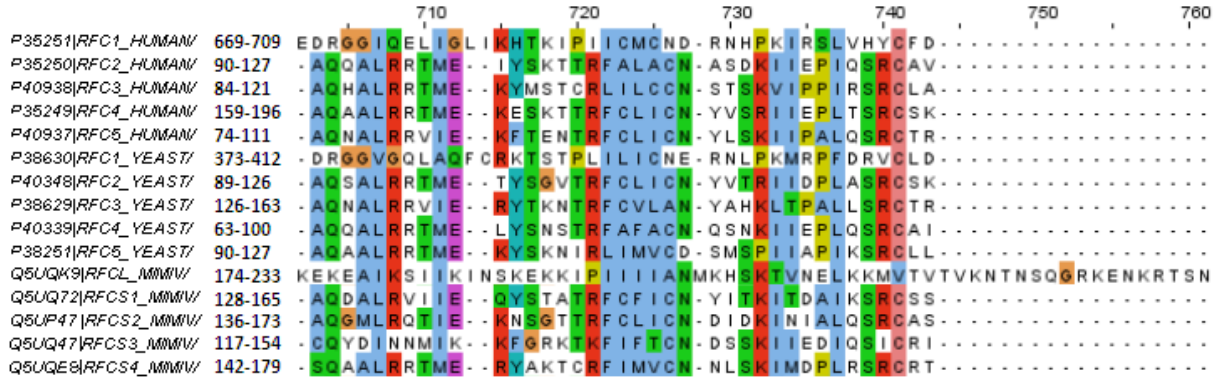


Figure 1.11: Multiple sequence alignment of PCNA proteins belonging to two eukaryotic organisms, Homo sapiens and *S. cerevisiae,* and the three giant viruses of interest, Mimivirus, Marseillevirus and CroV. The protein alignments were performed using the ClustalO 1.1.0 bioinformatics software program, while the results were visualised using Jalview 2.8. The ClustalX colour scheme was applied to conserved residues according to specific criteria set by the software (http://www.jalview.org/help/html/colourSchemes/clustal.html).


580
590
600
610
620
630
P35251|RFC1_HUMAN/ 573-621
P35250|RFC2_HUMAN/ 7-51
P40938|RFC3_HUMAN/ 7-41
P35249|RFC4_HUMAN/ 69-114
P40937|RFC5_HUMAN/ 1-36
P38630|RFC1_YEAST/ 279-323
P40348|RFC2_YEAST/ 1-41
P38629|RFC3_YEAST/ 44-88
P40339|RFC4_YEAST/ 1-24
P38251|RFC5_YEAST/ 8-42
Q5UQK9|RFCL_MIMIV/ 76-124
Q5UQ72|RFCS1_MIMIV/ 38-82
Q5UP47|RFCS2_MIMIV/ 48-92
Q5UQ47|RFCS3_MIMIV/ 36-80
Q5UQE8|RFCS4_MIMIV/ 32-86

PIP-box


650
660
670
680
690
700
P35251|RFC1_HUMAN/ 622-668
P35250|RFC2_HUMAN/ 52-89
P40938|RFC3_HUMAN/ 42-83
P35249|RFC4_HUMAN/ 115-158
P40937|RFC5_HUMAN/ 37-73
P38630|RFC1_YEAST/ 324-372
P40348|RFC2_YEAST/ 42-88
P38629|RFC3_YEAST/ 89-125
P40339|RFC4_YEAST/ 25-62
P38251|RFC5_YEAST/ 43-89
Q5UQK9|RFCL_MIMIV/ 125-173
Q5UQ72|RFCS1_MIMIV/ 83-127
Q5UP47|RFCS2_MIMIV/ 93-135
Q5UQ47|RFCS3_MIMIV/ 81-116
Q5UQE8|RFCS4_MIMIV/ 91-141


710
720
730
740
750
760
P35251|RFC1_HUMAN/ 669-709
P35250|RFC2_HUMAN/ 90-127
P40938|RFC3_HUMAN/ 84-121
P35249|RFC4_HUMAN/ 159-196
P40937|RFC5_HUMAN/ 74-111
P38630|RFC1_YEAST/ 373-412
P40348|RFC2_YEAST/ 89-126
P38629|RFC3_YEAST/ 126-163
P40339|RFC4_YEAST/ 63-100
P38251|RFC5_YEAST/ 90-127
Q5UQK9|RFCL_MIMIV/ 174-233
Q5UQ72|RFCS1_MIMIV/ 128-165
Q5UP47|RFCS2_MIMIV/ 136-173
Q5UQ47|RFCS3_MIMIV/ 117-154
Q5UQE8|RFCS4_MIMIV/ 142-179

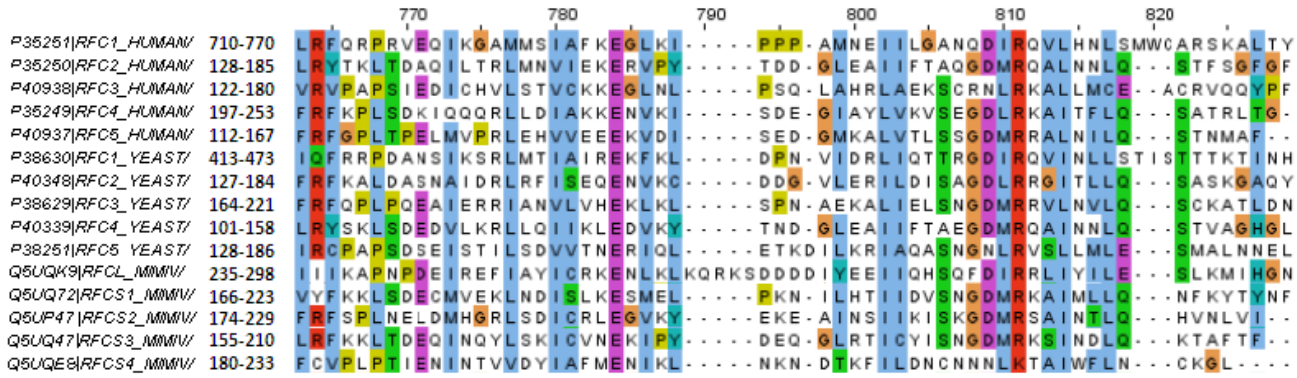

770
780
790
800
810
820
P35251|RFC1_HUMAN/ 710-770
P35250|RFC2_HUMAN/ 128-185
P40938|RFC3_HUMAN/ 122-180
P35249|RFC4_HUMAN/ 197-253
P40937|RFC5_HUMAN/ 112-167
P38630|RFC1_YEAST/ 413-473
P40348|RFC2_YEAST/ 127-184
P38629|RFC3_YEAST/ 164-221
P40339|RFC4_YEAST/ 101-158
P38251|RFC5_YEAST/ 128-186
Q5UQK9|RFCL_MIMIV/ 235-298
Q5UQ72|RFCS1_MIMIV/ 166-223
Q5UP47|RFCS2_MIMIV/ 174-229
Q5UQ47|RFCS3_MIMIV/ 155-210
Q5UQE8|RFCS4_MIMIV/ 180-233

Figure 1.12: Multiple sequence alignment of RFC proteins belonging to two eukaryotic organisms, Homo sapiens and *S. cerevisiae,* and the Mimivirus. The location of the potential RFC PIP-box domains is highlighted, although only part of the protein alignment is shown. The red boxes demonstrate full or part PIP-box motifs belonging to the different organisms, further revealing conserved residues. The distribution of PIP-box motifs in the various RFC subunits, for all three organisms studied, is in agreement with preceding studies that discovered their presence only in the RFC1, RFC3 and RFC5 subunits (*Kazlauskas D & Venclovas C, 2011*). The protein alignments were performed using the MUSCLE bioinformatics software program, while the results were visualised using the same software program mentioned previously.

Note: The Marseillevirus and CroV were not included in this protein alignment regarding the RFC protein. As previously discussed, none of the RFC protein subunits have been uncovered in the Marseillevirus genome. In the case of CroV, even though it has been established that it encodes for all five RFC subunits (*Kazlauskas D & Venclovas C, 2011*), these have not been fully characterised yet. Some CroV RFC proteins have been annotated, as can be seen in *Table 1.2*, but not much more has been acknowledged about them (i.e. what RFC subunits they actually demonstrate). Furthermore, only part of the protein alignment is depicted in *Figure 1.12*. The reason for this being the large sequence length of most proteins included in the alignment. However, the part of the alignment considered to be of most significance is illustrated. This specifically contains the majority of conserved residues between the different organisms, thus probably representing conserved protein sequence motifs such as PIP-box motifs, which are vital for interaction with the PCNA protein. The location of the potential RFC PIP-box domains highlighted in *Figure 1.12* has been predicted due to knowledge surrounding this type of motifs (*See Section 1.5*), and is also in accordance with a similar study conducted by *Kazlauskas D & Venclovas C, 2011* (*Refer to Figure 7 of their study*).

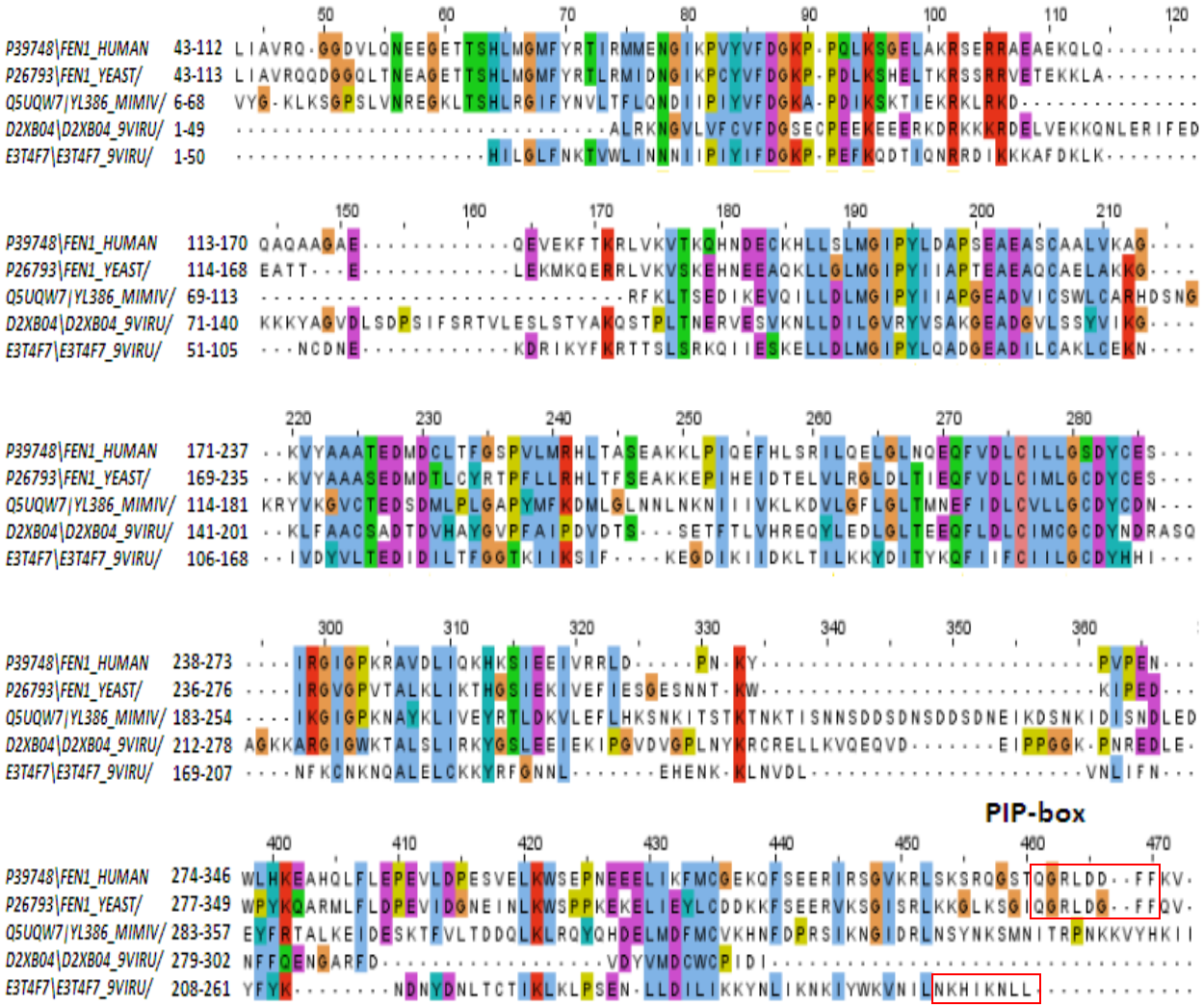


Figure 1.13: Multiple sequence alignment of FEN1 proteins belonging to two eukaryotic organisms, Homo sapiens and *S. cerevisiae,* and the three giant viruses of interest, Mimivirus, Marseillevirus and CroV. The location of the potential FEN1 PIP-box domains is highlighted, although only part of the protein alignment is shown due to long unconserved protein sequences present in some of the organisms that were not aligned. The red boxes demonstrate the PIP-box motifs belonging to the different organisms. In the case of Mimivirus and Marseillevirus, however, their PIP-box domains are not apparent in this figure. The protein alignments were performed using the ClustalO 1.1.0 bioinformatics software program, while the visualisation of the results was performed in the same manner as before.

Some general conclusion can be drawn from these three protein alignment figures constructed for each of the PCNA, RFC and FEN1 proteins (*Figures 1.11, 1.12 and 1.13,* respectively). All figures show that a good proportion of all three proteins are well conserved amongst the different organisms, eukaryotic and viral, as can be observed from the multiple regions of conserved amino acid residues. However, apart for the PCNA protein, both the RFC and FEN1 protein alignments revealed long stretches of either not very well conserved or fairly unconcerned protein regions (these have been cut out from *Figures 1.12 and 1.13*).

Regarding the PIP-box motif domains that were investigated in the RFC and FEN1 proteins; these do not appear to be significantly conserved in either protein between the organisms examined, and this fact may further justify their fairly imprecise alignment. Nevertheless, for the RFC PIP-box domains, the protein alignments undertaken have confirmed that these motifs only exist in RFC subunits 1 (RFC1), 3 (RFC3) and 5 (RFC5), as can be seen in the three organisms studied (i.e. Homo sapiens, *S. cerevisiae* and Mimivirus). Interestingly, the protein sequence alignments for FEN1 demonstrated a completely aligned and fully conserved PIP-box motif between Homo sapiens and *S. cerevisiae*. On the contrary, an obvious FEN1 PIP-box motif could not be determined for either Mimivirus or Marseillevirus. The CroV, however, did reveal a distinct PIP-box motif at the far end of its FEN1 protein sequence, a discussion of which follows in *Section 1.5*.

1.5 PCNA-Interacting Peptide (PIP) box

A number of important studies have revealed that the PCNA is a target for the binding of several proteins, with the different proteins apparently competing between them for this binding. The explicit PCNA-binding domain has particularly been identified and characterised in a large number of assorted proteins, e.g. p21, Cdt1, Topo IIa, DNA polymerases and ligase I, RFC, XPG (including FEN1) etc., which are all involved in various cellular mechanisms, from DNA replication and DNA repair to DNA methylation and cell cycle control (*Gomes XV & Burgers PMJ, 2000*). The interesting fact about this discovery is that the interaction formed appears to take place through a conserved motif. This typical motif found within the sequence of the various proteins, either on the far end of their N- or most commonly C- terminus –depending on the protein, may possibly contact the same site on the PCNA, that possibly being the hydrophobic cleft of the PCNA buried under the interdomain connecting loop (*Warbrick E, 1998*). The proposed consensus PCNA-binding motif specifically is Q*XX*Ψ*XX*A, where '*X*' represents any amino acid, 'Ψ' stands for L/I etc. (i.e. residues with moderately hydrophobic aliphatic side chains) and 'A' refers to amino acids containing hydrophobic aromatic side chains (i.e. F/W/Y). The Q, 'Ψ' and 'A' residues are all conserved, while the motif is usually followed by a non-conserved sequence. [*Maga G & Hubscher U, 2003*; *Vivona JB & Kelman Z, 2003*; *Scorah J et al, 2008*]. The consensus PIP-box sequence has been found to be conserved in *Archaea* and *Eukarya*, as well as in some viruses. This fact indicates the importance of these conserved residues during evolution, when they most likely had a regulatory role in coordinating aspects of DNA metabolism. [*Warbrick E, 2000*; *Bruning JB & Shamoo Y, 2004*; *Moldovan GL et al, 2007; Winter JA & Bunting KA, 2012*; *Mailand N et al, 2013*].

The specific effect of PCNA binding on various proteins, which takes place through their PIP-box motifs, still remains mostly unclear. Although, as these PCNA-protein partner interactions appear to be conserved through evolution, it is apparent that they must be highly essential for coordinated, probably enhanced, and therefore successful DNA synthesis and repair. For most proteins the only actual information available is whether or not they contain such a motif, and as such whether they are capable of binding to and interacting with the PCNA protein. For example, it has previously been discovered that in regards to DNA polymerases their PIP-box domains are located on the β-subunit of PolIII (*Bacteria*), and on the small (perhaps regulatory) subunits of Polδ [i.e. p12 and p66 in humans and p32 in yeast] and on two out of four subunits of Polε, including its catalytic Pol2 subunit (*Eukarya*). By forming an interaction with these replicative polymerases, PCNA presumably provides them with the high processivity required for replicating an entire genome. In the case of RFC, as discussed earlier, PIP-box motifs are specifically positioned on the RFC1 (also known as RFC-A), RFC3 (RFC-C) and RFC5 subunits (three out of five subunits). FEN1 and DNA ligase I both contain PIP-box domains, but while binding of FEN1 to PCNA has been shown to stimulate its activity (*Hosfield DJ et al, 1998*; *Sakurai S et al, 2005*), its precise effect on DNA ligase I is unknown. Nevertheless, PCNA should presumably coordinate the action of these proteins in a stepwise reaction during Okazaki fragment maturation. In general, it has been shown that Polδ, Polε, and RFC show strongest interactions compared to FEN1 and DNA ligase 1 that bind with somewhat lower affinity to PCNA. Regarding the XPG nuclease protein, which is a structure-specific repair endonuclease similar to FEN1, a PIP-box motif has been identified commonly on the C-terminus of its protein sequence. The binding of XPG to PCNA is responsible for

nucleotide excision repair (NER) activity in cells; however, the exact way by which NER activity is promoted remains uncertain (*Tsurimoto T, 1999; Maga G & Hubscher U, 2003*; *Moldovan GL et al, 2007*).

The binding sites of the protein factors mentioned above onto the PCNA protein structure have already been mapped (*Figure 1.14*). Most PCNA-protein interaction take place on two major sites of the PCNA protein; those being the interdomain connecting loop and its C-terminal tail. In particular, proteins such as Polδ, FEN1 and DNA ligase I recognise the PCNA loop, whereas other proteins such as Polε and RFC bind to the PCNAs' C-terminus (*Maga G & Hubscher U, 2003*).

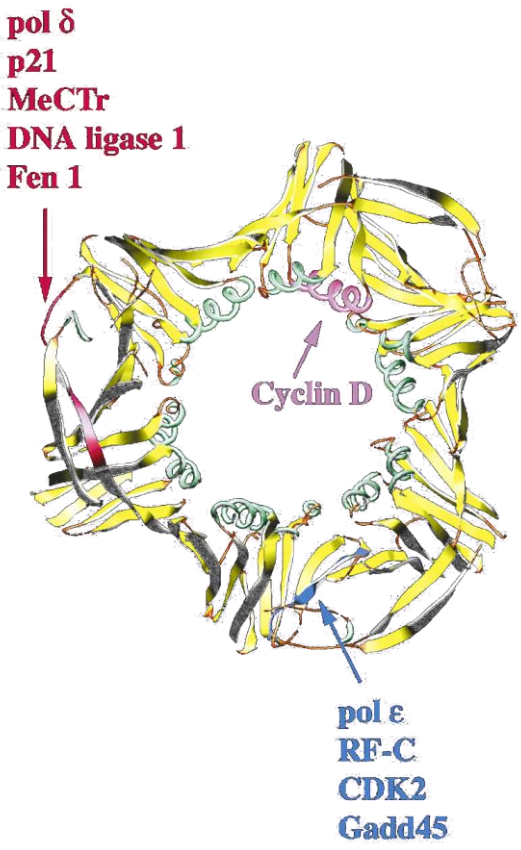


Figure 1.14: Sites of different proteins binding to the PCNA protein structure. Proteins marked in red bind to the interdomain connecting loop. Proteins marked in blue bind to the C-terminal tail. And the only protein marked in purple binds to the N-terminal tail. Figure obtained from *Maga G & Hubscher U, 2003.*

In terms of the FEN1 PCNA-binding peptide, which is of interest in this study; Characterisation studies of the PCNA-binding region of the FEN1 showed that the interaction between the two proteins took place in the presence of a specific 20aa sequence located at the C-terminus of the FEN1 protein, described as the FEN1 PIP-box motif (*Warbrick E, 2000*). The proposed sequence of the eukaryotic FEN1 PIP-box domain particularly is Q--L--FF (*Warbrick E, 1998*; *Zheng L et al, 2007*). The FEN1 PIP-box motif characteristics of giant viruses have not been studied yet, and generally not much is known about other proteins' PIP-box domains belonging to this group of organisms (*Note: The only study completed to date that mentions about a PIP-box sequence in giant viruses was conducted by Kazlauskas D & Venclovas C (2011) and particularly concerned the RFC PIP-box belonging to the Mimivirus and CroV; See Section 1.4.3*).

Consequently, during this project, an attempt was made to use a synthetic peptide containing a putative PIP-box domain as identified in the last 20aa of the CroV FEN1 protein, on its C-terminus (CroV037; *as obtained from UniProt and further highlighted in Figure 1.13*). The sequence of the hypothetical PIP-box motif was N--I--LL. It can be argued that this domain is comparable to the proposed consensus PCNA-binding motif, and subsequently to the eukaryotic FEN1 PIP-box domain, due to the specific sequence of the different amino acid groups. In particular, the motif begins with a neutral polar amino acid (i.e. N, which has the same properties as Q), continues with a hydrophobic amino acid (i.e. I) and ends with two hydrophobic residues (i.e. LL), which even though non-aromatic,

are still hydrophobic and, as such, not atypical of such a motif. The resulting synthetic peptide was used as a PCNA purification tool.

1.6 PCNA/FEN1 Protein Complex Interactions

Previous experimental studies have revealed that the PCNA sliding clamp can actually stimulate the action of the FEN1 nuclease; PCNA specifically recruits FEN1 at the site of branched DNA substrates near the replication folk, where it enhances its nuclease activity by 10- to 50- fold (*Hosfield DJ et al, 1998*; *Sakurai S et al, 2005*). By forming a direct interaction, an increase of FEN1 binding stability to the dsDNA is achieved, further allowing for a far greater cleavage specificity and efficiency (*Samson T et al, 2000*). The interaction between the two proteins is particularly formed when FEN1 interact with the PCNA molecule through a hydrophobic cleft located at the front part of the latter. Notably, the interaction domains of the PCNA and FEN1 have been pinned down to two regions belonging to either protein; the hydrophobic cleft, formed by the interdomain connector loop of PCNA, and a small sequence of conserved residues on C-terminus of FEN1. In the initial stages of this interaction, the PCNA-binding motif (PIP-box) (*Warbrick E, 1998 & 2000*) found on the C-terminus of FEN1 mediates the correct binding of the latter to the hydrophobic cleft of the former, but only in the absence of the double-stranded DNA. However, once the PCNA interacts and encircles the DNA molecule, FEN1 requires the formation of a C-terminus PCNA interaction so as to ensure its correct and accurate function. Furthermore, as PCNA interacts at the same time with DNA polymerases, it suggests that FEN1 is especially recruited to the DNA replicase complex bound to the

replication fork (*Figure 1.15*). [*Hosfield DJ et al, 1998*; *Jonsson ZO et al, 1998*; *Gomes XV & Burgers PMJ, 2000*; *Maga G & Hubscher U, 2003*; *Chapados BR et al, 2004*; *Sakurai S et al, 2005*; *De Biasio A et al, 2012*].

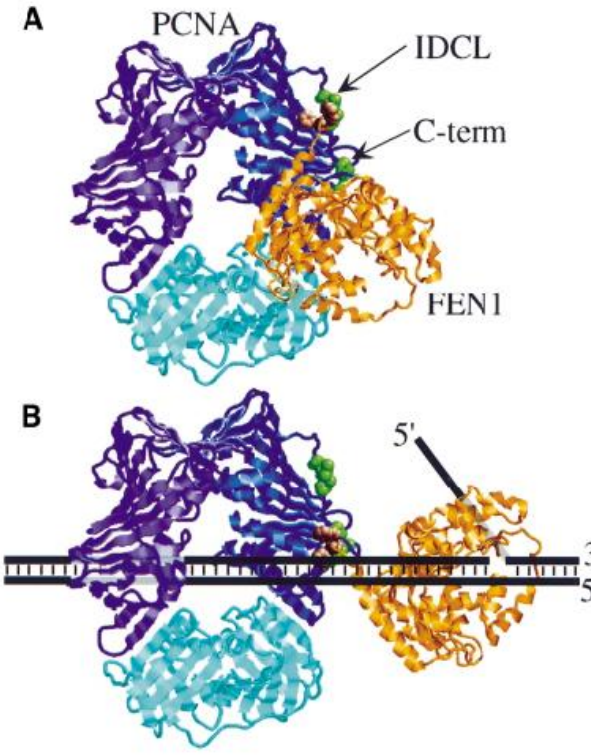


Figure 1.15: PCNA and FEN1 interaction model. (A) A typical heterotrimeric PCNA molecule is displayed, containing its interdomain connector loop (IDCL). Attached to the PCNA is the one-subunit FEN1 molecule. Binding occurs between the hydrophobic cleft, formed under the IDCL of the PCNA, and the C-terminus of the FEN1. (B) PCNA shown to encircle the DNA strand, carrying with it the FEN1 enzyme. 'A model of the interaction on a FLAP structure is shown'. Figure obtained from *Gomez XV & Burgers PMJ, 2000*.

It has been proved that mutations in the FEN1 protein disrupt its interaction capacities with the PCNA protein, therefore significantly decreasing FEN1 cleavage efficiency; at the same time they are accountable for a series of diseases seen to develop in mice (*Zheng L et al, 2007*; *Zheng L et al, 2011*). The latter outcome can be justified by the fact that correct FEN1/PCNA interaction is critical for faithful and efficient Okazaki fragment processing, hence for completing DNA replication and repair further supporting the stability of the genomic integrity in all organisms.

1.7 Project Aims

In general terms; the aim of this project was to understand the molecular 'make-ups', as well as the properties and functions, of two viral proteins regarded as key components of the giant virus replication machinery; the PCNA sliding clamp and the FEN1 endonuclease. Both these proteins have been shown to be conserved evolutionarily, and together they form a greater complex that is critical for DNA replication and repair. Therefore, the presence of both these proteins in whatever organism is absolutely necessary as it ensures the maintenance of the genomic integrity by passing safely the genetic information from generation to generation.

To complete this project, an extensive biochemical analysis of the Mimivirus PCNA, the Marseillevirus PCNA, and the CroV PCNA and FEN1 was performed. At the same time, however, the probable PCNA/FEN1 interaction was also investigated. The general goal was to attempt to gain additional insights into the lifestyle and evolution of these unique viruses.

On completion of this project, it was generally hoped that a better understanding would have been reached on how replication and genetic material propagation is accomplished in the giant viruses. Thus providing further knowledge as to how these giant viruses have expanded to such a great extent in genomic size, during the history of evolution; this fact has also made their visualisation under a microscope possible, as they are comparable in size to bacteria.

To conclude; the discovery of these extraordinary viruses has began to change the science community's and probably the public's view on evolution and the whole history of life. Debates regarding the nature of these viruses have been ongoing ever since they were first discovered, and still no one knows for certain how these discussions will end and what conclusions will be derived from them. For these reasons mentioned, it is imperative and of considerable value to continue this type of research. So far, however, the majority of research has mostly focused on the computation analysis ─both bioinformatic and phylogenetic studies─ of these types of viruses, struggling to decipher the entire history of their evolution, as well as whether or not they should be included in the 'tree of life'. On the contrary, though, since the giant viruses' discovery there has not been a great deal of research focusing on their actual biological nature. Scrutinising the actual molecular biology 'make-up' of these viruses may optimistically result in gaining a better understanding of their life cycle and a more intense insight into their own evolutionary hallmark. The outcome of this will hopefully be that one day in the near future the world will have a more conclusive and maybe definite answer regarding the nature of the giant viruses, and as such a better outlook of the evolution of this planet as we know it.

-CHAPTER TWO-

Materials and Methods

2.1 Materials

2.1.1 Reagents

The majority of reagents utilised for this study were purchased from Biogene (UK), Bioline (UK), BioRad (UK), Fermentas (UK), Fisher Scientific (UK), Formedium (UK), GE Healthcare (UK), IBA (Germany), Macherey-Nagel (UK), Melford (UK), New England Biolabs (NEB, UK), Novagen (UK), Promega (UK), Qiagen (UK), Sigma-Aldrich (UK) and ThermoScientific (UK), unless otherwise stated.

2.1.2 Synthetic Genes

The synthetic genes for the Mimivirus PCNA (APMV_L108; *Appendix A.1.1*), Marseillevirus PCNA (MAR_ORF212; *Appendix A.1.2*), CroV PCNA (CroV_219; *Appendix A.1.3*) and CroV FEN1 (CroV_037; *Appendix A.1.4*) were ordered from Genscript (UK). The maps of these constructs are attached in the Appendix (*A.1, Construct Maps of the Synthetic Genes*), as provided by the supplier.

2.1.3 Vectors & Bacteria Cells

The vectors used for cloning the Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212), CroV (CroV_219) PCNA and CroV FEN1 (CroV_037) genes were

purchased from IBA (UK) and Novagen (UK), while some were designed 'in-house'. In addition, the variety of bacteria cells used for the purpose of this study were obtained from Dr. MacNeills' *E.coli* plasmid library.

2.1.4 Primers

The primers used for the amplification of the Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212), CroV (CroV_219) PCNA and CroV FEN1 (CroV_037) genes were ordered from Eurofins MWG (UK) DNA Oligo Synthesis Company.

2.1.5 Restriction Enzymes

The enzymes utilised during the various cloning procedures were bought from NEB (UK), Fermentas (UK) or Promega (UK).

2.1.6 Solutions and Buffers

	Composition:
Agarose Running Buffer; Tris-Acetate EDTA buffer (50x TAE) For 500ml:	121g Tris-Hydrogen Chloride (HCl) 28.6ml glacial acetic acid 50ml 0.5M Ethylene-Diamine-Tetra-Acetic acid (EDTA) (pH 8.0)
Bacteria Freezing Medium (BFM)	Luria-Bertani (LB) medium 30% glycerol
Blue Loading Dye; 2x Sample Buffer (SB) (for SDS-PAGE)	100mM Tris-HCl (pH 6.8)4% SDS 20% Glycerol 0.05% Bromophenol blue 0.25% Dithiothreitol (DTT) H_20

Lysis Buffer; Buffer A (pH 6.5 - 8.5)	50mM Sodium Dihydrogen Phosphate (NaH_2PO_4)/ Sodium Phosphate Dibasic (Na_2HPO_4) 150-600mM Sodium Chloride (NaCl) 1mM β-mercaptoethanol 10-40mM Imidazole	
Lysis Buffer; Buffer W	100mM Tris-HCl (pH 8) 150-500mM NaCl 1mM EDTA [-/+] 0.5-2% Tween	
Lysis Buffer Phosphate-buffered Saline (PBS)-based	PBS 100mM Tris-HCl (pH 8) 1mM EDTA 1mM β-mercaptoethanol [-/+] 0.5% Tween [-/+] 0.5% Glycerol	
SDS-PAGE; Gels	<u>Resolving Gel</u> Acrylamide (40%) 1.5ml 1M Tris-HCl (pH 8.8) 20µl 20% SDS 20µl 10% Ammonium Peroxodisulfate (APS) 3.4µl Tetramethylethylenediamine (TEMED) H_20	<u>4% Stacking Gel</u> 0.4ml 40% Acrylamide 0.5ml 1M Tris-HCl (pH 6.8) 20µl 20% SDS 20µl 10% APS 5µl TEMED 3.05ml H_20
SDS-PAGE; 5x Running Buffer For <u>1L</u>:	15g Tris-HCl 72g Glycine 5g SDS	
SDS-PAGE; Fix Solution	20% Ethanol (EtOH) 7.5% Acetic Acid (HAc)	
SDS-PAGE; Staining Solution	0.1% PAGE Blue G90 4% Perchloric acid	
Transformation Buffer (TSB)	LB 5% Dimethyl Sulfoxide (DMSO) 10% Polyethylene Glycol (PEG) 3350 50mM Mg^{2+} (pH 6.5)	
Western Blot; 10x Transfer Buffer For <u>1L</u>: 1x Transfer Buffer	30g Tris-HCl 144g Glycine + 15% Methanol	

Table 2.1: **Solutions and Buffers used during the experimental procedures.**

2.1.7 Equipment

***AKTA* Purification System**	'900 Series', GE Healthcare, UK
Cell Culture Incubators	1. Incubator Shaker Series, Innova® 44 and I26, New Brunswick Scientific, UK 2. Orbital Incubator SI50, Stuart Scientific, UK
Centrifuges	1. Biofuge fresco, Heraeus, UK 2. Centrifuge 5810, Eppendorf, UK 3. Beckman Coulter, Optima™ L-90K Ultracentrifuge, UK 4. Sorvall Evolution RC Centrifuge, ThermoScientific, UK
Gel Imaging; Agarose and SDS-PAGE	U:Genius, Syngene, UK
Heating Block	Dri-Block® DB-2D, Techne, UK
Polymerase Chain Reaction (PCR) ThermalCycler	PIKO, ThermalCycler, ThermoScientific, UK
Shaker	Model R100, Rotatest Shaker, Luckham, UK
Sonicators	1. Soniprep 150, MSE, UK 2. Ultrasound Processor UP2005, Hielscher Ultrasound Technology, UK
Spectrophotometer	SP-50 Spectrophotometer, Sanyo, UK
UltraViolet (UV) Transilluminator	High Performance Ultraviolet Transilluminator, UVP, LLC, UK
Western Blot Developer	Kodak X-Omat 1000 Processor, UK

Table 2.2: **Equipment used during the experimental procedures.**

2.2 Methods

2.2.1 General Protocols

2.2.1.1 Bacteria Transformation of PCNA/FEN1 constructs into *E.coli* cells

Fresh competent *E.coli* cells were made each time they were required by inoculating 20ml of LB with 200µl of the appropriate *E.coli* bacteria culture. This was incubated at 37^{o}C in a shaking incubator for approximately ~3 hours until the optical density (OD) of the bacteria culture reached 0.5-0.6. At that point, the bacteria culture was centrifuged at 3,000 rpm for 10 minutes to collect the *E.coli* cells. The bacteria pellet was next resuspended in 1 ml of TSB buffer (*See Table 2.1*) and placed on ice for 30 minutes. *E. coli* competent cells were produced.

To transform the PCNA and FEN1 constructs; 100µl of the competent cells were mixed with 10µl of the previously carried out ligation reaction. The bacteria transformation reaction was initially kept on ice for 30 minutes, while subsequently a further 200µl of TSB buffer were added to it and finally this was incubated at 37^{o}C for an hour with shaking.

The PCNA and/or FEN1 bacteria transformations were plated on appropriate antibiotic resistance agar plates and placed overnight in a 37^{o}C incubator.

2.2.1.2 Bacteria Colony PCR for PCNA/FEN1 transformants

This protocol involved standard PCR techniques. Specifically, after transforming the PCNA and/or FEN1 plasmids, a variety of bacteria colonies were selected and screened for correct transformants.

The protocol undertaken was as follows:

Bacteria Colony PCR Protocol	
Reagents	**Final Volumes**
Bacteria Colony (Diluted in 50-100µl H_2O)	1.5µl
Red PCR Master Mix (5x) (MyTaqRed, Bioline, UK) - dNTPs - $MgCL_2$ - MyTaqRed DNA Polymerase + stabilizers/enhancers	 5mM 15mM 1.25 units
-Forward Primer (100µM) -Reverse Primer (100µM)	0.25µl 0.25µl
Total Volume:	4.5µl
Final Volume	6µl

Table 2.3: **Example of a typical 6**µl **volume Bacteria Colony PCR master mix.** The red dye incorporated into this DNA polymerase mix allows for direct visualization of genomic bands produced during agarose gel electrophoresis.

The typical cycling conditions were as follows:

Cycling Conditions for Bacteria Colony PCR			
Cycle		**Temperature**	**Time**
Initial Denaturation		94^{o}C	3’
30 cycles	Denaturation	94^{o}C	10’’
	Annealing	55 ^{o}C	10’’
	Polymerisation	72^{o}C	10’’

Table 2.4: **Example of the typical cycling conditions for a Bacteria Colony PCR reaction.**

The correct transformants were selected and grown on a small scale preparation. This was carried out by inoculating 10µl of the previously diluted bacteria colony, used for the PCR reaction, into 10ml of LB medium containing an appropriate antibiotic. This was then left to grow overnight by shaking in a 37^{o}C incubator. Finally, the bacteria plasmids were purified using the QIAprep Spin Miniprep kit (NEB, UK) or the GeneJet Plasmid Miniprep kit (Fermentas, UK).

2.2.1.3 Protein Expression Experiments for PCNA/FEN1 proteins

Following the transformation of the Mimivirus PCNA (APMV_L108), Marseillevirus PCNA (MAR_ORF212), CroV PCNA (CroV_219) and CroV FEN1 (CroV_037) into the appropriate *E.coli* host strain(s), a single transformed colony was selected and grown

overnight on a small scale preparation in a 37^{o}C incubator by shaking, as explained already.

To test the protein expression efficiency of the PCNA and/or FEN1 proteins, initially a mini scale bacteria preparation was carried out; this was done by inoculating 500µl of the bacteria culture previously grown overnight in 50ml LB medium (+ appropriate antibiotic).

[*For a midi scale bacteria preparation, 2ml of the bacteria culture grown overnight were further inoculated in 250ml LB medium (+ appropriate antibiotic). Additionally, for a large scale bacteria preparation, 8ml of the bacteria culture grown overnight were further inoculated in 1L LB medium (+ appropriate antibiotic).*]

This was then left to shake in an incubator set at 37^{o}C for ~3-4 hours until the OD of the bacteria culture reached ~0.5-0.6. At that stage, a small sample of the culture was taken before adding the appropriate amount (1,000 fold less than the actual bacteria culture) of anhydrotetracycline (AHT) or Isopropyl β-D-1-thiogalactopyranoside (IPTG) so as to induce PCNA and/or FEN1 protein expression. This sample was labeled 'U' for Uninduced because no protein expression would have taken place. After inducing protein expression, the bacteria culture was further incubated. However, at this point, the incubation time and temperature varied according to the particular experimental procedure being followed (i.e. overnight incubation was carried out for bacteria cultures growing at 15-25^{o}C, while a 4 hour incubation period was undertaken for cultures growing at 30-35^{o}C). At the end of the incubation period, another small sample of the culture was taken to represent the effect of the inducing reagents on protein expression levels. This sample

was labeled ‘I’ for Induced as protein expression should have occurred. Both the uninduced and induced samples were properly treated by taking a ~20µl sample of each and boiling it at 95^{o}C for 4 min in 2xSB buffer (*See Table 2.1*). These samples were finally used for SDS-PAGE analysis; always following standard techniques.

Finally, the PCNA/FEN1 bacteria cells were harvested by centrifugation. The cells would then be resuspended in a suitable lysis buffer and sonicated for the purpose of disrupting the cell membranes and hence releasing all the cellular contents (*See below Section 2.2.1.4 Preparing Soluble PCNA/FEN1 protein samples by Sonication procedures*), further allowing for analysis of protein expression levels by SDS-PAGE or Western Blotting.

As mentioned, SDS-PAGE and Western blot experiments were conducted following standard procedures. More specifically, for both types of experiments specific percentage resolving gels were prepared according to the size of the proteins to be detected (*See Table 2.1: Solutions and Buffers used during the experimental procedures*); these most commonly were 10% or 12.5% resolving gels. The various protein samples of interest were then loaded onto the gel and run at a particular voltage (i.e. ~200V for 1 hour), alongside a suitable protein marker used as a size standard (PageRuler™ Plus, or non-Plus version, Prestained Protein Ladder (Fermentas, UK)). In the case of the SDS-PAGE experiments, once the gel had finished running it was fixed in a fix solution, made from ethanol and acetic acid, for 10 mins on a bench shaker, stained with a stain solution (i.e. PAGE Blue G90, also containing perchloric acid) for ~5 mins and finally destained in water for a

suitable amount of time until the protein bands became significantly visible for further analysis. The SDS-PAGE gel images were captured using the ImageLab Program (UK). In the case of the Western blot experiments, after running the SDS-PAGE, the proteins run on the gel were transferred onto a Westran PVDF membrane by running the prepared transfer gel/membrane 'sandwich' for 1 hour at ~100V. The resulting membrane was then blocked by a typical BSA-based solution, for the purpose of blocking non-specific binding sites. Detection of the target proteins was achieved by specific to the experiment antibodies. A chemiluminescent signal for the antibody-detected proteins present on the membrane was produced by soaking the Westran PVDF membrane in Pierce-ECL Western Blotting substrate (ThermoScientific, UK). The light corresponding to the proteins of interest was finally detected by photographic films that were developed using a Kodak X-Omat 1000 processor (UK).

Note**:** It is important to mention the predicted molecular masses (i.e. kDaltons) at which each PCNA and FEN1 protein, for the different viruses, was supposed to run on a SDS-PAGE gel or Western blot; the Mimivirus PCNA (APMV_L108) protein has a 32 kDa molecular mass, the Marseillevirus PCNA (MAR_ORF212) protein has a molecular mass of 34 kDa, the CroV PCNA (CroV_219) protein has a 32 kDa molecular mass (same as the Mimivirus) and finally, the CroV FEN1 (CroV_037) protein has a molecular mass of 37 kDa. Nevertheless, most proteins do not run at the exact same molecular masses as estimated and specifically PCNA proteins appear to have higher molecular weights than those predicted.

2.2.1.4 Preparing Soluble PCNA/FEN1 protein samples by Sonication procedures

According to the specific mass/volume of PCNA and/or FEN1 bacteria cell pellet that was previously harvested; this was lysed in the appropriate amount of particular buffers (i.e. Buffer W; Tris-HCl base, or Buffer A; Phosphate-base, *See Table 2.1*). The residual suspension, representing the total protein sample, was then sonicated by performing 5x10 sec bursts (<20-40% sonicator power) with 30 sec cooling intervals. The sonication protocol was always carried out under ice-cooling conditions (4^oC).

Following sonication, the suspension was centrifuged at high speed (i.e. 13,000rpm) for 15 min again at low temperature conditions (4^oC). This would allow for the membrane, and generally all the insoluble, contents of the cells to form a pellet, while all the soluble protein contents would remain in the supernatant. This sample that was labeled 'S' for Soluble, as it contained the solubilised protein of interest, would then be treated accordingly and used for SDS-PAGE or Western blot analysis; run alongside with the Uninduced and Induced samples, which were collected during the 'Protein Expression Experiments' (*See Section 2.2.1.3; Protein Expression Experiments for PCNA/FEN1 proteins*).

2.2.1.5 PCNA/FEN1 Protein Purification experiments using an *AKTA* system (GE Healthcare, UK)

The *AKTA* is a system for purifying proteins, fast and efficiently, for further use in biochemical analysis studies. It allows the user to easily develop specific and simple

protocols, while at the same time permitting the optimisation of the methods utilised for protein purification according to the experimental needs. Furthermore, it allows for small-scale protein purification of either tagged or un-tagged proteins, making it extremely useful compared to other protein purification techniques. What is more, it simply works by loading the soluble protein sample of interest onto the system. During the protein purification, different protein fraction samples are collected and can be easily accessed by the UNICORN software program that, for this purpose, generates a graph illustrating where proteins are eluting off the particular column.

2.2.1.6 Mass Spectrometry (MS)

Mass spectrometry of the protein bands of interest was carried out by our in house mass spectrometry facilities (BSRC Mass Spectrometry and Proteomics Facility, University of St. Andrews, UK). More specifically, the protein band to be analysed was precisely excised from a SDS-PAGE gel, prepared carefully so as to minimise keratin contamination, and subjected to in-gel digestion by trypsin for peptide separation and extraction. Protein identification analysis was initially carried out on an AB Sciex 4800 MALDI (Matrix-Assisted Laser Desorption/Ionisation) TOF/TOF™ Analyser (AB Sciex, UK) instrument; while samples were then processed using MASCOT and the data was compared against the NCBI (National Center for Biotechnology Information) and the in-house BMS protein databases. In most instances, when no conclusive results could be produced by the MALDI analyser, then a higher sensitivity apparatus was used (i.e. ESI (ElectroSpray Ionisation) analyser). This particularly was the AB Sciex QStar XL

NanoLC-ESI qTOF™. All results were finally evaluated using the ProteinPilot™ software (version 4.0.8085, Applied Biosystems, UK) for protein identification and quantitation.

2.2.2 Proliferating Cell Nuclear Antigen (PCNA) protein

2.2.2.1 Cloning Techniques and Plasmid Construction for the Mimivirus, Marseillevirus and CroV PCNA proteins

The original synthetic genes for the Mimivirus PCNA (APMV_L108), Marseillevirus PCNA (MAR_ORF212) and CroV PCNA (CroV_219) were cloned into a pUC57 backbone by EcoRV restriction digest, as described by the provider company (*Appendix A.1; Construct Maps of the Synthetic Genes_A.1.1-A.1.3, respectively).* At the same time, unique *Bsa*I restriction sites were intentionally introduced on either side of these PCNA genes to facilitate further digestion and hence cloning procedures.

2.2.2.1.1 Cloning of the Mimivirus, Marseillevirus and CroV PCNA proteins into a pASK-IBA17plus vector backbone, containing a Strep-Tactin affinity tag (IBA, UK) for protein purification

Note: The pASK-IBA17plus vector contains the Strep-Tactin affinity tag (= ~30 bp) attached to a TEV (Tobacco Etch Virus) protease cleavage site (= ~20 bp). So cloning the PCNA gene of interest, from the different viruses, into this vector would add an additional ~50bp to the PCNA gene sequence or otherwise ~2 kDa to the PCNA protein sequence. This would increase the protein molecular masses of the three different viral PCNAs to: 34

kDa for the Mimivirus PCNA (APMV_L108) protein, 36 kDa for the Marseillevirus PCNA (MAR_ORF212) protein and 34 kDa for the CroV PCNA (CroV_219) protein (same as the Mimivirus).

The Mimivirus PCNA (pUC57_APMV_L108), Marseillevirus PCNA (pUC57_MAR_ORF212) and CroV PCNA (pUC57_CroV_219) genes, as well as the empty pASK-IBA17plus vector, were digested with *Bsa*I restriction enzyme. Restriction digests were carried out for 3 hours at 50^{o}C; this temperature being specific to the *Bsa*I restriction enzyme.

At 2.5 hours of *Bsa*I restriction digest, the 3' and 5' ends of the pASK-IBA17plus vector were cleaned from the phosphate (P)-groups using the Antarctic Phosphatase enzyme (NEB, UK) by placing the reaction for a final 30 minutes in a heating block set at 37^{o}C.

On the other hand, following the *Bsa*I restriction digest of the three different PCNA constructs, the different sized fragments produced (i.e. PCNA genes at around ~900 bp and empty pUC57 vector backbone at around ~1800 bp), were separated on standard 1% (w/v) agarose electrophoresis gels and gel purified. This permitted correct restriction digest confirmation and accurate size validation of each PCNA product, when compared to a suitable DNA marker (GeneRuler™ 1kb DNA Ladder, Fermentas, UK).

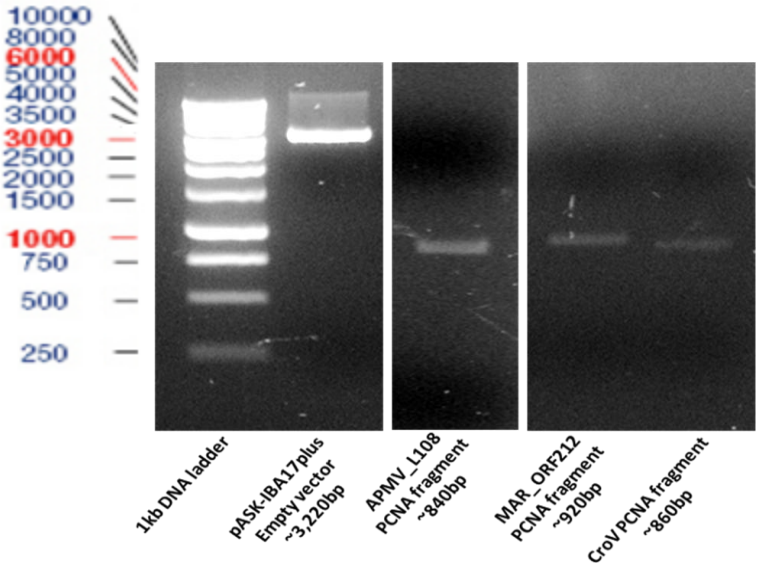


Figure 2.1: <u>Example of 1% Agarose Gel Photograph showing the products of a restriction digest</u>; This 1% agarose gel shows the products generated after *Bsa*I restriction enzyme digest of pASK-IBA7plus empty vector (3,220 bp), Mimivirus pUC57_APMV_L108 PCNA (840 bp), Marseillevirus pUC57_MAR_ORF212 PCNA (920 bp) and CroV pUC57_CroV_219 PCNA (860 bp). All PCR amplification products were analysed against the GeneRuler™ 1kb DNA Ladder, and the images were captured using the ImageLab Program (UK).

All restriction digest reactions were purified using the Qiagen MinElute Gel Extraction kit (Qiagen, UK).

The three different ~900 bp *Bsa*I digested PCNA fragments for the Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV-219) were ligated into the pASK-IBA17plus vector, using the Quick Ligation kit (NEB, UK). The ligation reactions were left for 20 minutes at room temperature (RT), after which they were transformed into DH5α *E.coli* cells on ampicillin resistance agar plates; [The pASK-IBA17plus vector confers ampicillin resistance] (*See Section 2.2.1.1; Bacteria*

Transformation of PCNA/FEN1 constructs into E.coli cells). DH5α *E.coli* cells are the most common strain of choice for routine cloning as they increase the insert stability and improve the quality of plasmid DNA prepared from minipreps, as explained by the manufacturer.

The selection of correct transformants containing the Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA gene fragments ligated into the pASK-IBA17plus vector was performed by bacteria colony PCR reaction (*See Section 2.2.1.2; Bacteria Colony PCR for PCNA/FEN1 transformants*). The primers used for this PCR reaction were specific to the pASK-IBA17plus vector (*Table 2.5*), so as to allow for specificity confirmation of the PCR amplification process and accurate size validation of each PCR product, when compared to a suitable DNA marker (*Figure 2.2*).

Oligonucleotides (sequence 5' → 3')	**Ta^{o}C**	**PCR product size (bp)**
FW: GAGTTATTTTACCACTCCCT	48	1,150
Rev: CGTTTACCGCTACTGCG	49	

Table 2.5: **Oligonucleotides specific to the pASK-IBA17plus vector; for amplification of the Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA gene fragment.**

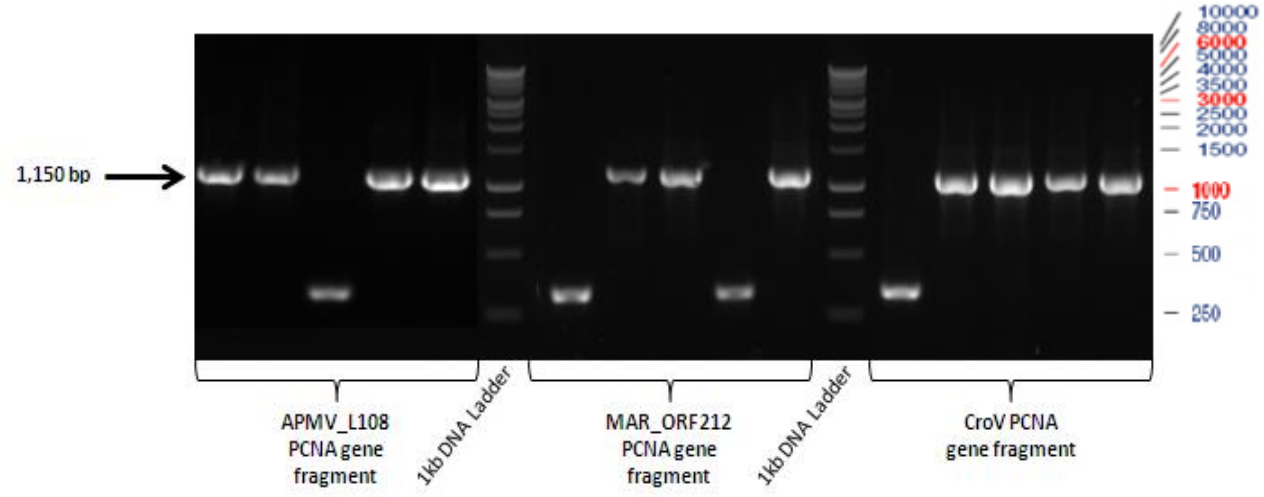


Figure 2.2: <u>Example of 1% Agarose Gel Photograph showing the amplification products of a bacteria colony PCR reaction</u>; This 1% agarose gel shows the PCR products acquired when using as a DNA template bacteria colonies that grew after the ligation and transformation of Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA gene fragments into the pASK-IBA17plus vector. The primers employed were specific for the pASK_IBA17plus vector. The correct transformants generate PCR products of 1,150 bp in size.

However, to verify with certainty that the *Bsa*I digested Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA gene fragments had been inserted into the pASK-IBA17plus vector, another restriction digest was undertaken with the *Xba*I and *Hind*III enzymes, following standard procedures. The expected sizes for all three viruses were around ~1,000 bp.

Once the correct pASK-IBA17plus Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA constructs were confirmed (*Appendix A.2; Construct Maps of the pASK-IBA17plus Mimivirus (APMV_L108), Marseillevirus*

(MAR_ORF212) and CroV(CroV_219) PCNAs), these were further transformed into Rosetta2(DE3)(pLysS) *E.coli* cells on ampicillin/chloramphenicol resistance agar plates, following the same method as described previously. This Rosetta2(DE3)(pLysS) strain allows for a higher level of protein expression, as it is specifically "designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*", as described by the supplier (Novagen, UK).

The Rosetta2(DE3)(pLysS)_pASK-IBA17plus Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA plasmids were used to perform protein expression studies.

2.2.2.1.2 Cloning of the Mimivirus, Marseillevirus and CroV PCNA proteins into a pEHISTEV vector backbone, containing a PolyHistidine (or HexaHis/6xHis) affinity tag for protein purification

Note: The pEHISTEV vector contains the 6xHis affinity tag (= ~20 bp) attached to a TEV protease cleavage site (= ~20 bp). So cloning the PCNA gene of interest, from the different viruses, into this vector would add an additional ~40bp to the PCNA gene sequence or otherwise ~2 kDa to the PCNA protein sequence. This would increase the protein molecular masses of the three different viral PCNAs, in the same manner as was the case for the pASK-IBA17plus vector (i.e. the Mimivirus PCNA (APMV_L108) protein molecular mass increases to 34 kDa, the Marseillevirus PCNA (MAR_ORF212) protein

molecular mass increases to 36 kDa and the CroV PCNA (CroV_219) protein molecular mass increases to 34 kDa (same as the Mimivirus).

Similar procedures were followed as explained in *Section 2.2.2.1.1; Cloning of the Mimivirus, Marseillevirus and CroV PCNA proteins into a pASK-IBA17plus vector backbone, containing a Strep-Tactin affinity tag (IBA, UK) for protein purification*.

More precisely, the CroV PCNA (pASK-IBA17plus _CroV_219; *Appendix A.2.3*) gene, as well as the empty pEHISTEV vector, were digested with *Nco*I and *Hind*III restriction enzymes for 3 hours at 37^{o}C. The pEHISTEV vectors' 3' and 5' ends were once again cleaned from the P-groups using the Antarctic Phosphatase enzyme, as previously described, while the digested CroV PCNA construct was run on an 1% agarose gel electrophoresis and the two bands of interest (= 460 + 400 = 860 bp; former band generated from single *Nco*I digest, while latter from double *Nco*I/*Hind*III, i.e. CroV PCNA contains two NcoI sites and one *Hind*III site resulting in the production of two different fragments) cut out and gel purified. The restriction digest reactions were specifically purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, UK). Next, the total 860 bp *Nco*I/*Hind*III digested CroV PCNA fragment was ligated into the pEHISTEV vector, by a two-step ligation procedure (i.e. one fragment at a time), and transformed into DH5α *E.coli* cells on kanamycin resistance agar plates. [The pEHISTEV vector confers kanamycin resistance].

In the case of the Mimivirus PCNA (pASK-IBA17plus _APMV_L108; *Appendix A.2.1*) and the Marseillevirus PCNA (pASK-IBA17plus _MAR_ORF212; *Appendix A.2.2*); for the introduction of the particular PCNA gene fragments into the pEHISTEV vector, a novel *Nco*I site had to be generated to allow for *Nco*I/*Hind*III restriction digest. The Mimivirus and Marseillevirus PCNA gene sequence, in contrast to the CroV PCNA gene sequence, do not have a naturally occurring *Nco*I restriction site. Therefore, to create this novel *Nco*I site, specific primers had to be ordered containing a site-directed mutation forming the *Nco*I restriction site of interest. The new restriction site would purposely be introduced by a long-range PCR reaction, where the forward primers of the Mimivirus and Marseillevirus PCNA gene sequence, already containing the *Nco*I restriction site, together with the reverse primer of the pASK-IBA17plus vector would amplify the specific Mimivirus (APMV_L108) and Marseillevirus (MAR_ORF212) PCNA gene fragment introducing the new *Nco*I restriction site, which would then be utilised in a *Nco*I/*Hind*III restriction digest reaction.

The sequence of the site-directed mutated primers containing the *Nco*I restriction site, were as follows:

Constructs	**Oligonucleotides (sequence 5' → 3')**	**Ta°C**	**PCR product size (bp)**
APMV	FW: ATGCACTCA *CCATGG* CT ACGAGCTGTGCTGACAAC	68	840
MAR	FW: ATGCACTCA *CCATGG* CT TCATTCGTGGGCTCACTG	68	920
pASK-IBA17plus	Rev: CGTTTACCGCTACTGCG	49	

Table 2.6: **Site-directed *Nco*I mutated oligonucleotides specific to Mimivirus PCNA (pASK-IBA17plus _APMV_L108) and the Marseillevirus PCNA (pASK-IBA17plus _MAR_ORF212); these primers introduce a novel *Nco*I restriction site within the PCNA gene sequence. This will be exploited in a *Nco*I/*Hind*III restriction digest reaction.**

The PCR protocol undertaken was as follows:

Long-Range PCR Protocol	
Reagents	**Final Volumes**
DNA Template (pASK-IBA17plus APMV/MAR PCNA constructs)	~1µl
10x Buffer (+MgCl)	5µl
dNTPs mix	1µl
DMSO	2.5µl
APMV/MAR Fw Primers	0.5µl
pASK-IBA17 plus Rev Primer	0.5µl
Enzyme mix (Long-PCR enzyme, Fermentas, UK)	0.25µl
H2O	39.25µl
Total Volume	50µl

Table 2.7: **Example of a Long-Range PCR Protocol.**

The cycling conditions were as follows:

Long-Range PCR Cycling Conditions			
Cycle	**Cycle**	**Temperature**	**Time**
	Initial denaturation	95^{o}C	3’
35	Denaturation	95^{o}C	30’’
	Annealing	58^{o}C	30’’
	Extension	72^{o}C	2’ 30’’
	Extension	72^{o}C	6’ 40’’
	Final Extension	72^{o}C	10’’

Table 2.8: **Example of Long-Range PCR cycling conditions.**

The Mimivirus PCNA (pASK-IBA17plus _APMV_L108) and the Marseillevirus PCNA (pASK-IBA17plus _MAR_ORF212) PCR products, after being successfully sequenced to guarantee no mutations had been introduced by the PCR reaction, were purified and digested with *Nco*I and *Hind*III restriction enzymes. The 840 bp and 920 bp generated from the *Nco*I/*Hind*III restriction digest for the Mimivirus PCNA (pASK-IBA17plus _APMV_L108) and the Marseillevirus PCNA (pASK-IBA17plus _MAR_ORF212), respectively, were also ligated into the pEHISTEV vector and transformed into DH5α *E.coli* cells on kanamycin resistance agar plates.

The selection of correct transformants containing the Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA gene fragments ligated into the pEHISTEV vector was performed by another *Nco*I/*Hind*III restriction digest. This

digest would allow the direct confirmation of the presence or absence of the Mimivirus, Marseillevirus and CroV PCNA fragments of interest cloned within the pEHISTEV vector. However, the transformants corresponding to the CroV PCNA (CroV_219) gene, ligated into the pEHISTEV vector, had to also be checked for their correct (i.e. sense) *Nco*I fragment orientation. This is due to the fact that the CroV PCNA (CroV_219) gene sequence contains two naturally occurring *Nco*I restriction sites, so when *Nco*I/*Hind*III digested it would have resulted in the appearance of two fragments instead of one, as is the case for Mimivirus (APMV_L108) and Marseillevirus (MAR_ORF212) PCNA. The confirmation that the *Nco*I/*Hind*III digested CroV PCNA fragments were cloned in a sense orientation in the pEHISTEV vector was achieved by another double SalI/PstI restriction digest, generating a fragment of 330 bp. [If the *Nco*I digested CroV PCNA fragment had been inserted in an anti-sense orientation then the SalI/PstI digest would have produced a fragment of 730 bp, as seen on a 1% agarose gel].

Once the correct pEHISTEV Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA constructs were confirmed (*Appendix A.3; Construct Maps of the pEHISTEV Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV(CroV_219) PCNAs*), these were further transformed into Rosetta2(DE3)(pLysS) *E.coli* cells on kanamycin/chloramphenicol resistance agar plates. The resulting plasmids would then be used for various protein expression studies.

Note: In addition, only the CroV PCNA (pASK-IBA17plus _CroV_219) gene was also transformation into a variety of different other *E.coli* host strains (following a similar study as the one described by *Busso D et al, 2011*). So far all constructs created were only transformed into the Rosetta2(DE3)(pLysS) *E.coli* strain hoping for high efficiency of protein expression. The additional strains examined were: BL21(DE3), BL21(DE3)(pLysS), Rosetta2(DE3), Arctic Express(DE3)RP and C43(DE3). The BL21 strains are "all-purpose strains for high-level protein expression and easy induction, while the *pLysS* one provides tighter control of protein expression for expression of toxic proteins", as explained by the supplier (Stratagene, UK). The Arctic Express strain is usually used to overcome any protein insolubility and misfolding, while the C43 strain is effective for expressing toxic proteins.

2.2.2.1.3 Cloning only of the CroV PCNA protein into pETDuet-1 & pCDFDuet-1 vector backbones, containing NO affinity tags for protein purification

Note: As the pETDuet-1 and pCDFDuet-1 vectors do not contain any sort of affinity tag, in this case the molecular masses of the PCNA proteins, for the three different viruses, remain intact.

The CroV PCNA (pEHISTEV _CroV_219; *Appendix A.3.3*) gene, as well as the empty pETDuet-1 and pCDFDuet-1 vectors (not containing any affinity tags to aid in protein purification procedures), were digested with *Nco*I and *Hind*III restriction enzymes, treated suitably and purified for further use. Subsequently, the *Nco*I/*Hind*III fragments of interest

digested from the pEHISTEV _CroV_219_PCNA construct were ligated into the also digested pETDuet-1 and pCDFDuet-1 vectors and transformed into DH5α *E.coli* cells on ampicillin or streptomycin resistance agar plates; pETDuet-1 is ampicillin resistant, while pCDFDuet-1 is streptomycin resistant. The selection of correct transformants containing the CroV PCNA (CroV_219) gene fragment ligated into the pETDuet-1 and pCDFDuet-1 vectors was conducted by another *Nco*I/*Hind*III restriction digest, but also by a SalI/NdeI double digest for reasons explained previously. Once the correct pETDuet-1 (*Appendix A.5.1*) and pCDFDuet-1 (*Appendix A.5.2*) CroV PCNA (CroV_219) constructs were confirmed (*Appendix A.5; Construct Maps of the pETDuet-1* and *pCDFDuet-1 CroV PCNA (CroV_219) and CroV FEN1 (CroV_037)*), these were further transformed into Rosetta2(DE3)(pLysS) *E.coli* cells on ampicillin or streptomycin/+chloramphenicol resistance agar plates.

However, these CroV PCNA (CroV_219) constructs were additionally co-transformed with the equivalent CroV FEN1 (CroV_037) constructs in the same Rosetta2(DE3)(pLysS) *E.coli* cells, so as to examine if and how these two proteins interact when run on the *AKTA* system. More specifically, two co-transfections took place; these were: pETDuet-1_CroV_219_PCNA + pCDFDuet-1_CroV_037_FEN1 and pCDFduet-1_CroV_219_PCNA + pETDuet-1_CroV_037_FEN1 (See also *Sections 2.2.1.5; PCNA/FEN1 Protein Purification experiments using an AKTA system, and 2.2.3.1.2; Cloning of the CroV FEN1 protein into pETDuet-1 & pCDFDuet-1 vector backbones, containing NO affinity tags for protein purification*). The resulting plasmids would then be used for various protein expression studies.

2.2.2.2 Protein Expression Experiments for the Mimivirus, Marseillevirus and CroV PCNA proteins

The exact method is described in *Section 2.2.1.3; Protein Expression Experiments for PCNA/FEN1 proteins*. Nevertheless, some adjustments were made; one of them being that protein expression for the pASK-IBA17plus Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA constructs is only achieved by the addition of anhydrotetracycline (AHT) protein expression inducer.

On the contrary, protein expression for the pEHISTEV Mimivirus (APMV_L108), Marseillevirus (MAR_ORF212) and CroV (CroV_219) PCNA constructs is only induced by Isopropyl β-D-1-thiogalactopyranoside (IPTG) reagent.

2.2.2.3 Preparing Soluble PCNA protein samples from the Mimivirus, Marseillevirus and CroV by Sonication procedures

The precise protocol is described in *Section 2.2.1.4; Preparing Soluble PCNA/FEN1 protein samples by Sonication procedures*. However, for the different Mimivirus, Marseillevirus and CroV PCNA constructs produced different lysis buffers were tested, so as to evaluate the efficiency of PCNA protein expression. Below is a table with all the lysis buffers assayed (*Table 2.9*).

	Ingredient Concentration
pASK-IBA17plus PCNA constructs	1. 100mM Tris-HCl pH8, 150/250/500mM NaCl, 1mM EDTA
	2. 100mM Tris-HCl pH8, 150/250/500mM NaCl, 1mM EDTA + 1% Tween
	3. 100mM Tris-HCl pH8, 150/250/500mM NaCl, 1mM EDTA + 2% Tween
pEHISTEV & pEHISGFPTEV PCNA constructs	1. 50mM Na2HPO4, 400mM NaCl, 1mM β-mercaptoethanol, pH8
	2. 50mM Na2HPO4, 250/400/500/600mM NaCl, 10/20/40/65mM Imidazole, 1mM β-mercaptoethanol, pH8
	3. 50mM Na2HPO4, 250/400/500/600mM NaCl, 10/20/40/65mM Imidazole, 1mM β-mercaptoethanol, pH8 + 0.5% Tween
	4. PBS, + 0.5% Tween
	5. PBS, 100mM Tris-HCl pH8, 10mM Imidazole, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol
pETDuet-1 & pCDFDuet-1 PCNA constructs	1. 50mM NaH2PO4 (pH 6.5/ 7.5/ 8.5, with Na2HPO4), 0/ 150/ 300 mM NaCl, 1mM EDTA and 1mM β-mercaptoethanol (-/+ 0.5% Tween) 2. 100mM Tris-HCl pH8, 150mM NaCl, 1mM EDTA (-/+ 0.5% Tween and/or 1mM β-mercaptoethanol) 3. PBS, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol (-/+ 100mM Tris-HCl pH8)

Table 2.9: The different Lysis Buffers tested for Mimivirus, Marseillevirus and CroV PCNA protein expression efficacy

2.2.2.4 Protein Purification Experiments for the Mimivirus, Marseillevirus and CroV PCNA proteins

2.2.2.4.1 Streptavidin pull-down of PCNA Solubilised protein samples for the Mimivirus, Marseillevirus and CroV, with the use of strept-tactin magnetic beads (*only for the pASK-IBA17plus PCNA constructs*)

As already mentioned, the pASK-IBA17 plus vector has attached onto it a Strep-Tactin affinity tag to assist protein purification. In this case, to test the Mimivirus, Marseillevirus and CroV PCNA purification efficiency, ~20µl of the Strep-Tactin magnetic beads were added directly into each soluble PCNA protein sample, following the manufacturer's instructions (IBA, Germany). The samples were mixed for an hour on a wheel at 4^{o}C, followed by ~4-6 washes with the specific lysis buffer previously used for sonication. Finally, they were treated by boiling them at 95^{o}C for 5 min in 2xSB buffer. These were then run on a SDS-PAGE gel for analysis of the effectiveness and purity of PCNA purification. In addition, these Strep-Tactin pulled-down PCNA protein samples were also run on a Western blot, for comparison reasons. The Western blot protocol carried out can be found in the '*Expression and purification of proteins using Strep-tag and/or 6xHistidine-tag*' manual from the IBA (UK) website.

On another occasion, however, the Strep-Tactin beads were packed so as to create a Strep-tag column. The column would always be equilibrated with the specific lysis buffer previously used for sonication. The soluble PCNA protein lysates prepared by the

sonication procedure would then travel through it and the different flow-through PCNA protein fractions would be collected for further analysis of protein purification.

2.2.2.4.2 Pull-down of PCNA Solubilised protein samples for the Mimivirus, Marseillevirus and CroV, with the use of Nickel (Ni-NTA) affinity agarose beads (*only for the pEHISTEV PCNA constructs*)

As mentioned earlier, the pEHISTEV vector has attached onto it a 6xHis affinity tag to assist protein purification. Therefore, to examine the Mimivirus, Marseillevirus and CroV PCNA purification efficiency, the exact same technique as described in *Section 2.2.2.4.1; Streptavidin pull-down of PCNA Solubilised protein samples for the Mimivirus, Marseillevirus and CroV, with the use of strept-tactin magnetic beads (only for the pASK-IBA17plus PCNA constructs)* was undertaken. The only difference in this case being that Ni-NTA agarose beads (Qiagen, UK) were used instead of Strep-Tactin magnetic beads. Furthermore, a Ni-NTA agarose bead packed column was also used for Mimivirus, Marseillevirus and CroV PCNA purification. The samples were collected, properly treated and run on a SDS-PAGE gel for analysis of the PCNA protein purity.

2.2.2.4.3 EZview™ Red Streptavidin pull-down of CroV PCNA Solubilised protein sample, with the use of EZview™ Red Streptavidin affinity agarose gel beads and the help of a *Fen1* peptide (*only for the pETDuet-1 PCNA construct*)

The pETDuet-1 vector differs from most vectors due to the fact that it contains no affinity tag for easy and straight-forward protein purification. Consequently, in the case of this experiment, an intermediate agent had to be utilised so as to assist purification of the PCNA protein. This was a biotinylated *Fen1* designed oligopeptide, which would eventually be pulled-down with the help of the EZview™ Red Streptavidin affinity agarose gel beads (Sigma, UK). This type of beads strongly trap biotinylated target proteins on their N-terminus.

The *Fen1* peptide was produced by GenScript (UK). It was specifically 20aa in length and its exact sequence was "IKNKIYWKVNIL**N**KH**I**KN**LL**", while it contained the biotin tag on its N-terminus. The peptide was deliberately designed to correspond to the last 20aa of the CroV FEN1 (CroV_037) protein sequence, consequently also forming the specific PIP-box motif domain (i.e. **N--I--LL**) for precise interaction with the PCNA protein.

The protocol carried out was as mentioned in *Section 2.2.2.4.1; Streptavidin pull-down of PCNA Solubilised protein samples for the Mimivirus, Marseillevirus and CroV, with the use of strept-tactin magnetic beads* (*only for the pASK-IBA17plus PCNA constructs*), the only difference being that both the *Fen1* peptide and the red beads were added

simultaneously to the soluble PCNA protein sample. The protein samples were collected, treated and run on a SDS-PAGE gel for analysis of the PCNA protein purity.

2.2.2.4.4 Protein Purification Experiments for CroV PCNA using an *AKTA* system

During the very first *AKTA* purification attempt for the CroV PCNA protein cloned into the pEHISTEV vector (i.e. pEHISTEV_CroV_219_PCNA construct), a simple one-step purification protocol was created. The column used for trapping the CroV PCNA protein of interest was a histidine binding His-Trap™ column containing Ni Sepharose™. During this procedure, two different salt (NaCl) concentrations were tested for the purpose of evaluating their effect on the PCNA protein purity. The lysis buffers specifically chosen can be seen in *Table 2.9* (*pEHISTEV & pEHISGFPTEV PCNA constructs; Lysis Buffer No.3*); both buffers contained 40mM Imidazole, while one contained 250mM NaCl and the other 500mM NaCl. The majority of the fractions collected were run on a SDS-PAGE gel.

The CroV PCNA sample that had previously been resuspended in the 250mM NaCl lysis buffer was subsequently used in a gradient purification reaction. This time the variant being examined was the Imidazole concentration, which was varied between 40mM to 500mM. These concentrations were assayed so as to evaluate how the gradual increase of Imidazole would affect the detachment of the PCNA protein from the His-Trap™ column used to originally trap it and finally to assess how pure the detached protein would be when collected. The fractions that produced a small peak, as seen on the graph generated by the UNICORN software, were run both on an SDS-PAGE gel and a Western blot.

The third attempt to purify the CroV PCNA protein, cloned this time into the pETDuet-1 vector (i.e. pETDuet-1_CroV_219_PCNA), was another gradient purification protocol assaying different salt (NaCl) concentrations varying from 150mM to 1M (*pETDuet-1 PCNA construct Lysis Buffer No.2, Tables 2.9*). A Hi-Trap™ Q (anion) HP column was used, as this column effectively binds negatively charged proteins, such as the PCNA protein. The most significant fractions collected (i.e. the ones that appeared as peaks on the UNICORN software graph) were run on a SDS-PAGE gel.

2.2.3 Flap Structure-specific Endonuclease 1 (FEN1) protein

2.2.3.1 Cloning Techniques and Plasmid Construction for the CroV FEN1 protein

The original synthetic gene for the CroV FEN1 (CroV_037) was cloned into a pUC57 backbone by EcoRV restriction digest, as described by the provider company (*Appendix A.1.4; Construct Maps of the Synthetic Genes_CroV (CroV_037) FEN1).*

2.2.3.1.1 Cloning of the CroV FEN1 protein into a pEHISTEV and pEHISGFPTEV vector backbone, both containing a 6xHis affinity tag for protein purification, while the latter also contains a GFP-tag

Note: Both the pEHISTEV and pEHISGFPTEV vectors contains the 6xHis affinity tag (= ~20 bp) attached to a TEV protease cleavage site (= ~20 bp). So cloning the FEN1 gene of interest, from CroV, into these vectors would add an additional ~40bp to the FEN1 gene

sequence or otherwise ~2 kDa to the FEN1 protein sequence. This would increase the CroV FEN1 (CroV_037) protein molecular mass to 39 kDa. Moreover, the pEHISGFPTEV vector also includes an eGFP-tag (= ~730 bp or 28 kDa). This would further increase the CroV FEN1 (CroV_037) protein molecular mass to 67 kDa.

The precise same protocol, explained in *Section 2.2.2.1.2; Cloning of the Mimivirus, Marseillevirus and CroV PCNA proteins into a pEHISTEV vector backbone, containing a PolyHistidine (or HexaHis/6xHis) affinity tag for protein purification*, was undertaken.

The CroV FEN1 (pUC57 _CroV_037_FEN1; *Appendix A.1.4*) gene, as well as the empty pEHISTEV and pEHISGFPTEV vectors, were digested with *Nco*I and *Hind*III restriction enzymes. The pEHISTEV and pEHISGFPTEV vectors' 3' and 5' ends were treated with the Antarctic Phosphatase enzyme, while the digested CroV FEN1 (CroV_037) construct was directly purified. The pUC57 vector is ampicillin resistance, so it will not grow on kanamycin resistance plates that are required for pEHISTEV and pEHISGFPTEV plasmid growth. Subsequently, the 970 bp *Nco*I/*Hind*III digested CroV FEN1 (CroV_037) fragment was directly ligated into the pEHISTEV and pEHISGFPTEV vectors and transformed into DH5α *E.coli* cells on kanamycin resistance agar plates.

The selection of correct transformants containing the CroV FEN1 (CroV_037) gene fragment ligated into the pEHISTEV and pEHISGFPTEV vectors was once again

performed by a *Nco*I/*Hind*III restriction digest, revealing either the presence or the absence of the FEN1 gene of interest. Moreover, a bacteria colony PCR reaction was also carried out to confirm that the transformants selected were indeed correct. The primers used for this PCR reaction were the T7 forward and reverse primers, while the products generated were a ~1,200bp fragment for the pEHISTEV_CroV_037_FEN1 construct and a ~2,200bp fragments for the pEHISGFPTEV_CroV_037_FEN1 construct.

Once the correct pEHISTEV and pEHISGFPTEV CroV FEN1 (CroV_037) constructs were confirmed (*Appendix A.4; Construct Map of the pEHISTEV and pEHISGFPTEV CroV FEN1 (CroV_037)*), these were also transformed into Rosetta2(DE3)(pLysS) *E.coli* cells on kanamycin/chloramphenicol resistance agar plates. The resulting plasmids were used for different protein expression studies.

Note: The CroV FEN1 (pEHISTEV _CroV_037_FEN1) construct was also transformed into various other *E.coli* host strains (based on a similar study performed by *Busso D et al, 2011*); BL21(DE3), BL21(DE3)(pLysS), Rosetta2(DE3), Arctic Express(DE3)RP and C43(DE3).

2.2.3.1.2 Cloning of the CroV FEN1 protein into pETDuet-1 & pCDFDuet-1 vector backbones, containing NO affinity tags for protein purification

Note: As the pETDuet-1 and pCDFDuet-1 vectors do not contain any sort of affinity tag, in this case the molecular mass of the CroV FEN1 (CroV_037) protein remains unchanged.

The CroV FEN1 (pEHISTEV_CroV_037_FEN1; *Appendix A.4*) gene, as well as the empty pETDuet-1 and pCDFDuet-1 vectors, were digested with *Nco*I and *Hind*III restriction enzymes, treated suitably and purified for further use. Subsequently, the *Nco*I/*Hind*III fragments of interest digested from the pEHISTEV _CroV_037_Fen1 construct were ligated into the also digested pETDuet-1 and pCDFDuet-1 vectors and transformed into DH5α *E.coli* cells on ampicillin or streptomycin resistance agar plates; pETDuet-1 is ampicillin resistant, while pCDFDuet-1 is streptomycin resistant. This cloning technique resulted in the creation of un-tagged protein. The selection of correct transformants containing the CroV FEN1 (CroV_037) gene fragment ligated into the pETDuet-1 and pCDFDuet-1 vectors was conducted by another *Nco*I/*Hind*III restriction digest. Once the correct pETDuet-1 (*Appendix A.5.3*) and pCDFDuet-1 (*Appendix A.5.4*) CroV FEN1 (CroV_037) constructs were confirmed (*Appendix A.5; Construct Maps of the pETDuet-1* and *pCDFDuet-1 CroV (CroV_219) PCNA and CroV FEN1 (CroV_037)*), these were further transformed into Rosetta2(DE3)(pLysS) *E.coli* cells on ampicillin or streptomycin/+chloramphenicol resistance agar plates. The resulting plasmids would then be used for various protein expression studies.

As described earlier, the pETDuet-1_CroV_037_FEN1 and pCDFDuet-1_CroV_037_FEN1 constructs were also co-transformed with pCDFduet-1_CroV_219_PCNA and pETDuet-1_CroV_219_PCNA, respectively, in the Rosetta2(DE3)(pLysS) *E.coli* strain. The reason for this being to check for any protein-protein interactions between these two different proteins that are both involved in DNA replication. These experiments were conducted on the *AKTA* system. [See also *Sections*

2.2.1.5; PCNA/FEN1 Protein Purification experiments using an AKTA system, and 2.2.2.1.3; Cloning of the CroV PCNA protein into pETDuet-1 & pCDFDuet-1 vector backbones, containing NO affinity tags for protein purification].

2.2.3.2 Protein Expression Experiments for the CroV FEN1 protein

The same protocol described in *Section 2.2.1.3; Protein Expression Experiments for PCNA/FEN1 proteins w*as followed. For both pEHISTEV and pEHISGFPTEV CroV FEN1 (CroV_037) constructs, protein expression was induced by the IPTG reagent.

2.2.3.3 Preparing Soluble FEN1 protein samples from the CroV by Sonication procedures

The protocol has already been described in *Section 2.2.1.4; Preparing Soluble PCNA/FEN1 protein samples by Sonication procedures*. For the pEHISTEV and pEHISGFPTEV CroV FEN1 (CroV_037) constructs produced, different lysis buffers were tested so as to evaluate the protein expression efficacy and levels of the FEN1 protein. Below is a table with the various lysis buffers assayed (*Table 2.10*).

	Ingredient Concentration
pEHISTEV & pEHISGFPTEV FEN1 constructs	1. 50mM Na2HPO4, 300mM NaCl, 1mM β-mercaptoethanol, 20mM Imidazole, pH8
	2. 50mM Na2HPO4, 500mM NaCl, 1mM β-mercaptoethanol, 30mM Imidazole, pH8 +0.5% Tween
	3. 100mM Tris-HCl pH8, 150/250mM NaCl, 10/20mM Imidazole, 0.5% Tween, 1mM β-mercaptoethanol
	4. PBS, +0.5% Tween or 0.5% glycerol
	5. PBS, 100mM Tris-HCl pH8, 10mM Imidazole, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol

Table 2.10: The different Lysis Buffers tested for CroV FEN1 protein expression efficacy

2.2.3.4 Protein Purification Experiments for the CroV FEN1 protein

2.2.3.4.1 Pull-down of FEN1 Solubilised protein samples for the CroV, with the use of Nickel (Ni-NTA) affinity agarose beads (*pEHISTEV & pEHISGFPTEV FEN1 constructs*)

The same technique as explained in *Section 2.2.2.4.2; Pull-down of PCNA Solubilised protein samples for the Mimivirus, Marseillevirus and CroV, with the use of Nickel (Ni-NTA) affinity agarose beads (only for the pEHISTEV PCNA constructs)* was undertaken. The samples were collected, properly treated and run on both a SDS-PAGE gel and a Western blot for analysis of the FEN1 protein purity.

-CHAPTER THREE-

Results

3.1 Tagged Protein Expression of Giant Virus PCNA in *E.coli*

In summary, the Mimivirus PCNA (APMV_L108_32 kDa; *Appendix A.1.1*), Marseillevirus PCNA (MAR_ORF212_34 kDa; *Appendix A.1.2*) and CroV PCNA (CroV_219_32 kDa; *Appendix A.1.3*) gene fragments of interest were cloned into two different expression vectors containing an affinity tag; the first such vector was the pASK-IBA17plus vector containing a Strep-tag (~2 kDa), while the other one was a pEHISTEV vector containing a 6xHis-tag (~2 kDa).

3.1.1 Mimivirus

3.1.1.1 Cloning of the Mimivirus PCNA into the pASK_IBA17plus Strep-tag vector for Protein Expression Experiments

Refer to *Section 2.2.2.1.1; Cloning of the Mimivirus, Marseillevirus and CroV PCNA proteins into a pASK-IBA17plus vector backbone, containing a Strep-Tactin affinity tag for protein purification*.

The protocol resulted in the generation of construct pASK-IBA17plus_ Mimivirus(APMV_L108)_PCNA (*Appendix A.2.1*).

3.1.1.1.1 Protein Expression and Purification experiments for the Mimivirus PCNA, cloned into the pASK_IBA17plus Strep-tag vector and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

Refer to *Section 2.2.2.2; Protein Expression Experiments for the Mimivirus, Marseillevirus and CroV PCNA proteins*, as well as *Section 2.2.2.3 Preparing Soluble PCNA/FEN1 protein samples from the Mimivirus, Marseillevirus and CroV by Sonication procedures*.

First Mimivirus PCNA protein expression study in the pASK_IBA17plus Strep-tag vector:

For this first PCNA protein expression study, a mini-scale preparation of the Rosetta2 (DE3) (pLysS)_pASK-IBA17plus_Mimivirus (APMV_L108)_PCNA plasmid was incubated for ~3 hours at 37^{o}C, until the OD of the bacteria culture reached 0.5-0.6 (represented by the 'U'_APMV 'uninduced' sample on *Figure 3.1*). At that OD, PCNA protein expression was induced by the addition of AHT. The bacteria culture was left to incubate for a further 4 hours at 37^{o}C (represented by the 'I'_APMV 'induced' sample on *Figure 3.1*). The cells were harvested by centrifugation and resuspended in ~2ml of Lysis Buffer W (*Table 2.1 & 2.9*; *100mM Tris-HCl pH8, 150mM NaCl, 1mM EDTA*). The total cell extract was finally sonicated and the soluble cell extracts collected for protein expression analysis (represented by the 'S'_APMV 'soluble' sample on *Figure 3.1*). The samples collected were run on a 12.5% SDS-PAGE gel (*Figure 3.1*).

In addition, the same exact procedure was undertaken for the Rosetta2(DE3)(pLysS)_pASK-IBA17plus empty vector, for use as a control (*Figure 3.1*).

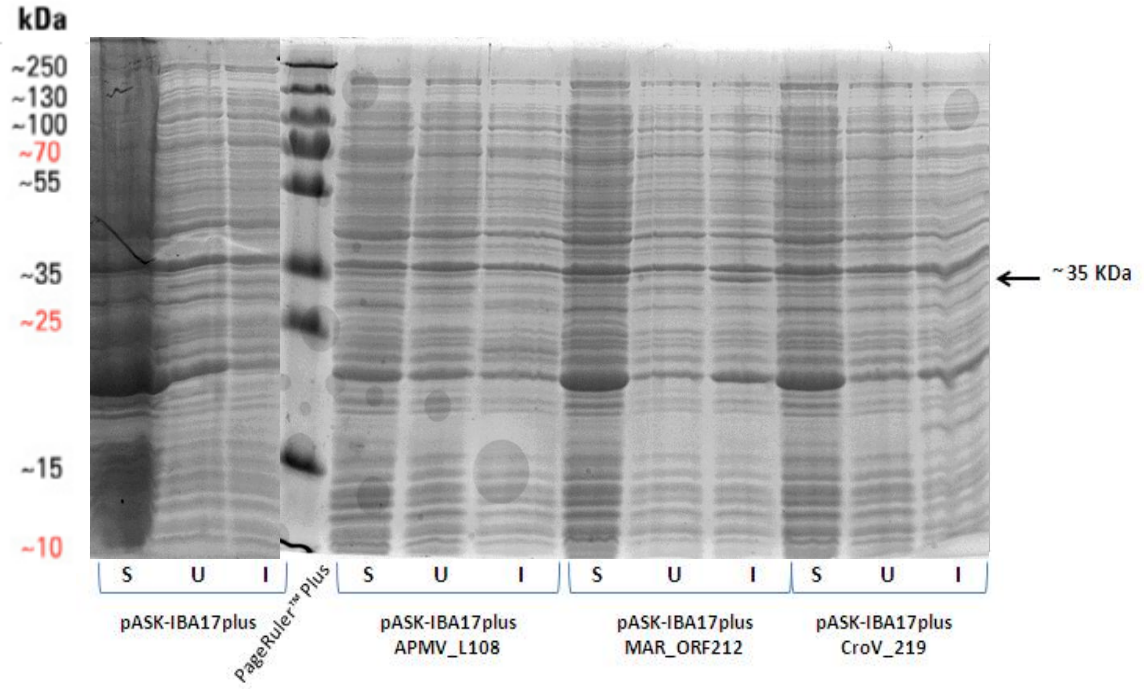


Figure 3.1: SDS-PAGE analysis of first Mimivirus, Marseillevirus and CroV PCNA protein expression study in the pASK_IBA17plus Strep-tag vector**: 12.5% SDS-PAGE Gel Photograph showing PCNA protein expression efficiency and levels for the pASK-IBA17plus Mimivirus, Marseillevirus and the CroV PCNA constructs, respectively, after PCNA protein expression induction was carried out for 4 hours at 37^{o}C;** (S): soluble, (U): uninduced, (I): induced. The expected PCNA protein molecular masses were 34 kDa for the Mimivirus, 36 kDa for the Marseillevirus and 34 kDa for the CroV. An extra band is only slightly more apparent for the Marseillevirus PCNA protein at ~34 kDa. The PageRuler™ Plus Prestained Protein Ladder (Fermentas, UK) was used as a size standard for comparison, while the pASK-IBA17plus empty vector protein samples were used as a control. The image was captured using the ImageLab Program.

The same samples, for the pASK-IBA17plus_Mimivirus (MAR_ORF212) PCNA (*Figure 3.1*), were used to perform a Western blot during which the Strep-tagged PCNA protein was directly detected using Strep-Tactin™ horse radish peroxidase (HRP) conjugate. The protocol conducted can be found in the "*Expression and purification of proteins using Strep-tag and/or 6xHistidine-tag*" manual (IBA, UK). However, no chemiluminescence signal was generated when developing the film. Therefore, the question that arose was whether or not the PCNA protein was actually being expressed.

The soluble protein samples for the Mimivirus, as well as the Marseillevirus and CroV, PCNA were also used to perform an initial streptavidin pull-down, with the use of Strep-Tactin magnetic beads (*For protocol see Section 2.2.2.4.1; Streptavidin pull-down of PCNA Solubilised protein samples for the Mimivirus, Marseillevirus and CroV, with the use of strept-tactin magnetic beads (only for the pASK-IBA17plus PCNA constructs*). No figure is attached because the streptavidin pull-down did not work. It could be questioned whether or not the Strep-tag that had been attached to these PCNA constructs was in some way obscured.

3.1.1.1.2 Summary of Results

As a general observation, the Mimivirus protein expression band patterns, as seen in *Figure 3.1*, are rather similar to the protein band patterns generated by the empty Rosetta2(DE3)(pLysS)_pASK-IBA17plus vector. Additionally, no distinct protein band

running at ~34 kDa (i.e. expected PCNA protein molecular mass) was produced by this organism. Therefore, the fact that no noticeable PCNA protein expression could be identified, not even when conducting a Western blot, led to the initial belief that most probably the protein expression conditions had to be optimised. However, instead of optimising the protein expression conditions, it was decided to clone the Mimivirus PCNA sequence into a different vector containing another protein tag hoping to obtain quicker better quality results. This decision would also save time in the case that the pASK-IBA17 plus Mimivirus PCNA construct was for some reason non-functional.

3.1.1.2 Cloning of the Mimivirus PCNA into the pEHISTEV 6xHis-tag vector for Protein Expression Experiments

Refer to *Section 2.2.2.1.2; Cloning of the Mimivirus, Marseillevirus and CroV PCNA proteins into a pEHISTEV vector backbone, containing a PolyHistidine (or HexaHis/6xHis) affinity tag for protein purification*.

The resulting construct was pEHISTEV_Mimivirus(APMV_L108)_PCNA (*Appendix A.3.1*).

3.1.1.2.1 Protein Expression and Purification experiments for the Mimivirus PCNA, cloned into the pEHISTEV 6xHis-tag vector and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

(See Sections 2.2.2.2 & 2.2.2.3)

First Mimivirus PCNA protein expression study in the pEHISTEV 6xHis-tag vector:

For this second Mimivirus PCNA protein expression study, a midi-scale preparation of the Rosetta2 (DE3) (pLysS)_pEHISTEV_Mimivirus (APMV_L108)_PCNA plasmid was incubated and grown following standard protocols. When the bacteria culture reached the desirable OD, PCNA protein expression was induced by the addition of IPTG (*not AHT*). The bacteria culture was left to incubate at 25^{o}C overnight. The cells were harvested by centrifugation and resuspended in ~2ml of the same Lysis Buffer W (*Table 2.1 & 2.9*; *100mM Tris-HCl pH8, 150mM NaCl, 1mM EDTA*). The total cell extract was finally sonicated and the soluble cell extracts collected for protein expression analysis. This time, however, the insoluble cell extracts were also used in the protein expression analysis, to allow the determination of whether or not the Mimivirus PCNA protein is soluble (represented by the 'P'ellet samples in *Figure 3.2*). All the samples collected were run on a 10% SDS-PAGE gel (*Figure 3.2*).

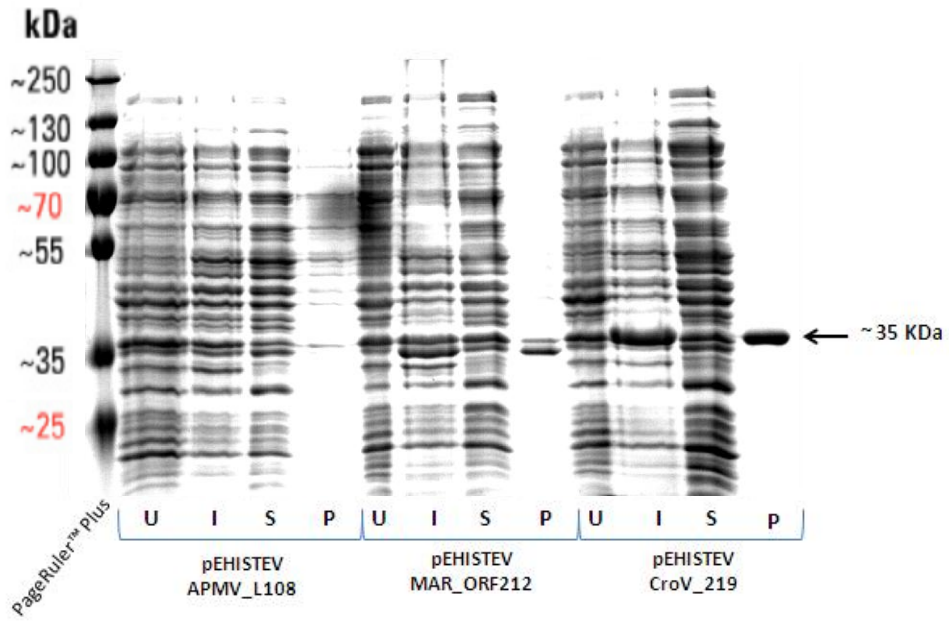


Figure 3.2: SDS-PAGE analysis of first Mimivirus PCNA protein expression study in the pEHISTEV 6xHis-tag vector: **10% SDS-PAGE Gel Photograph showing PCNA protein expression efficiency and levels for the pEHISTEV Mimivirus, Marseillevirus and the CroV PCNA constructs, respectively, after PCNA protein expression induction was carried out at 25^{o}C overnight;** (U): uninduced, (I): induced, (S): soluble, (P): pellet. The hypothesised PCNA proteins for the Marseillevirus and CroV appeared to be insoluble (i.e. visible in the 'P' samples).

The soluble PCNA protein samples for the Mimivirus, as well as the Marseillevirus and CroV (as seen in *Figure 3.2*), were also utilised in a pull-down procedure, with the use of Ni-NTA agarose beads (*For protocol see Section 2.2.2.4.2; Pull-down of PCNA Solubilised protein samples for the Mimivirus, Marseillevirus and CroV, with the use of Nickel (Ni-NTA) affinity agarose beads (only for the pEHISTEV PCNA constructs*). Pull-down refers to a column purified protein sample. The results of this pull-down experiment can be seen in *Figure 3.3*.

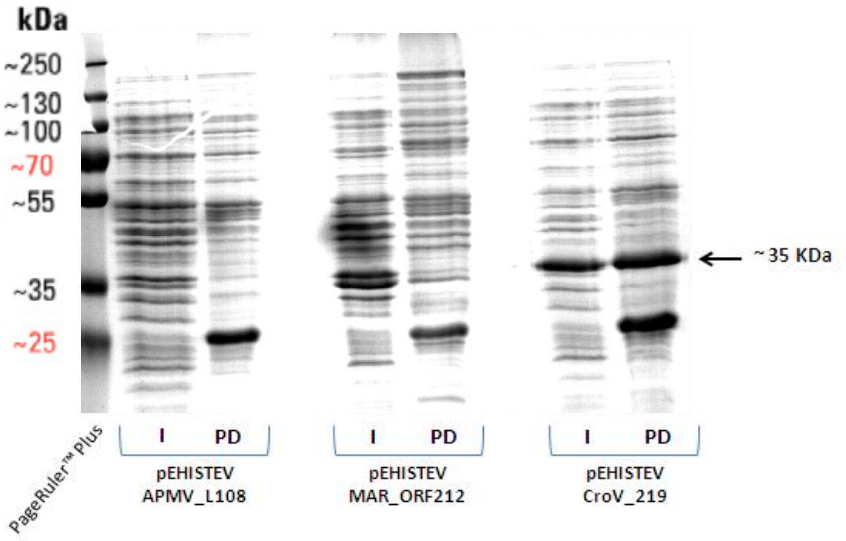


Figure 3.3: SDS-PAGE analysis of first Mimivirus PCNA protein expression study in the pEHISTEV 6xHis-tag vector, gel No.2**: 10% SDS-PAGE Gel Photograph showing soluble PCNA protein samples being pulled-down with Ni-NTA agarose beads for the pEHISTEV Mimivirus, Marseillevirus and the CroV PCNA constructs;** (I): induced (*as seen in Figure 3.2*), (PD): pull-downs. A strong PCNA protein signal was produced for the CroV, as seen from the 'PD' sample, and was estimated to be running at ~35-36 kDa.

3.1.1.2.2 Summary of Results

During this experiment an additional step was carried out; the pellet sample was collected after sonication and used so as to determine whether the Mimivirus PCNA protein was actually insoluble. If this was the case, a fairly distinct band running at ~34 kDa should have been obvious for this sample. However, this was not the case and, same as what was seen in the previous experiment, no noticeable PCNA protein expression could be identified in any case, not even when conducting a Ni-NTA agarose bead pull-down experiment. At this point, it was assumed that something was wrong either with the purchased Mimivirus PCNA protein sequence or with both the pASK-IBA17 plus and

pEHISTEV Mimivirus PCNA constructs generated, while it was also possible that the protein tags found within the vectors chosen for these experiments may have had a negative effect on the particular proteins expression by probably skewing its structure. As a result, no further experiments were carried out with the PCNA protein sequence belonging to this organism.

3.1.2 Marseillevirus

3.1.2.1 Cloning of the Marseillevirus PCNA into the pASK_IBA17plus Strep-tag vector for Protein Expression Experiments

(*See Section 2.2.2.1.1.*)

The construct created was pASK-IBA17plus_Marseillevirus(MAR_ORF212)_PCNA (*Appendix A.2.2*)

3.1.2.1.1 Protein Expression and Purification experiments for the Marseillevirus PCNA, cloned into the pASK_IBA17plus Strep-tag vector and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

(*See Sections 2.2.2.2 and 2.2.2.3*)

First Marseillevirus PCNA protein expression study in the pASK_IBA17plus Strep-tag vector:

For the protein expression protocol of the Rosetta2 (DE3) (pLysS)_pASK-IBA17plus_Marseillevirus (MAR_ORF212)_PCNA plasmid, refer to the corresponding 'Mimivirus PCNA protein expression' section, while the results can be seen in *Figure 3.1*.

In *Figure 3.1*, specifically for the pASK-IBA17plus_Marseillevirus (MAR_ORF212), the presence of an extra band in the induced and soluble protein sample at around ~35 kDa was quite noticeable, compared to the other viruses' samples. Therefore, it was presumed that this extra band might indeed be the Marseillevirus PCNA protein being expressed. To verify this assumption, further protein expression studies were conducted for the Marseillevirus (MAR_ORF212) PCNA.

Conclusions

The Marseillevirus protein expression band patterns, as seen in *Figure 3.1*, are rather comparable to the protein band patterns generated by the empty Rosetta2(DE3)(pLysS)_pASK-IBA17plus vector, as well as with the protein band patterns produced by both the Mimivirus and CroV. In contrast, however, to the Mimivirus and CroV, the Marseillevirus produced a slightly more distinct protein band, running at what was estimated to be ~34-35 kDa. This band though was only present in the sample collected after protein expression had been induced. This outcome indicated that even if

this band did indeed represent the Marseillevirus PCNA, this protein was present in minute levels and was most possibly not greatly soluble. Nevertheless, the fact that a somewhat positive result was produced led to the decision to optimise the protein expression conditions for the Marseillevirus PCNA protein.

Second Marseillevirus PCNA protein expression study in the pASK_IBA17plus Strep-tag vector:

In this experiment, the same procedure as followed during the first Marseillevirus protein expression was undertaken, the only difference being that a midi-scale preparation of the Rosetta2 (DE3) (pLysS)_pASK-IBA17plus_Marseillevirus (MAR_ORF212)_PCNA plasmid was prepared. The aim was to separate the original bacteria culture into four equal amounts just after inducing protein expression. Each equally separated sample was grown under different temperature conditions, and hence different time periods. The temperatures chosen were 20^{o}C (overnight protein expression), 25^{o}C (overnight protein expression), 30^{o}C (4 hour protein expression) and 37^{o}C (4 hour protein expression).

Following Marseillevirus PCNA protein expression induction, the bacteria cells from each sample were harvested. Each sample was further divided into two additional samples; one that was then treated with 1% Tween, while the other one did not contain any detergent. Treatment with Tween was conducted by mixing on a rotating wheel for 30 minutes at 4^{o}C. The results can be seen in *Figure 3.4*.

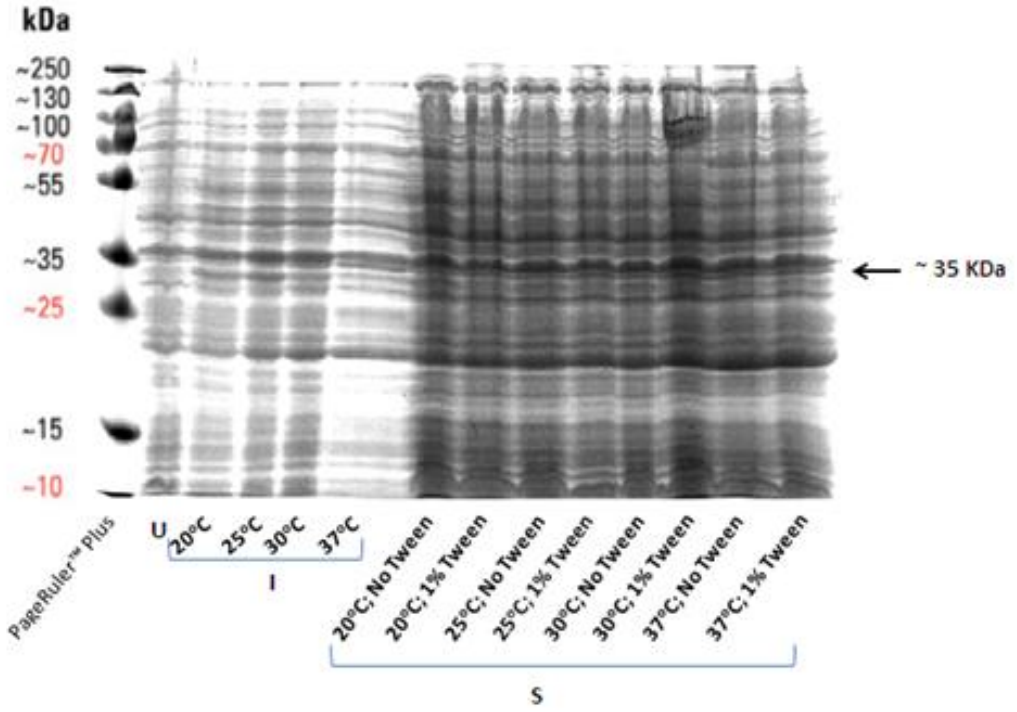


Figure 3.4: SDS-PAGE analysis of second Marseillevirus PCNA protein expression study in the pASK_IBA17plus Strep-tag vector**: 12.5% SDS-PAGE Gel Photograph showing PCNA protein expression efficiency and levels for the cloned pASK-IBA17plus Marseillevirus PCNA construct, after inducing protein expression in the original culture at four different temperatures and lysing the total protein samples in two different buffers;** Protein expression induction temperatures: 20^{o}C, 25^{o}C, 30^{o}C and 37^{o}C. Lysis buffers: 100mM Tris-HCl pH8, 150mM NaCl, 1mM EDTA, -/+ 1% Tween. The expected Marseillevirus PCNA protein molecular mass was 36 kDa. Maybe a PCNA protein was slightly expressed at around that molecular mass.

The samples depicted in *Figure 3.4* were subsequently used to perform a Western blot. The antibody utilised was the Strep-Tactin™ HRP conjugate, making it a one-step probing/detection procedure. The outcome of the Western blot can be seen in *Figure 3.5*.

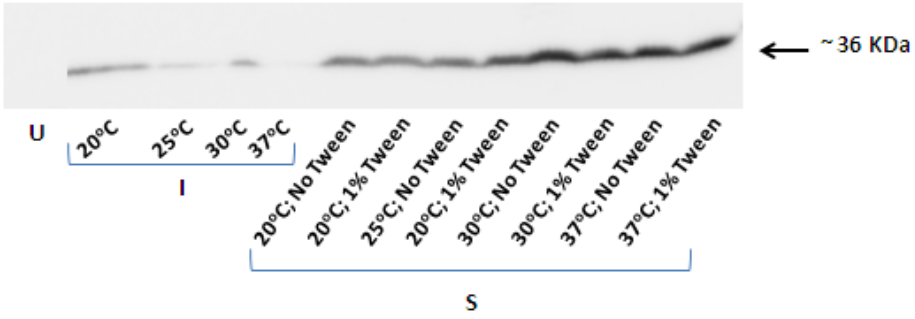


Figure 3.5: Western blot analysis of second Marseillevirus PCNA protein expression study in the pASK_IBA17plus Strep-tag vector**: Western blot showing PCNA protein expression efficiency and levels for the cloned pASK-IBA17plus Marseillevirus PCNA construct, after inducing protein expression in the original culture at four different temperatures and lysing the total protein samples in two different buffers;** The samples were run on a 12.5% resolving gel. The Western blot does indeed reveal the presence of the Marseillevirus PCNA protein being expressed and estimated to be running at approximately 36 kDa, while at the same time it can be pointed out that the protein appears to solubilise somewhat better when grown at 30-37^{o}C. The image was captured using a Kodak processor (UK).

Conclusions

The initial results obtained after running the samples on a SDS-PAGE gel were uninterpretable, as all samples produced identical protein bands of the same intensity. Consequently, the same samples were used in a Western blot experiment were the differences between each sample were more noticeable, with the different growth conditions and lysis buffers having varying effects on the PCNA proteins' expression and solubility. In conclusion; the Western blot revealed that the Marseillevirus PCNA was indeed expressed and running at the estimated size of ~36 kDa. As a result, it was decided to continue trying to optimise the protein expression and solubilisation conditions of the Marseillevirus PCNA.

Third Marseillevirus PCNA protein expression study in the pASK_IBA17plus Strep-tag vector:

Another midi-scale preparation of the Rosetta2 (DE3) (pLysS)_pASK-IBA17plus_Marseillevirus (MAR_ORF212)_PCNA plasmid was prepared. The aim this time was to separate the final bacteria culture, which had already been induced for PCNA protein expression for 4 hours at 37^{o}C, into a total of six equal volume samples. The cells harvested from each equally separated sample were resuspended in different lysis buffers (*see also Table 2.9*). All buffers set up had a Tris-base but contained varying amounts of salt (NaCl), from 150mM to 500mM (also testing an intermediate of 250mM), while the bacteria cell samples were also treated with 1% or 2% Tween detergent. The samples prepared were utilised in a Western blot experiment, and the proteins of interest were detected with the use of Strep-Tactin™ HRP conjugate (*Figure 3.6*).

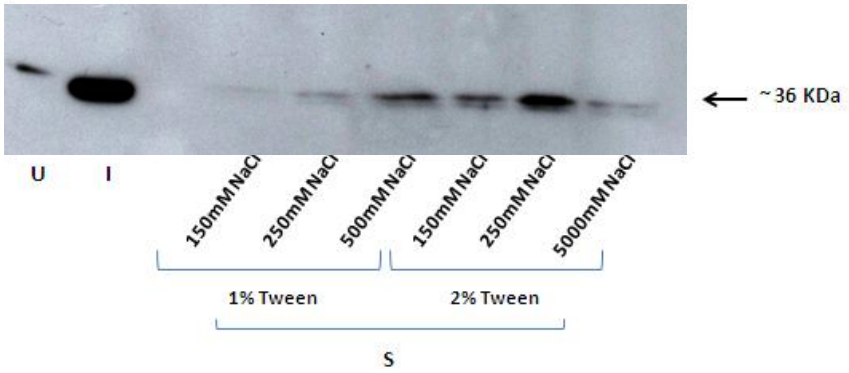


Figure 3.6: Western blot analysis of third Marseillevirus PCNA protein expression study in the pASK_IBA17plus Strep-tag vector**: Western blot showing PCNA protein expression efficiency and levels for the cloned pASK-IBA17plus Marseillevirus PCNA construct, after inducing protein expression for 4 hours at 37^{o}C and lysing the total protein samples in six different buffers;** Lysis buffers: 100mM Tris-HCl pH8, 150/250/500mM NaCl, 1mM EDTA, 1% or 2% Tween. The samples were run on a 10% resolving gel. The Western blot under examination indicates that the Marseillevirus PCNA protein solubilises a bit better when lysed in a buffer containing 250mM NaCl and 2% Tween.

Conclusions

The various samples produced during this experiment were once again utilised in a Western blot procedure were the results were rather positive, as it was revealed that different lysis buffers had a diverse effect on the PCNA proteins' solubility, with some buffers being responsible for greatly improved protein solubility. However, the total levels of PCNA protein expression and solubility were not sufficient for use in further biochemical analysis of this protein, thus it was concluded to clone the Marseillevirus PCNA sequence in a different vector containing another protein tag anticipating to finally obtain greater levels of protein expression and solubility that could be applied in further experiments.

3.1.2.1.2 Summary of Results

The first Marseillevirus PCNA protein expression experiment revealed the presence of a protein band running at the estimated and anticipated size of ~35-36 kDa. This was a very positive result as it was expected that the specific band quite likely represented the Marseillevirus PCNA protein. However, even if that was indeed the case, it was obvious from the results obtained that the protein was only present in very low levels and probably not that soluble. As a consequence, during the second Marseillevirus PCNA protein expression experiment different growth conditions were tested so as to determine their effect on the protein expression of the Marseillevirus PCNA, while at the same time two different lysis buffers were used for protein solubilisation; one containing detergent and another one not containing any detergent. This method did not generate high-quality results

as all samples produced similar protein band patterns making the data uninterpretable. However, a Western blot did reveal the presence and expression of the Marseillevirus PCNA protein. Finally, the third Marseillevirus PCNA protein expression experiment, during which a greater variety of lysis buffers were tested for the purpose of once again increasing the proteins' solubility, confirmed that even though different buffers had a varying effect on the PCNA proteins' solubility, nevertheless they did not result in greatly improved protein solubilisation levels hence not allowing for further biochemical analysis of this protein. Consequently, the protein was cloned into a different vector hoping for more positive results.

3.1.2.2 Cloning of the Marseillevirus PCNA into the pEHISTEV 6xHis-tag vector for Protein Expression Experiments

(*See Section 2.2.2.1.2*)

The resulted construct was pEHISTEV_Marseillevirus(MAR_ORF212)_PCNA (*Appendix A.3.2*)

3.1.2.2.1 Protein Expression and Purification experiments for the Marseillevirus PCNA, cloned into the pEHISTEV 6xHis-tag vector and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

(*See Sections 2.2.2.2 and 2.2.2.3*)

First Marseillevirus PCNA protein expression study in the pEHISTEV 6xHis-tag vector:

For the protein expression procedure of the Rosetta2 (DE3) (pLysS)_pEHISTEV_Marseillevirus (MAR_ORF212)_PCNA plasmid, refer to the equivalent 'Mimivirus PCNA protein expression' section, while the results can be seen in *Figure 3.2*.

The soluble protein samples for the Marseillevirus PCNA protein (as seen in *Figure 3.2*), were also used to perform a pull-down, with the use of Ni-NTA agarose beads (*See Figure 3.3*).

3.1.2.2.2 Summary of Results

The slightly distinct band running at ~35 kDa assumed to be the Marseillevirus PCNA, previously seen in the protein expression induced sample in *Figure 3.1*, was also visible in the same sample in *Figure 3.2*. In this experiment, however, a band of the same size and intensity was additionally visible in the insoluble 'pellet' sample. This outcome, as well as the fact that no band of interest was observed after performing a Ni-NTA agarose bead pull-down experiment (*Figure 3.3*), led to the conclusion that the Marseillevirus PCNA protein was in general insoluble. As a consequence, no further experiments were conducted for the PCNA protein belonging to this organism.

3.1.3 CroV

3.1.3.1 Cloning of the CroV PCNA into the pASK_IBA17plus Strep-tag vector for Protein Expression Experiments

(*See Section 2.2.2.1.1.*)

The construct generated was pASK-IBA17plus_CroV(CroV_219)_PCNA (*Appendix A.2.3*).

3.1.3.1.1 Protein Expression and Purification experiments for the CroV PCNA, cloned into the pASK_IBA17plus Strep-tag vector and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

First CroV PCNA protein expression study in the pASK_IBA17plus Strep-tag vector:

For the protein expression protocol of the Rosetta2 (DE3) (pLysS)_pASK-IBA17plus_CroV (CroV_219)_PCNA plasmid, refer to the appropriate 'Mimivirus PCNA protein expression' section. The results can be seen in *Figure 3.1*.

The samples depicted in *Figure 3.1* for the pASK-IBA17plus_CroV (CroV_219) PCNA were additionally used in a Western blot experiment, during which the Strep-tagged PCNA protein was directly detected by Strep-Tactin™ HRP conjugate. The results of this experiment can be seen below (*Figure 3.7*).

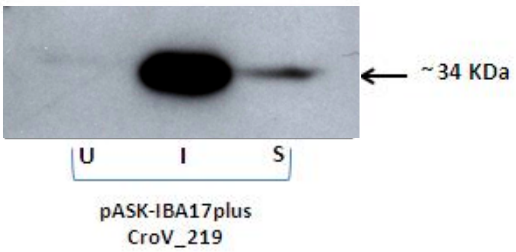


Figure 3.7: Western blot analysis of first CroV PCNA protein expression study in the pASK_IBA17plus Strep-tag vector**: Western blot showing PCNA protein expression efficiency and levels for the cloned pASK-IBA17plus CroV PCNA construct, after inducing protein expression for 4 hours at 37^{o}C;** The samples were run on a 10% resolving gel (as seen in *Figure 3.1*). From the Western blot it can be concluded that even though the CroV PCNA protein seems to be expressed at a high level, nevertheless it is not significantly soluble (less than ¼ of the expressed PCNA protein has actually solubilised when sonicated in the particular lysis buffer).

Conclusions

As seen in *Figure 3.1*, the CroV protein expression band patterns are highly comparable to the protein band patterns generated by the empty Rosetta2(DE3)(pLysS)_pASK-IBA17plus vector. Furthermore, as was the case for the Mimivirus, no intensively distinct protein band running at ~35 kDa (i.e. expected PCNA protein molecular mass) could be observed. However, when the exact same samples were used in a Western blot experiment, the results were clearer and fairly positive; the CroV PCNA was indeed being expressed and running at the anticipated size, but was not exceptionally soluble. Consequently, the next stages of the project involved trying to optimise the CroV PCNA proteins' expression and solubilisation conditions by testing a range of different growth conditions and lysis buffers.

Second CroV PCNA protein expression study in the pASK_IBA17plus Strep-tag vector:

This time, a midi-scale preparation of the Rosetta2 (DE3) (pLysS)_pASK-IBA17plus_CroV (CroV_219)_PCNA plasmid was prepared. During this second CroV PCNA protein expression attempt, the main aim was to test a variety of different bacteria growth temperatures, as well as different lysis buffers for resuspension of the CroV bacteria plasmid cells.

More specifically, the original bacteria culture was divided into four equal amounts just after inducing protein expression. Each equally separated sample was grown under different temperature conditions, and hence grown for different time periods. The temperatures chosen were once again 20^{o}C (overnight protein expression), 25^{o}C (overnight protein expression), 30^{o}C (4 hour protein expression) and 37^{o}C (4 hour protein expression). Following the completion of CroV PCNA protein expression induction, the bacteria cells from each sample were harvested. Each such sample was however further divided into nine samples, which were subsequently resuspended in various lysis buffers (*see also Table 2.9*). The buffers had a Tris-base but all of them varied in the amount of salt (NaCl) they contained [NaCl concentrations ranged from 150mM to 500mM, testing also at 250mM]. In addition, the bacteria cell samples were or were not treated with 1% or 2% Tween detergent (i.e. a 'No Tween' lysis buffer was also prepared). The samples prepared were directly used in a Western blot experiment, where the target proteins were detected by using Strep-Tactin™ HRP conjugate. Even though prominent signals were produced in all

cases, even for the un-induced protein sample, no figure is shown as no definite conclusions could be drawn from this experiment, a reason being that the different temperatures and lysis buffers tested did not appear to have a distinguishable effect on the proteins' expression and solubility, respectively.

All the soluble protein samples resuspended in 250mM NaCl and treated with no, 1% or 2% Tween were subsequently run on a 10% SDS-PAGE gel, for the purpose of checking if more obvious protein bands could be detected and the difference between them determined. In the meantime, the differently lysed bacteria cultures, for which PCNA protein expression was induced at 25^{o}C overnight and which were thought to have produced somewhat the best signals during the Western blot experiment, were repeated, exactly the same way as described above, and run on a 10% SDS-PAGE gel. However, no clear differences between the bands generated could once again be distinguished for either SDS-PAGE carried out (*no figures included*).

Some of the soluble protein samples, which were previously used for the Western blot and SDS-PAGE analysis, were also used for a streptavidin pull-down (*See Section 2.2.2.4.1*). No figure is attached because the streptavidin pull-down did not work optimally. This was the second Strep-tag pull-down attempt for a specific pASK-IBA17plus PCNA construct. As a result, the question was still whether or not the Strep-tag attached to the CroV PCNA, and generally the Strep-tag found in all three pASK-IBA17plus PCNA constructs, was in some way obscured hence affecting PCNA protein expression and solubilisation.

Conclusions

The results obtained after running the samples prepared on both SDS-PAGE gels and Western blots were unfortunately inconclusive, as all samples produced identical protein band patterns of the same intensity and hence no relative differences could be distinguished between them. Nevertheless, the Western blot experiments did provide some evidence of the presence and expression of the CroV PCNA protein. This outcome had as a result the cloning of the CroV PCNA protein sequence into a different vector containing another protein tag, thus expecting to achieve greater and more distinguishable levels of protein expression and solubility.

3.1.3.1.2 Summary of Results

The first CroV PCNA protein expression experiment, during which the samples prepared were run both on a SDS-PAGE gel and a Western blot, exposed the CroV PCNA protein being expressed and running at the anticipated size of ~34-35 kDa. However, the results also revealed that the protein was not exceptionally soluble. Therefore, for the second CroV PCNA protein expression experiment a variety of different growth conditions and lysis buffers were tested so as to determine their varying effect on the solubility of the CroV PCNA protein. The samples were once again run on both a SDS-PAGE gel and a Western blot; similar results were obtained. As a consequence, the protein was cloned into a different vector hoping for more positive results.

3.1.3.2 Cloning of the CroV PCNA into the pEHISTEV 6xHis-tag vector for Protein Expression Experiments

(*See Section 2.2.2.1.2*)

The construct generated was pEHISTEV_CroV(CroV_219)_PCNA (*Appendix A.3.3*).

3.1.3.2.1 Protein Expression and Purification experiments for the CroV PCNA, cloned into the pEHISTEV 6xHis-tag vector and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

(*See Sections 2.2.2.2 and 2.2.2.3*)

First CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector:

For the protein expression procedure of the Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV (CroV_219)_PCNA plasmid, again refer to the corresponding 'Mimivirus PCNA protein expression' protocol. The results can be seen in *Figure 3.2*.

In *Figure 3.2*, specifically for the pEHISTEV_CroV (CroV_219), the presence of a pronounced band in the insoluble protein sample (i.e. 'p'ellet sample) at around ~35-36 kDa was evident. Therefore, it was presumed that this band was indeed the CroV PCNA protein being expressed but not being solubilised. To verify this assumption, further

protein expression studies were conducted, while at the same time different conditions were tested so as to try to improve the CroV PCNA protein solubility.

The soluble protein samples for the CroV PCNA protein (as seen in *Figure 3.2*), were also used to perform a pull-down, with the use of Ni-NTA agarose beads (*See Figure 3.3*).

Conclusions

Similarly to the Marseillevirus; a previously undetectable band running at ~35 kDa was visible for the CroV in the insoluble 'pellet' sample, as seen in *Figure 3.2*, leading to the initial consideration that the CroV PCNA protein may also be insoluble. In contrast though to the Marseillevirus; the CroV Ni-NTA agarose bead pull-down experiment produced rather contradictory results to the previous ones discussed, specifically revealing a distinct and prominent band of the anticipated size also in the soluble sample purified by a 'pull-down' technique. These results led to the assumption that the CroV PCNA protein could be significantly purifiable under the correct conditions and hence, following this outcome, numerous attempts were made to optimise the CroV PCNA protein expression and solubilisation conditions.

Second CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector (first Ni-NTA agarose bead column purification attempt):

The same procedure as carried out during the 'First CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector' was performed, i.e. the IPTG protein induced bacteria culture was grown overnight at 25^{o}C. The difference was that for this protein expression experiment, a different lysis buffer was used for the resuspension of the harvested CroV PCNA bacteria plasmid cells. This was conducted for the purpose of evaluating the effect of the new lysis buffer on CroV PCNA protein solubility. The buffer, which was Phosphate-based, was named Buffer A and can be seen in *Tables 2.1 and 2.9* (*i.e. 50mM Na2HPO4, 400mM NaCl, 10mM Imidazole, 1mM β-mercaptoethanol, pH8*).

Following the sonication of the total CroV PCNA protein extract, which had previously been resuspended in Lysis Buffer A, the soluble protein sample was collected by centrifugation and filtered through a column containing Ni-NTA agarose beads, attempting a small-scale PCNA protein purification (*For protocol see Section 2.2.2.4.2*). The column had been equilibrated with Buffer A prior to starting the experimental procedure. The flow-through protein sample was collected for further analysis.

Once the soluble CroV PCNA protein extract was filtered through the Ni-NTA agarose bead column, the column was washed with a second buffer (i.e. Buffer B: *50mM*

Na2HPO4, 400mM NaCl, 30mM Imidazole, 1mM β-mercaptoethanol, pH8) and the resulting sample was collected for further analysis.

Finally, the proteins bound to the Ni-NTA agarose beads were eluted with a third buffer (i.e. Buffer C: *50mM Na2HPO4, 400mM NaCl, 300mM Imidazole, 1mM β-mercaptoethanol, pH8*). Different fractions of this sample were collected and maintained for further analysis.

All samples collected throughout this procedure were run on an SDS-PAGE gel (*no figure displayed*). Nevertheless, the attempted protein purification of the CroV PCNA protein was unsuccessfully and no conclusive outcome could be deduced, apart from the fact that the CroV PCNA protein did not preferentially bind to the column set up and, in addition, was not greatly soluble.

Conclusions

The samples prepared during this experiment, and which were directly utilised in a Ni-NTA agarose bead packed column pull-down experiment, were run on a SDS-PAGE gel where it was apparent that the new phosphate-based lysis buffer had no improved effect on the proteins' solubility, as no protein could be solubilised whatsoever. Specifically, it was determined that most proteins did not bind to the column as the majority of them were

eluted at once. One of the eluted proteins, which produced a prominent band on the gel, seemed to be running at the PCNA anticipated size of ~35 kDa. This fact led to the conclusion that even if that was the protein of interest, then it was significantly insoluble. However, even though this experiment did not generate encouraging results, the optimisation of the CroV PCNA protein expression conditions was continued as previous results were rather optimistic.

Third CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector:

Another midi-scale preparation of the Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV (CroV_219)_PCNA plasmid was prepared. The aim this time though was to separate the final bacteria culture, which had similarly been induced for PCNA protein expression at 25^{o}C overnight, into a total of four equal volume samples. The bacteria cells harvested from each equally separated sample were resuspended in different lysis buffers (*see also Table 2.9*). This was once again the Phosphate-base buffer, but this time, containing varying concentrations of Imidazole, from 10mM to 40mM, while a 'No Imidazole' equivalent buffer was also prepared. The samples were sonicated and the soluble protein extracts were run on a 10% SDS-PAGE gel, which revealed fairly noticeable bands in all the soluble samples at around the predicted molecular weight for the CroV PCNA protein (*no figure available*). Furthermore, the soluble protein samples were used to perform a pull-down of the PCNA protein, hoping to achieve higher levels of protein solubility. The results of this Ni-NTA agarose bead pull-down, as well as the results from the whole

protein expression experiment, however were poor and no confident conclusions could be reached.

Conclusions

All samples obtained during this Ni-NTA agarose bead pull-down experiment were run on a SDS-PAGE gel. The results were rather inconclusive as all soluble protein samples produced identical protein band patterns of the same intensity, revealing no obvious difference between the lysis buffers employed. However, when observing the Ni-NTA agarose bead purified protein samples run on the gel it could be stated that the lysis buffer with the highest imidazole concentration produced slightly clearer and maybe more positive results (i.e. more soluble protein). Therefore, it was decided to continue with the phosphate-based lysis buffer containing a high amount of imidazole for any following experiments, while continuing to try and optimise the expression and solubilisation conditions of the CroV PCNA protein.

Fourth CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector (plus pEHISTEV empty vector):

After evaluating the results obtained from the previous experiment, another midi-scale preparation of the Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV (CroV_219)_PCNA plasmid was prepared. This time, however, the same procedures were concurrently carried out for the Rosetta2(DE3)(pLysS)_ pEHISTEV empty vector, for the purpose of using this

as a control sample for comparison reasons. Protein expression induction was undertaken again at 25^{o}C overnight, while the lysis buffer used for resuspension of the PCNA bacteria cells was: *50mM Na2HPO4, 400mM NaCl, 40mM Imidazole* (considered to produce the best results compared to other amounts of Imidazole, according to the previous experiment)*, 1mM β-mercaptoethanol, pH8*. A fraction of each soluble protein sample was used to carry out a pull-down experiment with Ni-NTA beads, which specifically bind to the 6xHis affinity tag found in the pEHISTEV vector. Specifically, 20μl of the beads were directly added into each sample, the samples were incubated at 4^{o}C on a wheel for one hour, they were then washed 4-6 times with the lysis buffer previously used for sonication and finally were treated by boiling them at 95^{o}C for 5 min in 2xSB buffer. A description of the exact pull-down procedure undertaken can be found in *Section 2.2.2.4.2; Pull-down of PCNA Solubilised protein samples for the Mimivirus, Marseillevirus and CroV, with the use of Nickel (Ni-NTA) affinity agarose beads (only for the pEHISTEV PCNA constructs)*. The results of these experiments can be seen in *Figure 3.8*.

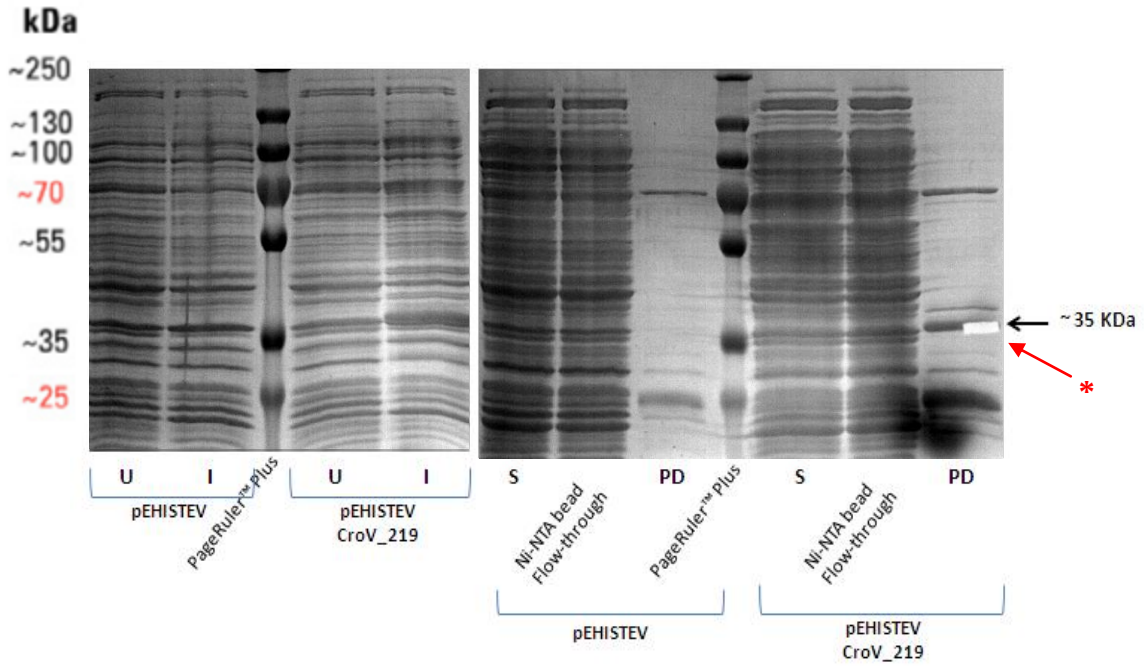


Figure 3.8: SDS-PAGE analysis of fourth CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector (plus pEHISTEV empty vector): **10% SDS-PAGE Gel Photograph showing PCNA protein expression efficiency and levels for the cloned pEHISTEV CroV PCNA construct, compared to the empty pEHISTEV vector. Both bacteria plasmid samples were induced for protein expression at 25°C overnight and the bacteria cells lysed in four different buffers. The different soluble protein extracts were used to perform a pull-down;** (U): uninduced, (I): induced, (S): soluble, (PD): pull-down. Once more, the anticipated CroV PCNA protein molecular mass was 34 kDa.

Note: The band represented by the red arrow (*) was sent for Mass Spectrometry, to determine whether or not it is the CroV PCNA protein.

Conclusions

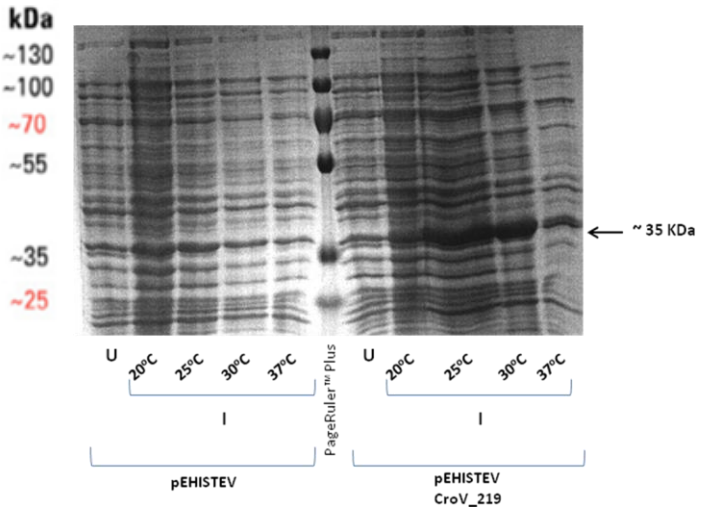
The results of this experiment were rather unexpected as no alterations to the previous protein expression and solubilisation conditions were undertaken. The only differing factors were that this Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV (CroV_219)_PCNA plasmid preparartion was performed at a slightly larger-scale than the previous one, and also the use of the empty vector as a control sample throughout the method. In particular,

the soluble PCNA protein samples were once again utilised in a Ni-NTA agarose bead pull-down procedure, which produced a very clear, distinct and prominent band running at the anticipated size for the CroV PCNA. This specific purified sample was subsequently sent for mass spectrometry analysis to verify its nature. The results were positive, but not significant; even though the PCNA protein of interest was the first hit identified by MS, the values corresponding to its score, specific peptides and sequence coverage were poor. Specifically, when using the ProteinPilot™ software system, the following data was retrieved for the protein band of interest analysed by ESI: ProteinPilot Score = 12 (at $p \leq 0.05$, based on the Paragon Algorithm™ used), Sequence Coverage (%) = 32, Significant Peptides (95%) = 6, Contribution of each Peptide = 2 at a 99% Confidence; the data was the same when compared against both the NCBI and BMS protein databases. As a consequence, the presence of the PCNA protein could not be identified with great confidence. Nevertheless, even though the PCNA protein was most likely present in the sample, the amount of protein was not sufficient for use in further biochemical analysis, and thus, the protein expression and solubilisation conditions had to be further optimised for the purpose of obtaining greater levels of clean and pure soluble protein. The exact reason as to why this experiment produced remarkably positive results still remains unknown, however the fact that it was a somewhat larger-scale plasmid preparation, compared to the previous experiment described, may have played a role.

Fifth CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector (plus pEHISTEV empty vector):

The aim of this experiment was to test a variety of different bacteria growth temperatures, as well as different lysis buffers for resuspension of the CroV PCNA bacteria plasmid cells (similarly as was previously performed for the Rosetta2 (DE3) (pLysS)_pASK-IBA17plus_CroV (CroV_219)_PCNA plasmid). This experiments were particularly performed for the Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV (CroV_219)_PCNA plasmid, alongside with the Rosetta2(DE3)(pLysS)_ pEHISTEV empty vector plasmid.

Therefore, as described elsewhere, four different temperature conditions, and hence four different time periods of bacteria culture growth, were examined. Following the completion of protein expression induction, the bacteria cells from each sample were harvested. This time, however, each individual sample was further divided into twelve samples, which were subsequently resuspended in various lysis buffers (*see also Table 2.9*). All buffers had a Phosphate-base but contained varying concentrations of salt (NaCl) (i.e. from 250mM to 600mM, with a 400mM intermediate) and Imidazole (i.e. 40mM and 65mM), while additionally the samples were either treated with 0.5% Tween or were not treated with any detergent whatsoever. All resulting samples were run on a 10% SDS-PAGE gel. Nonetheless, as the different lysis buffers did not have any pronounced effect on the CroV PCNA protein expression efficacy, only a couple of the SDS-PAGE gels that were run will be demonstrated here (*Figure 3.9*).



Protein Expression Induction at 30°C for 4 hours; Soluble (S) Protein Extracts

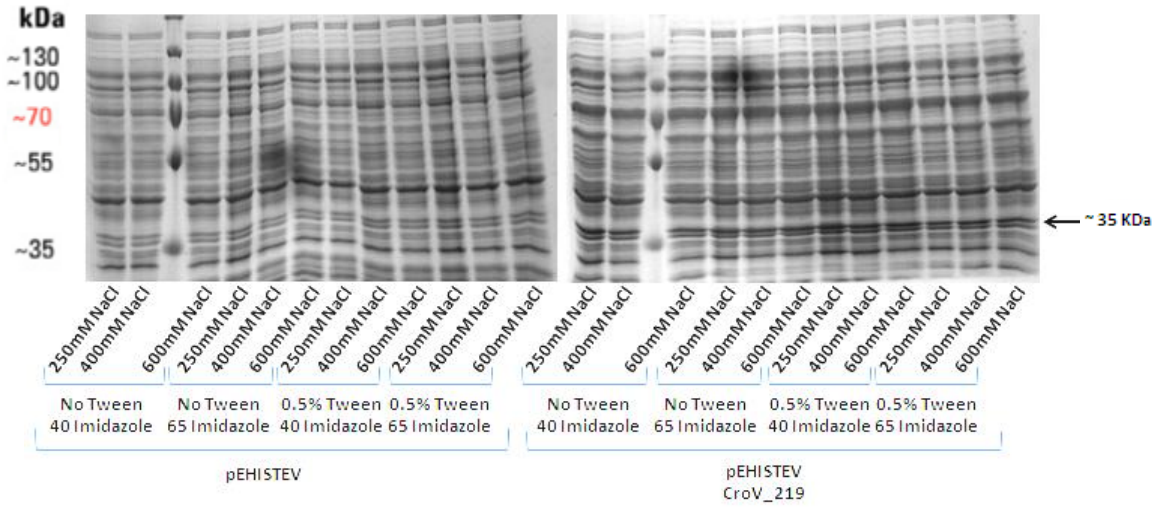


Figure 3.9: SDS-PAGE analysis of fifth CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector (plus pEHISTEV empty vector)**: 10% SDS-PAGE Gel Photograph showing PCNA protein expression efficiency and levels for the cloned pEHISTEV CroV PCNA construct, compared to the empty pEHISTEV vector, after inducing protein expression in the original bacteria cultures at four different temperatures and lysing the total protein samples in twelve different buffers;** Protein expression induction temperatures: 20°C, 25°C, 30°C and 37°C; This gel however only demonstrates the soluble protein extracts representing protein expression induction at 30°C for 4 hours. Lysis buffers: 50mM Na2HPO4, 250/400/650mM NaCl, 40/65mM Imidazole, 1mM β-mercaptoethanol, pH8, No or 0.5% Tween. Even though the anticipated molecular mass for the CroV PCNA protein was 34 kDa, in most gels run during this part of the experiment it appeared to have a somewhat higher molecular weight (i.e. around ~36 kDa). A distinguishable band was observed in the 'I' protein samples, whereas the same band could not easily be distinguished in the 'S' protein samples.

Conclusions

The only conclusion that could be drawn from this experiment was that the various growth temperatures and lysis buffers tested had no distinct effect on the CroV PCNA proteins' solubility. In general, all samples prepared and employed during this study produced exactly the same results (i.e. same pattern of protein bands), hence proving that the optimisation of the CroV PCNA protein expression and solubilisation conditions was more difficult than what expected. For this purpose, a different approach was taken whilst trying to further optimise the proteins' expression conditions. This specifically was studying PCNA protein purification using an *AKTA* purifier and thus hoping for better quality results.

Sixth CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector (for first *AKTA* purification attempt):

A large-scale preparation of the Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV (CroV_219)_PCNA plasmid was set up as usual. The bacteria culture was left to incubate at $27^{o}C$ overnight, after protein expression induction by IPTG. The final bacteria culture was divided into two equal volume samples; for each, the cells were harvested by centrifugation and resuspended in two different Lysis Buffers A (*Table 2.1 & 2.9*; *50mM Na2HPO4, 250/500mM NaCl, 40mM Imidazole, 0.5% Tween, 1mM β-mercaptoethanol, pH 8*). The total cell extracts were finally sonicated and the soluble cell extracts collected for a one-step PCNA protein purification analysis on the *AKTA* system allowing for further assessment of the CroV PCNA protein expression. The column utilised during this first

PCNA *AKTA* purification was a histidine binding His-Trap™ column (*See Section 2.2.2.4.4; Protein Purification Experiments for CroV PCNA using an AKTA system*). The samples of interest were run on a 10% SDS-PAGE gel (*no figure attached*).

Conclusions

The conclusion from this *AKTA* purification experiment for the CroV PCNA was that, even though there clearly was a satisfactory level of induced protein expression, the purification of the protein was completely unsuccessfully. That was the case for both lysis buffers examined. As a result, more CroV PCNA protein expression conditions were investigated this time by performing a gradient *AKTA* purification experiment and testing an even greater range of lysis buffers containing increasing concentrations of Imidazole.

Seventh CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector (for second *AKTA* purification attempt):

Similarly to what was mentioned beforehand, a midi-scale preparation of the Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV (CroV_219)_PCNA plasmid was set up. The bacteria culture, however, was left to incubate at $25^{o}C$ overnight (instead of $27^{o}C$) so as to induce PCNA protein expression. The bacteria cells were harvested and resuspended in Lysis Buffer A (*Table 2.1 & 2.9*; *50mM Na2HPO4, 250mM NaCl, 40mM Imidazole, 1mM β-mercaptoethanol, pH 8*). The lysis buffer containing 250mM NaCl, instead of 500mM NaCl, was selected for this experiment, after carefully evaluating the results observed in

the previous experiment. The total cell extracts were finally sonicated and the soluble cell extracts collected for PCNA protein purification analysis on the *AKTA* system allowing once again for further assessment of the CroV PCNA protein expression. The *AKTA* method set up was a gradient purification protocol, testing a range of Imidazole concentrations from 40mM to 500mM. The column utilised during this second PCNA *AKTA* purification attempt was the same His-Trap™ column (*See Section 2.2.2.4.4*). The samples considered to be of most importance were run on a 10% SDS-PAGE gel (*no data shown*). The purification of CroV PCNA protein was once again unsuccessful.

The same samples run on the SDS-PAGE gel were subsequently used to perform a Western blot; the reason being to check if a stronger CroV PCNA protein signal could be detected. For this Western blot method, in particular, two antibodies were utilised making it a two-step probing/detection procedure; initially the blocked membrane was incubated with a primary anti-HIS mouse monoclonal antibody that recognised and bound to the 6xHis-tag sequence, while the PCNA protein was further detected with the use of a secondary anti-mouse antibody specific to the primary antibody. However, no Western blot figure is attached as a strong chemiluminescence signal was only produced for the 'I'nduced CroV PCNA protein sample (i.e. total cell extract), while no other band(s) for any of the other samples were actually visible. The band generated was estimated to have a molecular mass of ~37 kDa (appeared between 35 kDa and 40 kDa, when compared to the protein ladder).

Conclusions

The gradient *AKTA* purification method also proved unsuccessful, resulting in a similar outcome to the first *AKTA* purification experiment attempted. This result was also confirmed by performing a Western blot. Therefore, no further *AKTA* purifications were attempted for the CroV PCNA. Instead it was decided to attempt a final Ni-NTA agarose bead pull-down experiment under the same protein expression conditions.

Eighth CroV PCNA protein expression study in the pEHISTEV 6xHis-tag vector (second Ni-NTA agarose bead column purification attempt):

A Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV (CroV_219)_PCNA plasmid bacteria culture was grown as usual. CroV PCNA protein expression induction was performed at 30^{o}C for 4 hours. The bacteria cells were harvested and resuspended in Lysis Buffer A (*Table 2.1 & 2.9*; *50mM Na2HPO4, 250mM NaCl, 40mM Imidazole, 1mM β-mercaptoethanol, 0.5% Tween, pH 8*). The total cell extracts were sonicated and the soluble cell extracts collected for PCNA protein purification on a Ni-NTA agarose bead column (*See Section 2.2.2.4.2*). The column had previously been equilibrated with the same Lysis Buffer A. The flow-through protein sample was collected. Next the column was washed with Lysis Buffer A that only contained 15mM Imidazole. The low Imidazole concentration buffer was used so as to prevent any CroV PCNA protein loss from the column. The wash-through protein sample was collected. Finally, the proteins bound to the Ni-NTA column were eluted with Lysis Buffer A that contained a much higher concentration of Imidazole (i.e. 300mM), to ensure all bound proteins would elute.

Different fractions of this elution process were collected and maintained for further analysis by SDS-PAGE. The outcome of this CroV PCNA protein purification method was not informative, as there was no obvious induction of protein expression.

Conclusions

This Ni-NTA agarose bead pull-down experiment generated contrasting results to the previous two *AKTA* purification attempts. More specifically, even though during the earlier *AKTA* purification studies PCNA protein expression induction was clearly noticeable, in the case of this experiment it was obvious that no protein expression was being induced. These results were rather confusing and contradictory to other positive results previously found, so as a consequence it was decided to change approach and instead express the specific CroV PCNA construct in a variety of different *E.coli* expression systems. By doing so it was anticipated that the different *E.coli* expression cells would have a varying effect on the expression and hopefully solubility of the CroV PCNA protein, thus allowing for more positive results and further biochemical analysis of the specific protein.

3.1.3.2.2 Summary of Results

The first CroV PCNA protein expression experiment revealed interestingly some positive results. Specifically, it was established that a good proportion of the PCNA protein being expressed could probably also be solubilised, even though a good majority of it still appeared to be insoluble. As a result, multiple attempts of optimising the proteins'

expression and solubilisation conditions were carried out, primarily for the purpose of obtaining greater quantities of the solubilised CroV PCNA protein for use in further biochemical analysis. The first such attempt evaluated the effect of a new lysis buffer on the CroV PCNA proteins' solubility; the new buffer was Phosphate-based, while all previous buffers utilised were Tris-based. The PCNA protein sample prepared was directly utilised in a Ni-NTA agarose bead packed column pull-down experiment, during which process a number of samples were collected for further analysis of the CroV PCNA protein solubility. The result of this study was that no protein actually bound to the column, suggesting either that the protein was obscured or that the column was problematic. However, in the eluted protein fraction collected, a protein running at the anticipated size was actually visible on the SDS-PAGE gel, suggesting that if it was the CroV PCNA protein then it was significantly insoluble. Following this, a variety of phosphate-based lysis buffers were tested so as to determine their probable varying effect on the solubility of the CroV PCNA protein; the differing factor between all buffers was the concentration of Imidazole reagent, while a buffer with no Imidazole was also prepared for use as a control. No great difference could be distinguished between the different samples, but it did appear that buffers with greater Imidazole concentrations may have produced slightly better results compared to the rest. The fourth CroV PCNA protein expression experiment generated the best results with a very prominent protein band being expressed at the anticipated PCNA size. This band was specifically seen in the purified protein sample. The mass spectrometry results showed that the PCNA protein was quite possibly present, but in very low levels within the sample. Therefore, the optimisation of the proteins' expression and solubilisation conditions was continued. The first attempt was to examine a range of

growth conditions and lysis buffers. Nevertheless, all conditions chosen appeared to have the same effect on the CroV PCNA proteins' expression and solubility. At this stage, it was decided to try and purify the protein with the use of an *AKTA* purifier. Nevertheless, both attempts failed, even though some of the protein was noticeably being expressed. The final attempt to purify the CroV PCNA protein was with the use of another Ni-NTA agarose bead column. This experiment did not produce any results. The optimisation of the CroV PCNA proteins' expression and solubilisation proved very difficult and unsuccessful. In general, it can be stated that the CroV PCNA protein was mostly insoluble, with only a few lysis buffers slightly affecting and maybe increasing its solubilility. The amounts of soluble protein obtained, however, were considered insignificant for use in further experiments.

3.1.3.2.3 Protein Expression and Purification experiments for the CroV PCNA, cloned into the pEHISTEV 6xHis-tag vector and expressed in a variety of *E.coli* cells

First CroV PCNA protein expression study in a variety of *E.coli* cells:

The *E.coli* cells selected for this part of the project were: BL21(DE3), BL21(DE3)(pLysS), Rosetta2(DE3), Arctic Express(DE3)RP and C43(DE3). More details of these protein expression strains can be found in *Section 2.2.2.1.2*. The main purpose of testing other *E.coli* host strains, apart from the usual Rosetta2(DE3)(pLysS), was to verify whether or not they would have a different effect on PCNA protein expression and solubility.

Small-scale bacteria cultures of the pEHISTEV_CroV (CroV_219)_PCNA plasmid, expressed in each of the BL21(DE3), BL21(DE3)(pLysS), Rosetta2(DE3), Arctic Express(DE3)RP and C43(DE3) *E.coli* host strains, were set up following standard techniques. CroV PCNA protein expression was induced at 37^{o}C for 4 hours. The CroV PCNA bacteria cells were harvested and resuspended in a PBS-based buffer, additionally containing *100mM Tris-HCl pH8, 10mM Imidazole, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol* (*See Table 2.1 and 2.9*). No data is shown, as the different *E.coli* host strains did not seem to have any noticeable effect on the PCNA protein expression efficiency.

Conclusions

The general conclusion from this experiment was that the different *E.coli* expression systems used in the hope of increasing the CroV PCNA proteins' expression and solubility, did not actually affect in any way those parameters; meaning that all the different bacteria expression cells utilised in fact produced comparable results none of which were significant. However, it should be noted that a distinct band of the anticipated size for the PCNA protein was evident on the gels, but this band was present in all the samples. Thus, this band was considered to be of no importance. Following this experiment, and as the outcome was not promising, a final attempt was made to try and optimise the conditions of the CroV PCNA proteins' expression and solubility, while using the different *E.coli* expression strains.

Second CroV PCNA protein expression study in a variety of *E.coli* cells:

The main scope of this second study was to investigate different protein expression induction temperatures for the CroV PCNA protein, previously expressed in a selection of *E.coli* host strains. The preferred expression constructs chosen for use in this experiment were the BL21(DE3)(pLysS), Arctic Express(DE3)RP and C43(DE3) pEHISTEV_CroV (CroV_219)_PCNA plasmids.

Those three CroV PCNA plasmids, to begin with, were grown on a rather small-scale preparation, exactly as mentioned beforehand. The difference in this case being that the CroV PCNA protein expression was induced at two different temperatures (i.e. 16^{o}C and 25^{o}C) overnight. In particular, the three different bacteria cultures after being induced for PCNA protein expression were divided into two equal volume samples; each sample was separately grown at the precise temperatures mentioned. After the CroV PCNA protein expression induction was finalised, the bacteria cells from each differently grown sample were harvested and resuspended in the same PBS-based buffer (*See Table 2.1 and 2.9*). The protein samples of interest were run on a SDS-PAGE gel, but once again the different *E.coli* host strains did not appear to have an evident effect on the PCNA protein expression efficiency (*no figure included*).

Conclusions

Similarly to the experiment discussed above, the different *E.coli* expression strains created, containing the CroV PCNA protein, did not have any positive effect on the proteins' expression and solubility, even when two different temperatures were tested for protein expression induction. All samples once again generated exactly the same results, which were uninterpretable and trivial.

3.1.3.2.4 Summary of Results

Both protein expression studies attempted with the various *E.coli* expression strains created, and containing the CroV PCNA protein, were uninformative. The different strains all produced exactly the same results; they specifically all generated the same protein band pattern and no difference could be distinguished between them. This was the case even during the second CroV PCNA protein expression experiment when two different protein expression induction temperatures were tested for the purpose of optimising the proteins' expression and solubilisation conditions. Therefore, no further experiments were performed using these strains. At this point, it was decided to focus for a while on another protein, the FEN1, which interacts strongly with the PCNA protein under the right conditions. The purpose and hypothesis being that through the formation of an interaction between the two proteins of interest, maybe it would be plausible to isolate and solubilise our primary protein of interest, the PCNA.

3.2 Tagged Protein Expression of Giant Virus FEN1 in *E.coli*

3.2.1 CroV

The CroV FEN1 (*CroV_037, Appendix A.1.4;* 37 kDa) gene fragment of interest was cloned into two fairly similar expression vectors both containing a 6xHis affinity tag (2 kDa); the first such vector was the typical pEHISTEV vector, while the other one was the equivalent pEHISGFPTEV vector that, in addition to the 6xHis-tag, also contained a GFP-tag (~28 kDa).

3.2.1.1 Cloning of the CroV FEN1 into the pEHISTEV 6xHis-tag and pEHISGFPTEV 6xHis-tag + GFP-tag vectors for Protein Expression Experiments

Refer to *Section 2.2.3.1.1; Cloning of the CroV FEN1 protein into a pEHISTEV and pEHISGFPTEV vector backbone, both containing a 6xHis affinity tag for protein purification, while the latter also contains a GFP-tag*.

The protocol resulted in the generation of two different constructs: pEHISTEV_CroV_037_FEN1 and pEHISGFPTEV_CroV_037_FEN1 (*Appendix A.4*).

3.2.1.1.1 Protein Expression and Purification experiments for the CroV FEN1, cloned into the pEHISTEV 6xHis-tag and pEHISGFPTEV 6xHis-tag + GFP-tag vectors and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

Refer to *See Section 2.2.3.2; Protein Expression Experiments for CroV FEN1 protein*, and *Section 2.2.3.3; Preparing Soluble FEN1 protein samples from CroV by Sonication procedures*.

First CroV FEN1 protein expression study in the pEHISTEV 6xHis-tag and pEHISGFPTEV 6xHis-tag + GFP-tag vectors:

Midi-scale preparations of the Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV_037_FEN1 and Rosetta2 (DE3) (pLysS)_pEHISGFPTEV_CroV_037_FEN1 plasmids were incubated for ~3 hours at 37^{o}C, until the OD of the bacteria cultures reached the expected point. At that specific OD, FEN1 protein expression was induced in both plasmids by the addition of IPTG reagent. The bacteria cultures were left to incubate at 25^{o}C overnight. The different plasmid bacteria cells were harvested by centrifugation and were resuspended in Lysis Buffer A (*Table 2.1 & 2.10*; *50mM Na2HPO4, 300mM NaCl, 20mM Imidazole, 1mM β-mercaptoethanol, pH 8*). The total cell extracts for both plasmid samples were sonicated and the soluble cell extracts collected for further FEN1 protein expression analysis. Moreover, the two different soluble protein samples, resulting from the two different expression vectors, were used to perform a pull-down of the FEN1 protein with Ni-NTA agarose beads (*For protocol see Section 2.2.3.4.1; Pull-down of FEN1 Solubilised protein samples for the CroV, with the use of Nickel (Ni-NTA) affinity agarose beads (pEHISTEV*

& pEHISGFPTEV FEN1 constructs). All the protein expression samples of interest, as well as the ones resulting from the pull-down process, were run on a 10% SDS-PAGE gel (*Figure 3.10*).

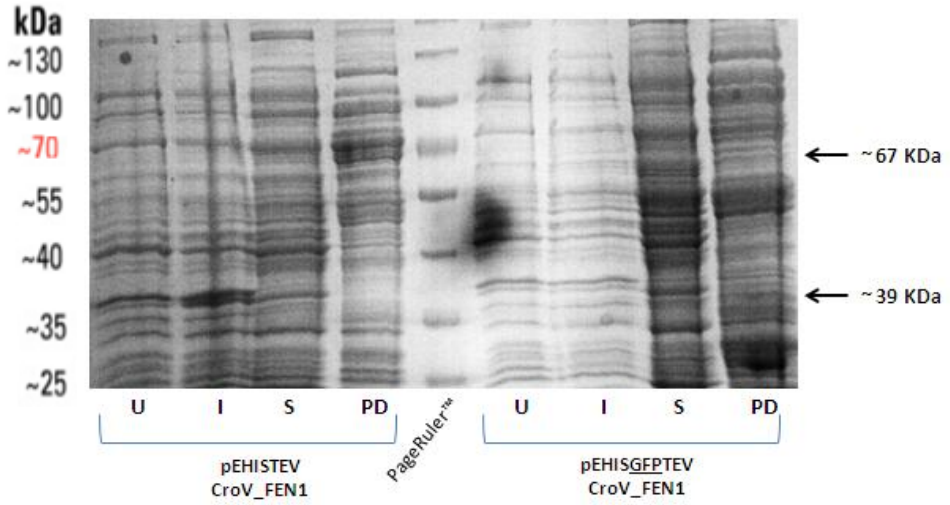


Figure 3.10: SDS-PAGE analysis of first CroV FEN1 protein expression study in the pEHISTEV 6xHis-tag and pEHISGFPTEV 6xHis-tag + GFP-tag vectors**: 10% SDS-PAGE Gel Photograph showing FEN1 protein expression efficiency and levels for the pEHISTEV and pEHISGFPTEV CroV FEN1 constructs, after FEN1 protein expression was inducted at 25°C overnight. The soluble FEN1 protein extracts were used to perform a Ni-NTA agarose bead pull-down;** The expected FEN1 protein molecular masses were 39 kDa for the pEHISTEV_CroV_037_FEN1, and 67 kDa for the pEHISGFPTEV_CroV_037_FEN1. The molecular weight of the former FEN1 protein seemed to be somewhat lower (~37 kDa) than what predicted, while the latter was not particularly apparent on this gel. The plain (i.e. not the 'Plus' version) PageRuler™ Prestained Protein Ladder (Fermentas, UK) was used as a size standard for comparison.

The exact same FEN1 protein samples, as represented in *Figure 3.10*, were additionally used to conduct a Western blot experiment, for the reason of hopefully obtaining a stronger CroV FEN1 protein signal. During this method, the blocked membrane containing the FEN1 protein samples was initially incubated with an anti-HIS mouse monoclonal antibody, after which process the proteins were targeted and detected with a secondary anti-mouse antibody. The outcome of the Western blot can be seen below (*Figure 3.11*).

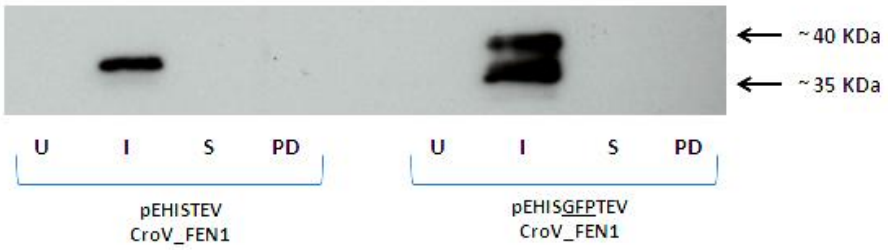


Figure 3.11: Western blot analysis of first CroV FEN1 protein expression study in the pEHISTEV 6xHis-tag and pEHISGFPTEV 6xHis-tag + GFP-tag vectors**: Western blot showing FEN1 protein expression efficiency and levels for the cloned pEHISTEV and pEHISGFPTEV CroV FEN1 constructs, after inducing protein expression at 25°C overnight. The soluble FEN1 protein extracts were used to perform a Ni-NTA agarose bead pull-down;** The samples were run on a 10% resolving gel (as seen in *Figure 3.10*). The pEHISGFPTEV_CroV_037_FEN1 protein sample appeared to be running at approximately the same molecular mass as the pEHISTEV_CroV_037_FEN1 protein sample, with two distinct bands being produced. This might have been a strong indication of the fact that the eGFP-tag (~28 kDa) had probably been cleaved of the FEN1 protein, but was somehow still being detected by the antibodies used.

Conclusions

When running the CroV FEN1 samples prepared on a SDS-PAGE gel, for both constructs generated, it was very difficult to distinguish any difference between the protein bands and hence interpret the results. Therefore, the same samples were subsequently run on a Western blot, so as to obtain a stronger protein signal. The outcome of this procedure was clearer, showing that in the case of the CroV FEN1 protein cloned into the pEHISTEV vector, containing no GFP tag, the protein was indeed being expressed at the correct size but was insoluble. On the other hand, in the case of the CroV FEN1 protein cloned into the pEHISGFPTEV vector, two protein bands were being expressed, one at the FEN1 anticipated size and the other one running a bit lower, but both were also insoluble. The fact that two bands were observed in the latter case could possibly be due to the GFP tag being cleaved off from the protein sequence, but nevertheless still being detected by the antibodies used for Western blotting. These results led to the decision of trying to optimise the FEN1 proteins' expression and solubilisation conditions.

Second CroV FEN1 protein expression study in the pEHISTEV 6xHis-tag and pEHISGFPTEV 6xHis-tag + GFP-tag vectors:

Midi-scale preparations of the Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV_037_FEN1 and Rosetta2 (DE3) (pLysS)_pEHISGFPTEV_CroV_037_FEN1 plasmids were prepared, following the same protocol as above. The difference this time being the temperature/time period of CroV FEN1 protein expression induction, as well as the lysis buffer in which the bacteria cells were eventually resuspended in.

The CroV FEN1 bacteria cultures were induced for protein expression and left to incubate at 37^{o}C for 4 hours. The plasmid bacteria cells were harvested for each culture and were resuspended in Lysis Buffer W (*Table 2.1 & 2.10*; *100mM Tris-HCl pH8, 250mM NaCl, 20mM Imidazole, 0.5% Tween, 1mM β-mercaptoethanol*). The lysis buffer chosen was a Tris-base one, instead of a Phosphate-base one. The samples collected throughout the procedure were run on a 10% SDS-PAGE gel. However, no figure is available as the resulting image was of very bad quality and none of the FEN1 proteins were particularly distinguishable.

The same FEN1 protein samples, used during the SDS-PAGE, were furthermore used to carry out a Western blot experiment, for the purpose of not only obtaining a stronger CroV FEN1 protein signal but also pin-pointing the exact molecular mass of the protein as run on the gel. The FEN1 protein samples contained within the blocked membrane were incubated with an anti-HIS mouse monoclonal antibody and the proteins were then detected with a secondary anti-mouse antibody specific to the primary one. Nevertheless, no figure is attached due to its general poor quality. Particularly, the development of the film did not generate notable chemiluminescent signals, probably further suggesting that the expected FEN1 protein cloned into the pEHISTEV and pEHISGFPTEV vectors was not actually being expressed in any of the protein samples.

Conclusions

For this experiment, two protein expression parameters were altered; the incubation temperature for protein expression and the lysis buffer for protein solubilisation. The protein samples prepared were utilised on a SDS-PAGE gel and subsequently on a Western blot. However, neither of the methods produced any results; in the case of the SDS-PAGE gel the protein bands were once again uninterpretable, while in the case of the Western blot no distinguishable chemiluminescent signal was generated. The latter results suggested that most likely the FEN1 protein, under the specific conditions chosen for protein expression, was in fact not being expressed. The optimisation of the CroV FEN1 proteins' expression and solubilisation conditions was continued.

Third CroV FEN1 protein expression study only in the pEHISTEV 6xHis-tag vector (plus pEHISTEV empty vector):

A Rosetta2 (DE3) (pLysS)_pEHISTEV_CroV_037_FEN1 plasmid bacteria culture was prepared, as usual. This time, however, the same procedure was in parallel carried out for the Rosetta2(DE3)(pLysS)_ pEHISTEV empty vector, for the purpose of using this as a control. Protein expression induction was undertaken at 37^{o}C for 4 hours, while a variety of lysis buffers were used for resuspension of the harvested bacteria cells (*See Table 2.10*). The selection of buffers tested specifically were: two PBS-based buffers (with either 0.5% Tween or 0.5% glycerol), a Phosphate-based buffer and a Tris-based buffer. The question being addressed was if completely different lysis buffers would have an effect on CroV FEN1 protein solubility. The results of this experiment can be observed in *Figure 3.12*.

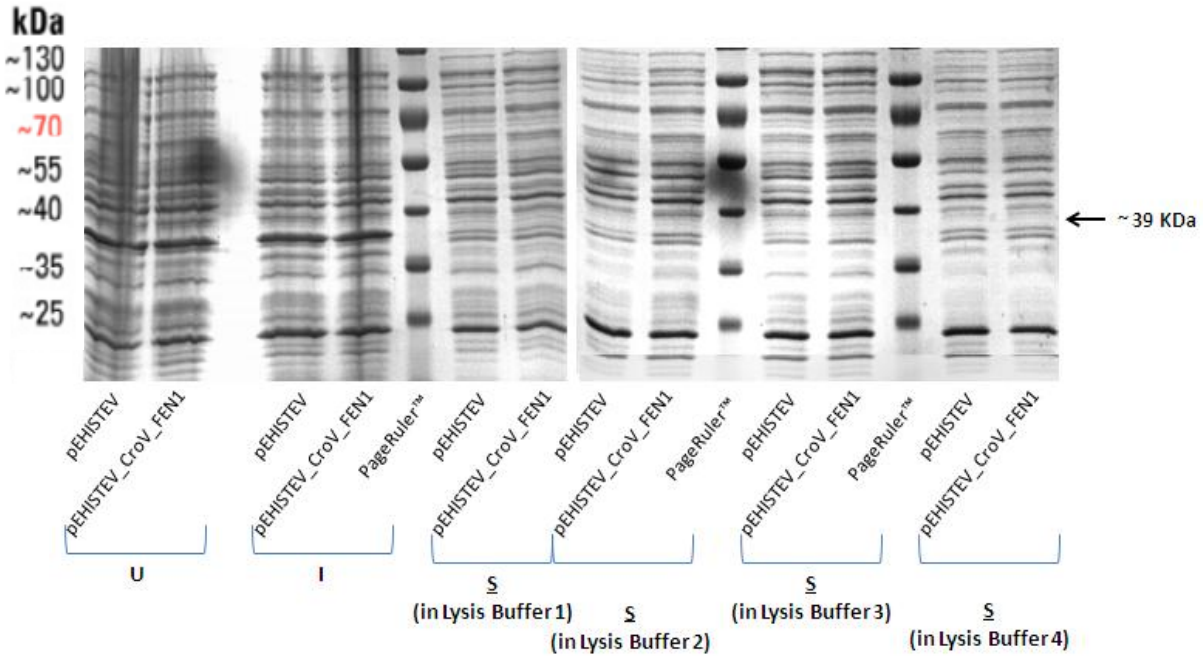


Figure 3.12: SDS-PAGE analysis of third CroV FEN1 protein expression study only in the pEHISTEV 6xHis-tag vector (plus pEHISTEV empty vector)**: 10% SDS-PAGE Gel Photograph showing FEN1 protein expression efficiency and levels for the cloned pEHISTEV CroV FEN1 construct, compared to the empty pEHISTEV vector. Both bacteria plasmid samples were induced for protein expression at 37°C for 4 hours and the bacteria cells lysed in four different buffers;** Lysis Buffer 1: PBS, 0.5% Tween; Lysis Buffer 2: 50mM NaH_2PO_4, 500mM NaCl, 30mM Imidazole, 0.5% Tween, 1mM β-mercaptoethanol, pH 8; Lysis Buffer 3: 100mM Tris-HCl pH8, 150mM NaCl, 10mM Imidazole, 0.5% Tween, 1mM β-mercaptoethanol; Lysis Buffer 4: PBS, 0.5% glycerol. The anticipated CroV FEN1 protein molecular mass was 39 kDa, but no obvious band could be detected on the gel.

Conclusions

The purpose of this experiment was to evaluate a number of different lysis buffers and their effect on the FEN1 proteins' solubility, while at the same time using the empty vector as a control sample for comparison. The results showed that the different buffers did not

have a varying effect on the proteins' expression and solubility, while most importantly when observing the gel it became apparent that almost certainly no protein was being expressed. Therefore, the next stage of the project was to express the CroV FEN1 construct in a variety of *E.coli* cells.

3.2.1.1.2 Summary of Results

While the first CroV FEN1 protein expression study revealed the presence of the protein being expressed but not soluble, all further experiments failed to produce any protein expression. This was most likely due to the fact that the bacteria growth conditions had been altered. The different lysis buffers did not appear to have an effect on the proteins' solubility either. The protein expressed in the primary experiment, however, was only present at a level detectable by blot.

3.2.1.1.3 Protein Expression and Purification experiments for the CroV FEN1, cloned into the pEHISTEV 6xHis-tag vector and expressed in a variety of *E.coli* cells

First CroV FEN1 protein expression study in a variety of *E.coli* cells:

Similarly as for the CroV PCNA, the *E.coli* cells picked were: BL21(DE3), BL21(DE3)(pLysS), Rosetta2(DE3), Arctic Express(DE3)RP and C43(DE3) (*See Section 2.2.3.1.1*). The main aim of testing other *E.coli* host strains, apart from the standard

Rosetta2(DE3)(pLysS), was to confirm whether or not they would have a different effect on FEN1 protein expression, solubility and purity.

In particular, following the exact same procedure as was carried out for the pEHISTEV_CroV (CroV_219)_PCNA plasmid (*See Section 3.1.3.2.2*); small-scale bacteria cultures of the pEHISTEV_CroV_037_FEN1 plasmid, expressed in each of the BL21(DE3), BL21(DE3)(pLysS), Rosetta2(DE3), Arctic Express(DE3)RP and C43(DE3) *E.coli* host strains, were set up using standard techniques. CroV FEN1 protein expression was induced at 37^{o}C for 4 hours. The CroV FEN1 bacteria cells were harvested and resuspended in the same PBS-based buffer, additionally containing *100mM Tris-HCl pH8, 10mM Imidazole, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol* (*See Table 2.1 and 2.10*). Nevertheless, no apparent differences between the diverse *E.coli* protein expression strains could be distinguished.

Conclusions

In conclusion, as was the case for the CroV PCNA protein, the different *E.coli* expression strains prepared and employed for the purpose of increasing the CroV FEN1 proteins' expression and solubility, did not in fact influence those parameters. Nevertheless, another attempt was made to try and optimise the conditions of the CroV FEN1 proteins' expression and solubility, while using these different *E.coli* expression strains.

Second CroV FEN1 protein expression study in a variety of *E.coli* cells:

The principle of this second study was to examine different protein expression induction temperatures for the CroV FEN1 protein, previously expressed in a range of *E.coli* host strains. The expression constructs selected for use in this experiment were the BL21(DE3)(pLysS), Arctic Express(DE3)RP and C43(DE3) pEHISTEV_CroV_FEN1 plasmids.

Once again, similarly to the pEHISTEV_CroV (CroV_219)_PCNA plasmid (*See Section 3.1.3.2.2*); the three CroV FEN1 plasmids were grown on a small-scale preparation. However, the resulting bacteria cultures after being induced for FEN1 protein expression were divided into two equal volume samples; each sample was separately grown either at 16^{o}C or at 25^{o}C overnight. Following CroV FEN1 protein expression induction at the chosen temperatures, the bacteria cells from each sample were harvested and resuspended in the same PBS-based buffer (*See Table 2.1 and 2.10*). Nonetheless, no visible differences could be determined between the *E.coli* protein expression plasmids or between the two protein expression induction temperatures (*no figure available*).

Conclusions

Similarly, the attempt to try and optimise the CroV FEN1 proteins' expression and solubilisation conditions, while making use of the different *E.coli* expression strains containing the CroV FEN1 protein, was completely uninformative, even when two different temperatures were tested for protein expression induction.

3.2.1.1.4 Summary of Results

As was the case for the CroV PCNA protein, the various *E.coli* expression strains generated to contain the CroV FEN1 protein, were not useful in that they did not allow for improved protein expression and solubilisation efficiency. Therefore, no further experiments were performed using these strains. At this point, it was decided to switch to another type of vector, one that did not contain any protein tags. The hypothesis being that maybe the presence of tags obscured the proteins' natural structure and hence affected its expression and subsequently its solubilisation. The use of these particular vectors was advantageous as they permitted the concurrent protein expression of the CroV PCNA and the CroV FEN1, a fact that in theory should prove useful when trying to isolate and purify the PCNA protein.

3.3 Un-Tagged Protein Expression and Co-Expression of Giant Virus PCNA and FEN1 in *E.coli*

3.3.1 CroV

The CroV PCNA (CroV_219; *Appendix A.3.3;* 32 kDa) and CroV FEN1 (CroV_037; *Appendix A.4;* 37 kDa) gene fragments of interest were cloned into two comparable expression vectors, none of which contained an affinity tag to assist in protein purification purposes. Namely these vectors were the pETDuet-1 and pCDFDuet-1. These vectors were selected as they facilitate co-expression (i.e. by containing two multiple cloning sites (MCS)) and co-transformation, in the same bacteria culture, of two target genes. As no

affinity tags were present within these vectors, thus no extra base pairs were added to the two gene sequences of interest and their protein molecular masses remained unaffected.

3.3.1.1 Cloning of the CroV PCNA and CroV FEN1 into the pETDuet-1 and pCDFDuet-1 vectors for Protein Expression Experiments

Refer to *Section 2.2.2.1.3; Cloning only of the CroV PCNA protein into pETDuet-1 & pCDFDuet-1 vector backbones, containing NO affinity tags for protein purification, and Section 2.2.3.1.2; Cloning of the CroV FEN1 protein into pETDuet-1 & pCDFDuet-1 vector backbones, containing NO affinity tags for protein purification*).

The constructs created were:

1. pETDuet-1_CroV(CroV_219)_PCNA (*Appendix A.5.1*)
2. pCDFDuet-1_CroV(CroV_219)_PCNA (*Appendix A.5.2*)
3. pETDuet-1_CroV(CroV_037)_FEN1 (*Appendix A.5.3*)
4. pCDFDuet-1_CroV(CroV_037)_FEN1 (*Appendix A.5.4*)

The resulting plasmids were:

1. Rosetta2 (DE3) (pLysS)_pETDuet-1_CroV (CroV_219)_PCNA
2. Rosetta2 (DE3) (pLysS)_pCDFDuet-1_CroV (CroV_219)_PCNA

3. Rosetta2 (DE3) (pLysS)_pETDuet-1_CroV(CroV_037)_FEN1

4. Rosetta2 (DE3) (pLysS)_pCDFDuet-1_CroV(CroV_037)_FEN1

Moreover, taking advantage of the fact that these two vectors that contain the target genes can actually be co-transformed in the same *E.coli* expression strain, the following plasmids were also set up:

5. Rosetta2 (DE3) (pLysS)_pETDuet-1_CroV (CroV_219)_PCNA +

   Rosetta2 (DE3) (pLysS)_pCDFDuet-1_CroV(CroV_037)_FEN1

6. Rosetta2 (DE3) (pLysS)_pCDFDuet-1_CroV (CroV_219)_PCNA +

   Rosetta2 (DE3) (pLysS)_pETDuet-1_CroV(CroV_037)_FEN1

The reason for doing the latter co-transformations was to verify any protein-protein interactions between the PCNA and FEN1 CroV proteins, which, according to the literature, are both involved in DNA replication. Specifically, they have been discovered to closely cooperate during this procedure, hence aiding in keeping it under control.

3.3.1.1.1 Protein Expression and Purification experiments for the CroV PCNA and CroV FEN1 (*each protein separately*), cloned into the pETDuet-1 and pCDFDuet-1 vectors and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

(*See Sections 2.2.2.2 and 2.2.2.3 for CroV PCNA, and Sections 2.2.3.2 and 2.2.3.3 for CroV FEN1*)

First CroV PCNA and CroV FEN1 protein expression study in the pETDuet-1 and pCDFDuet-1 vectors:

For this first study, small-scale preparations of the CroV PCNA/FEN1 plasmids were set up following standard protocols. PCNA and FEN1 protein expression was simultaneously induced in all diverse plasmids by the addition of IPTG reagent. The bacteria cultures were left to incubate at 37^{o}C for 4 hours. The different range of bacteria plasmid cells were harvested by centrifugation and were resuspended in a PBS-based lysis buffer (*Table 2.1, 2.9 & 2.10*; *PBS, 100mM Tris-HCl pH8, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol*). The total cell extract samples from each CroV plasmid were sonicated and the soluble cell extracts collected for further CroV PCNA and FEN1 protein expression analysis. All protein samples, collected throughout the process, were run on a 10% SDS-PAGE gel (*Figure 3.13*).

In addition, the same exact procedure was undertaken for the Rosetta2(DE3)(pLysS)_pETDuet-1 and Rosetta2(DE3)(pLysS)_pCDFDuet-1 empty vectors, for use as controls (*Figure 3.13*).

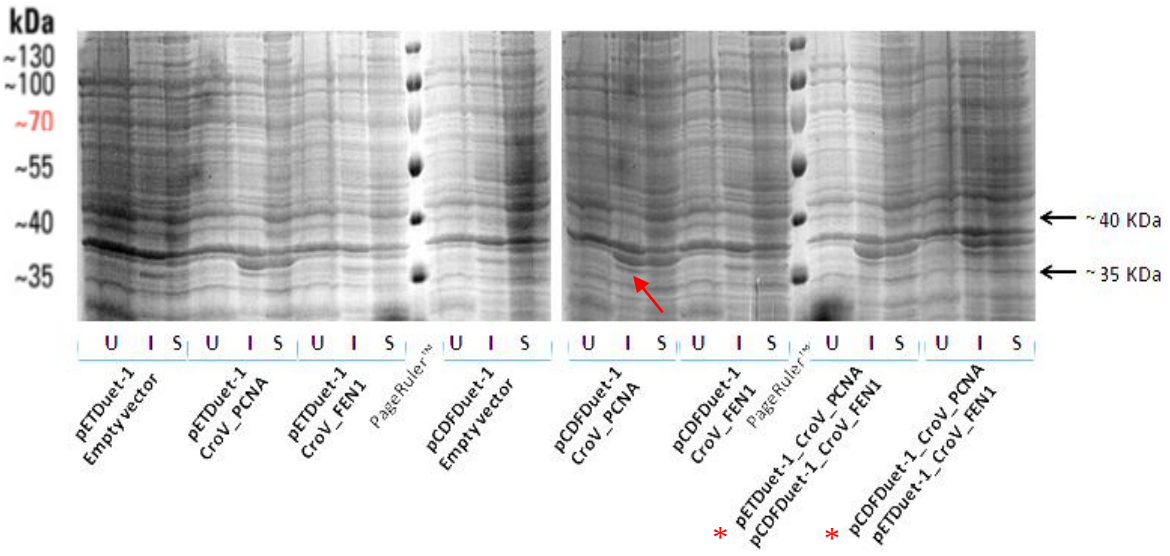


Figure 3.13: SDS-PAGE analysis of first CroV PCNA and CroV FEN1 protein expression study in the pETDuet-1 and pCDFDuet-1 vectors**: 10% SDS-PAGE Gel Photograph showing PCNA and FEN1 protein expression efficiencies and levels, for the CroV PCNA and CroV FEN1 proteins having been both cloned into the pETDuet-1 and pCDFDuet-1 vectors, compared to the empty pETDuet-1 and pCDFDuet-1 vectors. All the bacteria plasmid samples were induced for protein expression at 37^{o}C for 4 hours;** The samples indicated by the red asterisk (*) represent the PCNA and FEN1 constructs being co-transformed with each other in the same host strain. The anticipated CroV PCNA protein molecular mass was 32 kDa, and the expected CroV FEN1 protein molecular mass was 37 kDa (no affinity tags present). Even though, no obvious band could be detected for the CroV FEN1 protein at that particular molecular mass on the gel, a very strong and distinguishable band was produced for the CroV PCNA at roughly ~36 kDa (shown by a red arrow).

Note: The prominent band (*red arrow*), which was present both in the 'I' and the 'S' protein samples, was sent for Mass Spectrometry so as to determine whether or not it is indeed the CroV PCNA protein.

Conclusions

The results of this experiment, even though a bit divergent, were very positive and optimistic. In particular, strong protein bands of the anticipated sizes were evident for quite

a few of the PCNA and FEN1 protein constructs utilised during the experiment. Specifically, strong bands were obvious for all PCNA constructs and for all the constructs co-expressing the PCNA and FEN1 proteins, whereas no strong protein bands could be detected for the FEN1 constructs. Moreover, the protein bands that were being strongly expressed were also present in the soluble protein samples and were of the same intensity, revealing that probably all the protein being expressed could be solubilised. One of these strong protein bands was sent for mass spectrometry to determine its true nature. The results of this procedure were clear and most importantly showed that the PCNA protein was indeed present. Specifically, when using the ProteinPilot™ software system, the following data was retrieved for the protein band of interest sent for MS and analysed by ESI: ProteinPilot Score = 96.19, when using the BMS database, and 88.04, when using the NCBI database, ($p \leq 0.05$), Sequence Coverage (%) = 98.9, Significant Peptides (95%) = 96 (BMS database) and 101 (NCBI database). The results when analysing the same protein sample by MALDI were extremely poor. Even though these results were encouraging, the problem of producing high quantities of soluble and especially pure PCNA protein for use in additional biochemical analysis still remained, and thus the conditions of the proteins' expression and solubilisation had to be further optimised.

Second protein expression study only for CroV PCNA in the pETDuet-1 vector:

A pETDuet-1_CroV(CroV_219)_PCNA plasmid was set up, following the usual protocol. The aim of this experiment was to test the effect various lysis buffers had on the PCNA protein solubility. The variety of lysis buffers used for the resuspension and sonication of

the harvested bacteria plasmid cells can be seen in *Table 2.10*. The results of the experiment are shown in *Figure 3.14*.

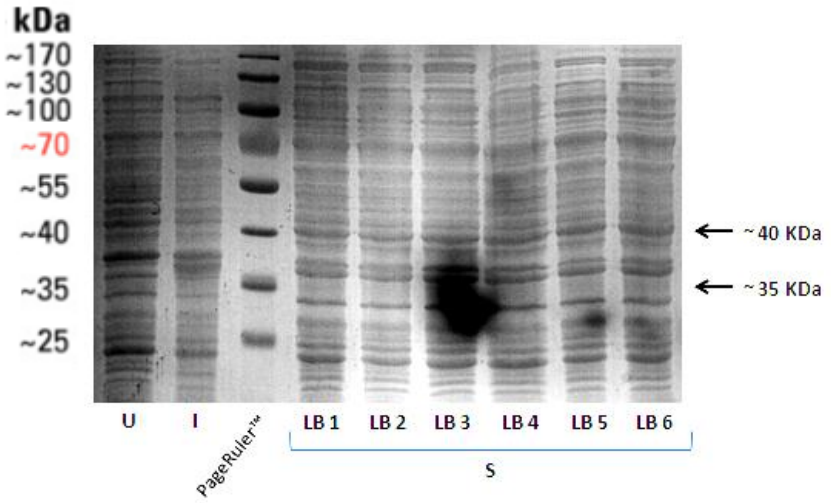


Figure 3.14: SDS-PAGE analysis of second protein expression study only of CroV PCNA in the pETDuet-1 vector**: 10% SDS-PAGE Gel Photograph showing PCNA protein expression efficiency and levels for the cloned pETDuet-1 CroV PCNA construct, after inducing protein expression at 37^{o}C for 4 hours. The bacteria plasmid cells were lysed in six different buffers;** (LB): Lysis buffer. LB 1: PBS, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol; LB 2: PBS, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol, 100mM Tris-HCl pH8; LB 3: 50mM NaH2PO4 (pH 6.5, with Na2HPO4), 150 mM NaCl, 1mM EDTA and 1mM β-mercaptoethanol; LB 4: 50mM NaH2PO4 (pH 6.5, with Na2HPO4), 150 mM NaCl, 1mM EDTA and 1mM β-mercaptoethanol, 0.5% Tween; LB 5: 100mM Tris-HCl pH8, 150mM NaCl, 1mM EDTA; LB 6: 100mM Tris-HCl pH8, 150mM NaCl, 1mM EDTA, 0.5% Tween and 1mM β-mercaptoethanol. The anticipated CroV PCNA protein molecular mass was 32 kDa. A more obvious band was noticed at approximately ~36 kDa, in both the 'I' and 'S' protein samples, but no obvious difference was visible between the various lysis buffers used for resuspension of the bacteria cells.

Conclusions

The aim of this experiment was to examine a number of different lysis buffers and evaluate their effect on the CroV PCNA proteins' solubility efficiency. All the different lysis buffers chosen had the same effect on the proteins' solubility, producing the same intensity protein bands. It appeared though that quite a significant amount of protein was still being solubilised, but this was still not adequate. The next stage of the project was to try and purify as much of the soluble CroV PCNA protein as possible with the use of the *AKTA* protein purifier.

Third protein expression and purification study only for CroV PCNA in the pETDuet-1 vector (*AKTA* purification with HiTrap™ Q HP column):

A pETDuet-1_CroV(CroV_219)_PCNA plasmid was set up, following the typical protocol. The soluble cell extract prepared from this plasmid was employed in a protein purification analysis using the *AKTA* system. The soluble cell extract had previously been resuspended in Lysis Buffer W (*Table 2.1 & 2.9*; *100mM Tris-HCl pH7.5, 150mM NaCl, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol*). This was directly loaded onto a Hi-Trap™ Q HP column, which interacts and binds negatively charged proteins. Protein elutions were assessed by performing a gradient purification of salt (NaCl) concentrations, ranging from 150mM to 1M. For the exact protocol see *Section 2.2.2.4.4; Protein Purification Experiments for CroV PCNA using an AKTA system*. The samples collected were run on a 10% SDS-PAGE gel, but only the important SDS-PAGE gel photographs are shown here (*Figure 3.15*).

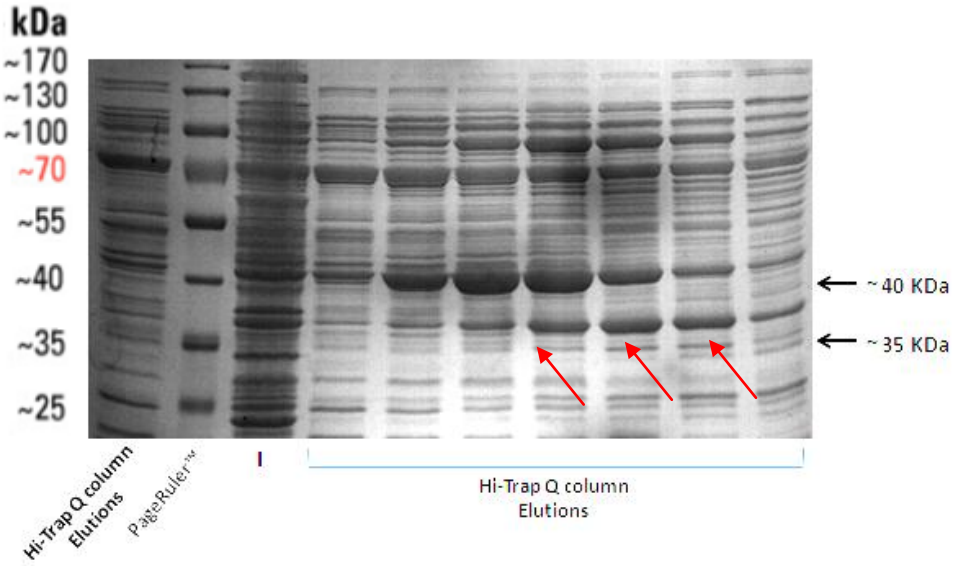


Figure 3.15: SDS-PAGE analysis of third protein expression and purification study only of CroV PCNA in the pETDuet-1 vector (*AKTA* purification with HiTrap™ Q HP column)**: 10% SDS-PAGE Gel Photograph showing PCNA protein purification efficacy for the cloned pETDuet-1 CroV PCNA construct, after inducing protein expression at 37^{o}C for 4 hours. The soluble protein extract was filtered through a Hi-Trap™ Q HP column on the *AKTA* system;** (I): induced. Lysis Buffers: 100mM Tris-HCl pH7.5, 150mM → 1M NaCl, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol. Two separate buffers were made up so as to conduct a gradient purification; one buffer contained 150mM NaCl, while the second one contained 1M NaCl. An incredibly prominent band was visible at roughly ~36 kDa in some of the eluted protein samples (illustrated by the red arrow).

Note: One of the bands represented by the red arrows was sent for Mass Spectrometry, to determine whether or not it is the CroV PCNA protein.

Conclusions

The result of this experiment was in general very positive, even though the CroV PCNA protein could not be greatly purified. More specifically, the PCNA protein was still being expressed and somewhat soluble, but when attempting to purify it did not emerge in a very pure form, i.e. it was eluted from the column together with a variety of other proteins, thus

making it difficult to isolate. In addition, this experiment verified that under certain conditions, and when employing certain elution buffers, greater amounts of PCNA protein were eluted. Additional attempts to purify the protein would have resulted in losing a significant amount of protein. One of the distinct protein bands was once again sent for mass spectrometry analysis. The results however were not as good a quality as anticipated. Specifically, when using the ProteinPilot™ software system, the following data was retrieved for the protein band of interest analysed by ESI: ProteinPilot Score = 6 ($p \leq 0.05$), Sequence Coverage (%) = 21.2, Significant Peptides (95%) = 5; the data was the same when compared against both the NCBI and BMS protein databases. These results were unexpected due to the original protein band size and intensity produced on the SDS-PAGE gel, a fact which meant that the PCNA protein could not be easily purified and isolated from other proteins. Following this experiment, all efforts were focused on trying to isolate the PCNA protein with the help of its FEN1 protein partner.

3.3.1.1.2 Summary of Results

In general, this set of experiments produced the most positive and optimistic results found during the course of this project. Initially, it was established that the new vectors, which lacked protein tags and were used to clone both the CroV PCNA and FEN1 proteins, produced higher-quality results in comparison to all previous vectors used. Therefore, the hypothesis that the presence of protein tags may in fact obscure the proteins structure may be to some extent correct. Additionally, from the first protein expression experiment it was apparent that most of the anticipated PCNA protein being expressed may also have been

soluble. The second protein expression experiment confirmed that the use of different lysis buffers for protein resuspension did not affect the proteins' solubility, as all samples produced the same protein band pattern of the same intensity. Finally, the protein purification study demonstrated that even though a good amount of the CroV PCNA protein could be eluted from the column under certain conditions, this was not pure and hence could not be easily isolated for use in additional biochemical analysis. For two of the experiments mentioned, the presence of the expected CroV PCNA protein was verified by mass spectrometry analysis, however only for the first experiment can we be 100% confident for the proteins' presence.

3.3.1.1.3 Protein Expression and Purification experiments for the CroV PCNA and FEN1 (*both proteins together*), cloned into the pETDuet-1 and pCDFDuet-1 vectors and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

Refer to *Section 2.2.1.5;PCNA/FEN1 Protein Purification experiments using an AKTA system.*

First simultaneous CroV PCNA and FEN1 protein purification study in the pETDuet-1 and pCDFDuet-1 vectors (*AKTA* purification with Heparin HP Trap™ column):

In this part of the study, one of the co-transformed CroV PCNA and FEN1 plasmid samples, that produced the extra protein band at ~36 kDa (as seen in *Figure 3.13*), was

obtained for the purpose of conducting the very first simultaneous PCNA and FEN1 protein purification experiment. In particular, the plasmid chosen was the "Rosetta2 (DE3) (pLysS)_pETDuet-1_CroV (CroV_219)_PCNA + Rosetta2 (DE3) (pLysS)_pCDFDuet-1_CroV (CroV_037)_FEN1".

Purification methodologies for this sample on the *AKTA* system, would allow the determination of any protein-protein interactions. By using particular protein trapping columns, a two-way interaction could possibly form, with the one protein binding onto the actual column while the second protein interacting and binding to the already bound first protein (if a protein-protein interaction does actually occur). Then the two proteins would detach from the column and elute together. At the same time, it could be hypothesised that by co-transfecting the two different proteins, each one might have an effect on the protein expression efficiency and levels, but also on the purification purity, of the other.

The soluble cell extract prepared from the plasmid mentioned above, expressing both CroV PCNA and FEN1 proteins, was thus employed in a protein purification analysis using the *AKTA* system. The soluble cell extract had previously been resuspended in the PBS-based lysis buffer (*Table 2.1, 2.9 & 2.10*; *PBS (usually contains 150mM NaCl), 100mM Tris-HCl pH8, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol*). This was loaded directly onto a Heparin HP Trap™ column, as this type of column is effective in binding proteins that interact with DNA. Therefore, the general hypothesis was that this column would allow the PCNA to bind to it through a FEN1 intermediate interaction (presuming that the

PCNA protein interacts with the FEN1 protein). More precisely, the FEN1 protein would presumably bind directly to the column, while the PCNA protein would also bind by forming a secondary interaction with the FEN1 protein, or vice versa. Eventually the two proteins would detach from the column and elute in the same fraction(s). Protein elutions were assessed by performing a gradient purification of salt (NaCl) concentrations, ranging from 150mM to 2M. The samples of interest were run on a 10% SDS-PAGE gel (*Figure 3.16*).

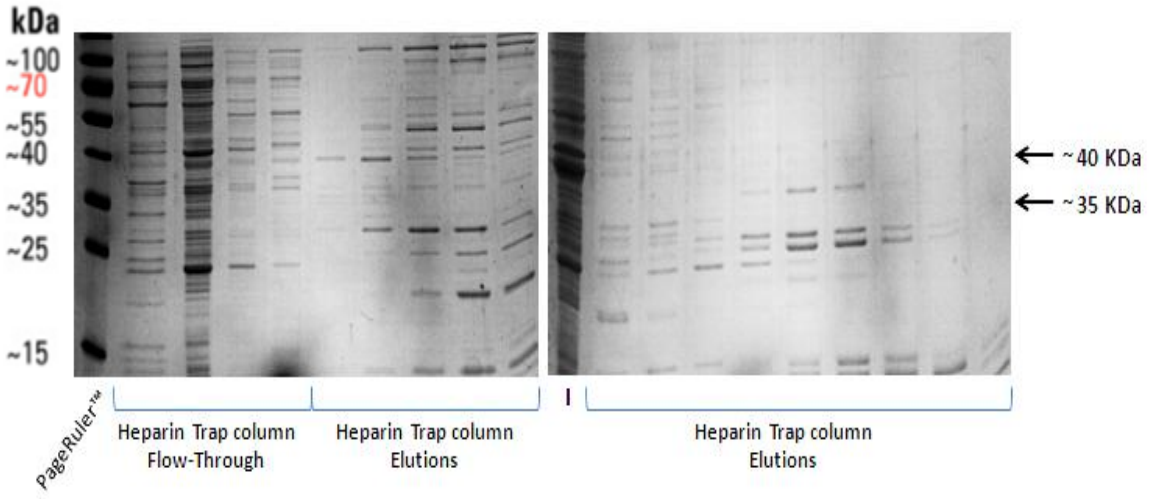


Figure 3.16: SDS-PAGE analysis of first simultaneous CroV PCNA and FEN1 protein purification study in the pETDuet-1 and pCDFDuet-1 vectors (*AKTA* purification with Heparin HP Trap™ column)**: 10% SDS-PAGE Gel Photograph showing PCNA and FEN1 protein purification efficacy for the co-transformed pETDuet-1 CroV PCNA and pCDFDuet-1 CroV FEN1 constructs. The soluble protein extract was filtered through a Heparin HP Trap™ column on the *AKTA* system for simultaneous PCNA/FEN1 protein purification;** Lysis Buffers: PBS (150mM NaCl → 2M NaCl), 100mM Tris-HCl pH8, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol. Two separate buffers were made up so as to conduct a gradient purification; one buffer already containing 150mM NaCl (amount present in all PBS-based buffers) and a second one containing a further ~2M NaCl. The soluble protein sample (*red asterisk*) utilised, as well as the (U): uninduced and (I): induced, can be seen in *Figure 3.13*.

3.3.1.1.4 Summary of Results

For the purpose of this experiment, a strain co-expressing the PCNA and FEN1 proteins was used. Protein expression was induced and the resulting sample was directly utilised in an *AKTA* purification experiment. However, the assumption that the two proteins would bond by forming an interaction between them and also with the particular column, and hence elute together under certain conditions did not take place. No noticeable bands could be detected for either the PCNA or the FEN1 CroV protein, throughout the gradient *AKTA* purification. Some further CroV PCNA protein purification experiments were attempted next.

3.4 Additional Trials of Protein Purification Experiments

3.4.1 Protein Purification experiments for the CroV PCNA, cloned into the pETDuet-1 vector and expressed in Rosetta2(DE3)(pLysS) *E.coli* cells

Refer to *Section 2.2.2.4.3; EZview™ Red Streptavidin pull-down of CroV PCNA Solubilised protein sample, with the use of EZview™ Red Streptavidin affinity agarose gel beads and the help of a Fen1 peptide (only for the pETDuet-1 PCNA construct).*

First CroV PCNA protein purification study in the pETDuet-1 vector (first EZview™ red streptavidin affinity agarose gel bead purification attempt):

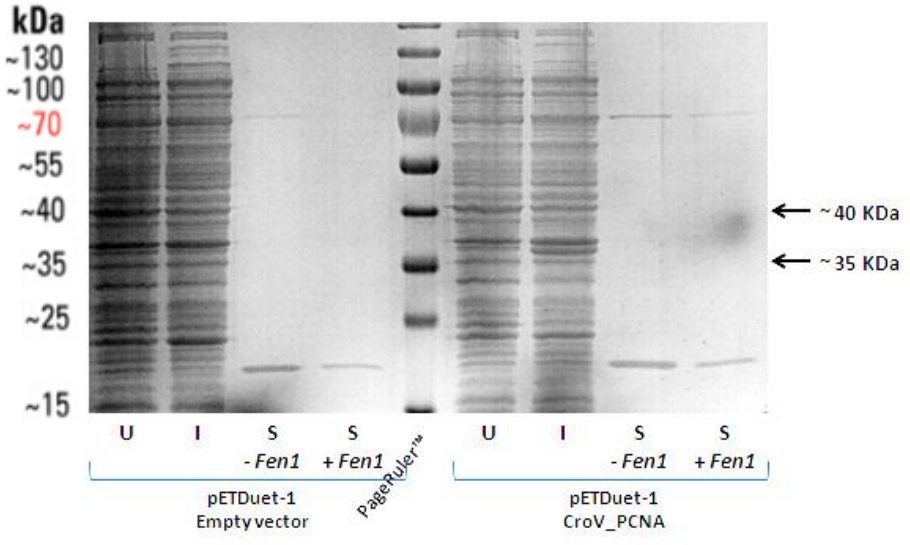
A small-scale preparation of the pETDuet-1_CroV (CroV_219)_PCNA plasmid was set up, following standard techniques. The bacteria culture was induced for PCNA protein

expression at 37^{o}C for 4 hours. The bacteria plasmid cells were harvested and were resuspended in the usual PBS-based lysis buffer (*Table 2.1, 2.9 & 2.10*; *PBS, 100mM Tris-HCl pH8, 0.5% Tween, 1mM EDTA and 1mM β-mercaptoethanol*). The total cell extract sample was sonicated and the soluble cell extract sample collected for further CroV PCNA protein purification analysis. In addition, the exact same procedure was undertaken for the pETDuet-1 empty vector, for use as a control sample.

In this part of the project, CroV PCNA protein purification was attempted by a slightly different streptavidin pull-down protocol. The completion of this technique required the use of a biotinylated *Fen1* peptide, as well as EZview™ red streptavidin affinity agarose gel beads. The *Fen1* peptide specifically included a putative PCNA-interacting peptide (PIP) box motif sequence (i.e. N--I--LL). The hypothesis was that the *Fen1* peptide could possibly bind to and interact with the PCNA protein through the specific PIP-box recognition sequence. Once an interaction between the peptide and the protein was formed, then the *Fen1*-PCNA complex could be pulled-down from the soluble *E.coli* cell extract by the EZview™ Red Streptavidin affinity agarose gel beads, which attract and attach to biotinylated targets.

For this purpose, the soluble cell extract, corresponding to either the control pETDuet-1 or the pETDuet-1_CroV (CroV_219)_PCNA plasmid, was further divided into two soluble cell extract samples of equal volumes; one sample was incubated with the *Fen1* peptide to allow potential interaction with the PCNA protein, while the other one was once again

used as a control (no *Fen1* peptide added). All samples, even the controls, were incubated with the EZview™ red streptavidin affinity agarose gel beads and washed a few times with the PBS-based buffer used for lysing the bacteria cells. The pull-down protein samples were checked for successful CroV PCNA protein purification by being run on a 10% SDS-PAGE gel (*Figure 3.17*).



.

Figure 3.17: SDS-PAGE analysis of first CroV PCNA protein purification study in the pETDuet-1 vector (first EZview™ red streptavidin affinity agarose gel bead purification attempt)**: 10% SDS-PAGE Gel Photograph showing PCNA protein purification efficacy for the cloned pETDuet-1 CroV PCNA construct, compared to the empty pETDuet-1 vector. The soluble protein extracts were used to perform a EZview™ red streptavidin affinity agarose gel bead pull-down, with the intermediate assistance of a *Fen1* peptide. Both bacteria plasmid samples were induced for protein expression at 37^{o}C for 4 hours;** An extra band was indeed noticeable in the 'I' sample of the pETDuet-1_CroV_PCNA construct but it appeared to have a molecular weight higher (~36 kDa) than what anticipated (~32 kDa). Nevertheless, the PCNA streptavidin purification with the *Fen1* peptide did not appear to have worked.

Conclusions

This experiment unfortunately did not produce significant results, apart from the fact that a protein band of the anticipated size, for the PCNA protein, was slightly being expressed but could not be purified. Therefore, the hypothesis that the PCNA protein can be purified and isolated with the help of a *Fen1* intermediate peptide may not be plausible, specifically under the conditions chosen. Nevertheless, an additional attempt was made to isolate the PCNA protein in this manner.

Second CroV PCNA protein purification study in the pETDuet-1 vector (second EZview™ red streptavidin affinity agarose gel bead purification attempt):

A pETDuet-1_CroV(CroV_219)_PCNA plasmid was set up as usual. However, the bacteria cells once harvested were resuspended in different lysis buffers, instead of the typical PBS-based buffer. The lysis buffers chosen for this procedure were all Phosphate-based, but each one was made up to a different salt NaCl concentration (ranging from no to 300mM NaCl; x3 salt concentrations) and was also measured to a different pH (ranging from pH 6.5 to pH 8.5; x3 pHs) [*See table 2.1 & 2.9; 50mM NaH2PO4 (pH 6.5/ 7.5/ 8.5, calibrated with Na2HPO4), 0/ 150/ 300 mM NaCl, 1mM EDTA and 1mM β-mercaptoethanol*]. This resulted in an overall of nine different total cell extract samples, due to the diverse combination of lysis buffers prepared. The soluble cell extracts collected were used in an EZview™ red streptavidin affinity agarose gel bead pull-down PCNA purification experiment, assisted by the intermediate biotinylated *Fen1* peptide (as described beforehand).

No SDS-PAGE gel photograph has been attached as the streptavidin pull-down of the PCNA soluble cell extract samples once more did not work. Only unspecific proteins were purified. However, the extra band running at ~36 kDa, presumed to be the PCNA protein, was still present in the 'I'nduced protein samples (same as previously seen in *Figure 3.15*). It should be noted though that the different lysis buffers applied had no distinguishable variation in their effect on the PCNA protein solubility.

Conclusions

During this experiment a variety of different lysis buffers were tested and their effect on the CroV PCNA proteins' solubility evaluated. The outcome was the same as previously, even though a protein band was being expressed at the anticipated size, no protein of interest could be purified. Moreover, the different buffers did not have a noticeable difference on the PCNA proteins' solubility.

3.4.2 Summary of Results

In this part of the project, purification of the CroV PCNA protein was attempted with the use of a *Fen1* peptide, hypothesised to interact through a PIP-box domain with the PCNA protein. This assumption, however, could not be proven as no PCNA protein could actually be purified and isolated, not even when employing a number of different lysis buffers for protein solubilisation. In conclusion, PCNA protein purification and isolation with the help

of intermediate proteins and peptides has proven to be very difficult, if not impossible, under the conditions employed during the course of this project.

-CHAPTER FOUR-

Discussion

4.1 Results of this study

The aim of this study was to perform a thorough biomolecular and biochemical investigation on the DNA replication machinery of the giant viruses Mimivirus, Marseillevirus and CroV. In particular, two key viral proteins involved in this fundamental procedure were taken into consideration; the PCNA sliding clamp and the FEN1 endonuclease. The main purpose of this project was to examine the properties and functions of these two vital DNA replicatory components. Upon completion of this study, it was anticipated that it would have generally been possible to gain a better understanding on the life-cycle of these viruses, which, during the course of history, have somehow acquired an enormous amount of genetic information granting them such a huge size that they cannot actually pass unnoticed. Therefore, to achieve the original goal set, a series of different protein expression protocols were undertaken; initially by cloning the synthetic PCNA genes encoded by the Mimivirus, Marseillevirus and CroV, and the synthetic FEN1 gene encoded only by the CroV, into a collection of suitable expression vectors and then expressing these into a range of *E.coli* host expression strains. In the final stages of the project, a few trial protein purification methods were attempted.

For the first part of this study; the synthetic PCNA genes encoded by the Mimivirus, Marseillevirus and CroV, and the synthetic FEN1 gene encoded only by the CroV, were cloned into a variety of expression vectors containing a specific tag. Cloning the gene of interest into a tagged-vector is usually extremely helpful, as it allows for easy and simple protein purification, while sometimes the presence of a tag may also increase the efficiency of the protein expression conferring larger amounts of the particular protein.

First of all, the Mimivirus, Marseillevirus and CroV PCNA synthetic genes were cloned into a pASK-IBA17plus vector backbone. This vector had been supplied by its manufacturer with a Strep-tag, which permits protein purification with the use of streptavidin beads. When a correct clone had been established for all three organisms, these were next expressed in a Rosetta2(DE3)(pLysS) *E.coli* host strain; a strain that offers enhanced protein expression levels, as stated by the producer.

The samples collected from an initial protein expression experiment were run on a simple SDS-PAGE gel (*Figure 3.1*); the results did not reveal a great difference between the protein expression patterns for the Mimivirus, Marseillevirus and CroV, while additionally a very distinct band running at ~ 35 kDa (i.e. expected PCNA protein molecular mass) could not be observed for either of the organisms. However, a slightly more noticeable band was seen for the Marseillevirus and this was estimated to be running at approximately 34 kDa. This suggested that the specific band may have possibly represented the PCNA protein; but this information was not convincingly positive. This band, nonetheless, was

only present in the protein induced sample and not in the soluble fraction, indicating that the protein if present might have been completely insoluble. The best way to increase a proteins' solubility, apart from changing the conditions under which protein expression is performed, is by modifying the lysis buffers in which it is resuspended for protein extraction. Different amounts or concentrations of reagents within a lysis buffer can have varying results on a proteins' expression and solubility. Therefore, a next step in the experimental plan could have been to try out different lysis buffers containing varying amounts of the appropriate reagents.

The fact that no expression was obtained for the Mimivirus and CroV PCNA protein, while only a small (if any) was obtained for the Marseillevirus, primarily hinted towards the fact that maybe the protein expression conditions (i.e. growth temperatures and time periods) were not optimised. It is a well known fact that different proteins require different conditions for their proper growth and competent expression; even the precise same protein found in diverse organisms will require different growth and expression conditions. Therefore, a next step during the experimental plan would have been to test other protein expression conditions, hoping to determine the right conditions that would allow for more efficient and greater PCNA protein expression, separately for the Mimivirus, Marseillevirus and CroV.

Before proceeding to test other protein expression conditions, however, a Western blot was performed for the Mimivirus, Marseillevirus and CroV PCNA protein using the same

samples previously run on the SDS-PAGE gel. The viral PCNA proteins were detected by incubating the samples of interest with Strep-Tactin™ horse radish peroxidase (HRP) conjugate, in a one-step protein detection process. The outcome of this procedure would determine whether or not PCNA protein expression was indeed present or absent; this process failed to expose the presence of any PCNA protein expression for the Mimivirus and the Marseillevirus, although for the latter it was hypothesised that PCNA expression had been produced to an extremely low level during the protein expression experimental procedure. This led to the belief that maybe, to some extent, it was quite possible that there was either something wrong with the vector used or, even if the vector was not problematic, then maybe there was a problem with the Strep-tag. It has previously been reported that the presence of a tag may have a bad effect on the protein of interest basically by obscuring its confirmation and folding. Additionally, the Strep-tag may itself have an altered confirmation thus somehow being concealed within the PCNA protein structure; this would mean that it would remain undetectable from any antibodies or beads. Nonetheless, a chumiluminescence signal was generated for the CroV PCNA protein, demonstrating a particularly high level of protein expression (molecular mass ~ 34 kDa), but unfortunately the protein was not greatly soluble (*Figure 3.7*).

Consequently, after taking into serious considerations the results obtained from the Western blot, it was decided: A) First of all, not to continue with the pASK-IBA17plus Mimivirus PCNA construct. B) Secondly, to continue with the pASK-IBA17plus Marseillevirus PCNA construct, as on the SDS-PAGE gel performed a tiny amount of what

was thought to be the PCNA protein was noticeable, and therefore the plan was to test a range of protein expression conditions so as to hopefully increase PCNA protein expression efficacy (*Figure 3.4*). However, the diverse growth conditions chosen did not appear to have any noticeable effect on the PCNA protein expression, as seen when the protein samples were run on an SDS-PAGE gel, where no difference could basically be identified between them. These samples were also run on a Western blot hoping to acquire more distinguishable results (*Figure 3.5*); this method demonstrated that the Marseillevirus PCNA protein was indeed being expressed (molecular mass of ~ 34 kDa) in all conditions selected, but some of the conditions (i.e. higher temperatures) did certainly seem to have a greater effect on its expression than others. The best experimental conditions were selected so as to perform another protein expression experiment this time attempting to increase the Marseillevirus PCNA proteins' solubility, thus miscellaneous lysis buffers were tested. The resulting protein samples were once again run on a Western blot (*Figure 3.6*). The various buffers were shown to have a diverse effect on the PCNA expression; some being responsible for improved and generally a greater level of protein expression (i.e. buffers containing double the amount of salt than usual, as well as considerably more detergent), while others were rather insufficient. C) Finally, it was decided to continue experiments with the pASK-IBA17plus CroV PCNA construct, as PCNA protein expression was confirmed but the proteins' solubility had to be largely improved. For this purpose, a wide variety of growth conditions and lysis buffers were checked. The resulting protein samples were run on a Western blot, where CroV PCNA expression appeared to be prominent with an obvious band running at ~34 kDa (*no data shown*). In general terms, it can be stated that

lower temperatures gave more satisfactory results, while the various lysis buffers did not seem to greatly affect the protein expression levels.

After extensively trying to gain good quality PCNA expression from the pASK-IBA17plus Mimivirus, Marseillevirus and CroV constructs, eventually the Mimivirus, Marseillevirus and CroV PCNA synthetic genes were cloned into another vector containing a different tag. By performing this procedure it was anticipated that more efficient PCNA protein expression would be achieved, mainly due to the different tag that would be present. The vector chosen specifically was the pEHISTEV, which contains a 6xHis-tag. The 6xHis-tag allows for protein purification with the use of Ni-NTA agarose beads. The approved Mimivirus, Marseillevirus and CroV clones were expressed once again in the Rosetta2(DE3)(pLysS) host strain.

The samples collected from this experimental protocol were run on a SDS-PAGE gel for PCNA protein expression confirmation (*Figures 3.2 & 3.3*). This time both the soluble and insoluble protein samples were examined, revealing that the Marseillevirus and CroV PCNA proteins, even though expressed at around ~ 35 kDa, were not exceptionally soluble. Nevertheless, a one-step Ni-NTA bead pull-down was performed for all the soluble Mimivirus, Marseillevirus and CroV PCNA fractions displaying that only for the CroV quite a large amount of what was considered to be the PCNA protein could be isolated but not in an incredibly pure form. This meant that, even though some CroV PCNA protein was found in the insoluble samples, still quite a significant amount of the

protein seemed to be purifiable. To increase the proteins purity a solution would have been to carry out multiple steps of the Ni-NTA pull-down, but unfortunately this solution would also have a negative effect, considering that after every purification step a significant amount of the desirable protein would have been lost. Concerning the Marseillevirus PCNA protein, no protein was actually pulled-down with the use of the Ni-NTA beads, meaning that the protein was probably all in the insoluble fraction.

After these results, two decisions were reached; a) As no results whatsoever could be produced for the Mimivirus PCNA protein, and the low level of PCNA protein that was expressed for the Marseillevirus was apparently insoluble in every condition tested, thus research on the PCNA protein in these two virus would be aborted; and b) The PCNA protein expression trials for the CroV would be continued, because for this virus PCNA expression was the most efficient of all.

As a consequence, following this, an even greater assortment of growth conditions and lysis buffers were investigated and analysed for the CroV PCNA (*Figures 3.9*). The general results were similar to those established for the pASK-IBA17plus CroV PCNA construct; even though some protein expression conditions had a somewhat notable effect on the PCNA protein expression, others had an insignificant effect on it. Moreover, not a huge difference could be detected between the data obtained from all the varying conditions, while in all cases the levels of protein expression were still considerably low. So it had become fairly obvious that no matter what protein expression conditions were

chosen, the PCNA expression was still not optimal. By this point, the solutions and options for optimising the CroV PCNA protein expression were becoming quite scarce; different vectors containing diverse tags were tested, a large variety of protein expression conditions had been examined and an even greater range of lysis buffers had been checked. The question needing to be answered was 'what was the real problem causing the PCNA protein to be expressing so poorly?'.

Some PCNA protein purification trials were also performed, mainly to check two things: a) if the PCNA could actually be purified (i.e. if the 6xHis-tag was not hidden within the PCNA protein structure and hence undetectable by the Ni-NTA agarose beads), and b) if the PCNA could indeed be purified, how pure was it and how much of it was actually being generated that could eventually be used in further biomolecular and biochemical analysis.

For the first protein purification trial a soluble fraction of the CroV PCNA was manually filtered through a column packed with Ni-NTA beads; different buffers were utilised for lysing and eluting the protein sample. The result of this trial was run on a SDS-PAGE gel (*no data shown*), which illustrated that although the PCNA protein was being expressed it was directly coming off the column in the 'flow-through' sample. A very similar outcome was also verified when another soluble CroV PCNA protein sample was automatically filtered through a His-trap column on an *AKTA* system (*no data shown*). This principally meant that the protein was not binding to the Ni-NTA beads, further suggesting that the

6xHis-tag was either problematic or concealed within the PCNA protein structure. In addition, it could also imply that the tag had been cleaved off the actual protein structure. The 6xHis-tag is quite small so it could not easily be identified if cleaved off the protein structure.

A second purification trial, however, did produce an evident band at around ~ 35 kDa that was believed to be the PCNA protein (*Figure 3.8*). In addition, this procedure seemed to accomplish a rather high degree of protein purity. This band was sent for mass spectrometry to confirm its nature. As a result, the presence of the CroV PCNA was confirmed, but only at a low level hence not permitting further biochemical analysis. As such, even though this purification experiment was considered successful in that it finally allowed us to identify the presence of the CroV PCNA, the fact, however, that the actual protein could only be purified and isolated in very low levels meant that the procedure required again further optimisation steps. Therefore, as time was running short, it was decided against trying to optimise the conditions of this protocol, but rather move on to new protein expression optimisation experiments hoping to get better quality results much faster, i.e. higher levels of purified CroV PCNA protein enough to perform a number of further biochemical analysis.

The bacteria colonies and cultures grown and used throughout this project were in all circumstances considered fresh, as they were not more than five days to a week old. However, in regards to the experiment just discussed, fresh bacteria transformations into

E.coli cells for both the CroV PCNA construct and the empty vector, used as a negative control, were performed on the day the specific experiment was initiated. Thus, the fresh bacteria cultures that were grown, which were only a day old, could have had a more efficient and enhanced protein expression level, justifying the positive results of this experiment. In addition, all reagents and solutions prepared and applied throughout this experiment were also fresh, having been prepared on the day. This could also be a possible explanation as to why this experiment worked more optimally compared to previous ones.

Following these protein purification trials, it was thought that maybe expressing the PCNA protein in various other *E.coli* host expression strains would possibly help to improve the effectiveness of the CroV PCNAs' expression levels. Therefore, once the CroV PCNA was expressed in a careful choice of *E.coli* host expression strains (each one of these strains being able to enhance protein expression efficiencies in a different way to the other), the plasmids generated were used in a number of protein expression trials. Once more, different protein expression conditions were studied.

The results acquired from the first such protein expression experiment demonstrated that the previous assumption made was in fact wrong; specifically, expressing the CroV PCNA in the different *E.coli* host strains had no effect whatsoever on the PCNA protein expression levels or efficiency. At the same time, it was fairly difficult to detect any kind of expression for the PCNA protein, as the band pattern of expressed proteins generated was similar for all plasmids (*no data shown*). However, when the same CroV PCNA

plasmids were grown at much lower temperatures, a somewhat greater effect on their PCNA protein expression levels could be detected, but still this was not considered to be of major significance.

At this point in the project, a second key protein involved in DNA replication was introduced into the study. This was FEN1 and it was only studied in CroV. The CroV FEN1 synthetic gene was initially cloned into the pEHISTEV vector, while the equivalent pEHISGFPTEV vector was also selected so as to check its efficacy and compare it to the previous one. The pEHISGFPTEV vector, apart from containing the 6xHis-tag, it additionally contains a GFP-tag. It was assumed that by increasing the overall molecular mass of the protein under investigation, maybe it would either prevent it from misfolding (if that was a cause for the poor protein expressions achieved so far), or due to the presence of two tags, one of which was quite considerable in size, maybe it would have been impossible for them to become concealed within the protein structure hence allowing them to be detectable by the binding substrate. The correctly checked CroV FEN1 clones were expressed in the Rosetta2(DE3)(pLysS) host expression strain.

The FEN1 protein samples collected from some of the earlier protein expression trials were run on a SDS-PAGE gel (*Figure 3.10*). As usual, exceptionally distinguishable bands were not observed either for the pEHISTEV CroV FEN1 construct or for the pEHISGFPTEV CroV FEN1 construct. Nevertheless, a slightly more visible band was seen for the pEHISTEV CroV FEN1 construct running approximately at the expected molecular mass

(~ 37 kDa), but it was not certain whether or not it did represent the FEN1 protein. The band was only noticeably present in the expression induced protein sample, while it was not obviously present either in the soluble or in the insoluble fractions. To obtain a clearer picture of these results the exact same samples were run on a Western blot (*Figure 3.11*), during which the target proteins were detected by incubating the blocked membrane with an initial anti-HIS mouse monoclonal antibody further targeted by a secondary anti-mouse antibody. In the case of the pEHISTEV CroV FEN1 construct the outcome was as expected; some FEN1 protein was being expressed but it could not be identified in either the soluble or the insoluble sample, which further suggested that it was somehow lost from the total protein sample or it may have degraded drastically over time. On the other hand, in the case of the pEHISGFPTEV CroV FEN1 construct the results were rather unexpected; it appeared as if the HIS-GFP-tag (~28 kDa) had been completely cleaved off the FEN1 protein structure, as two different protein fragments of an approximate ~ 35-40 kDa molecular mass were running one right below the other, and were only seen in the original protein sample were expression had been induced. Maybe the presence of protease enzymes could have been the reason for this outcome, as during all protein expression procedures no protease inhibitors were dialysed in the lysis buffers used.

Following these results, it was decided that: a) the research on the CroV FEN1 protein cloned into the pEHISGFPTEV vector would be terminated, as it became obvious that no correct results could be generated, while b) the FEN1 protein expression trials for the CroV

cloned into the pEHISTEV vector would be continued, this time analysing various protein expression conditions.

For the first such trial, a number of different lysis buffers were examined. However, when the protein samples were run on a SDS-PAGE gel (*Figure 3.12*), no noticeable difference could be seen in the protein expression patterns created between the control and the plasmid of interest, around the region on the gel corresponding to 37 kDa. Therefore, so as not to waste any time, it was concluded that the FEN1 protein would be expressed in a range of *E.coli* host strains (similarly to the CroV PCNA protein), hoping once again to elevate the CroV FEN1 protein expression efficiency and levels. When the FEN1 protein expressions were completed and the various plasmids created, additional protein expression trials were commenced. These resulted in the knowledge that expressing the CroV FEN1 in the different *E.coli* host strains has no effect whatsoever on the FEN1 protein expression efficiency or levels, while at the same time no FEN1 protein expression could be detected altogether (likewise to the CroV PCNA protein expression process). On growing similar CroV FEN1 plasmid samples but at much lower temperatures than the ones usually chosen, a somewhat greater effect on the FEN1 protein expression levels could be seen (*no data shown*). This, however, was not considered to be of major importance.

For the second part of this study; both the synthetic PCNA and FEN1 genes encoded by the CroV were cloned into two specific expression vectors. These vectors differed from the

previous ones used in that they did not contain any sort of tag. This protocol was carried out so as to validate whether or not the presence of the tag in all the formerly described experiments was the reason for the extremely poor PCNA and FEN1 protein expression levels and efficiency. Furthermore, these vectors are a good choice for studying protein-protein interactions, for two basic reasons: a) Firstly, they permit cloning of two different proteins in the same vector backbone, and b) Secondly, they allow co-expression of two different proteins in the same host expression strain, even if the two proteins have not been cloned in exactly the same but equivalent vectors.

In particular, the vectors mentioned were the pETDuet-1 and the pCDFDuet-1. Once the PCNA and FEN1 synthetic genes were cloned into these vectors, they were then expressed once more into the Rosetta2(DE3)(pLysS) *E.coli* host expression strain. This was the very first experimental procedure that unexpectedly provided the most distinctive and best quality results, proving the hypothesis already stated.

The protein samples collected during this method were all run on a SDS-PAGE gel (*Figure 3.13*). Surprisingly, exceptionally clear protein bands were visible for both the CroV PCNA and FEN1 proteins. These bands could easily be detected in both the total protein samples, as well as the soluble samples. Thus, it could positively be stated that the CroV PCNA and FEN1 proteins were being favourably expressed and in fact they were also highly soluble. Mass spectrometry also confirmed the presence of the CroV PCNA protein (FEN1 was not sent for mass spectrometry), in an incredibly high level. This information,

however, confirmed that the previous presence of an affinity tag somehow modified the proteins of interest resulting in the production of reduced expression levels and poor expression efficiencies.

Subsequent to these results and as there was little experimental time available, it was considered best to focus all the experimental efforts only on the CroV PCNA protein, expressed in the 'no-tag' vectors, as it was also the original protein of interest. The following protein expression trial aimed to examine whether or not the PCNA protein expression level and efficiency could be improved when lysing the cells in a mixture of buffers. The resulting CroV PCNA protein samples were run on a SDS-PAGE gel (*Figure 3.14*), where the PCNA and FEN1 protein bands of interest were still present. Nevertheless, the outcome depicted, not only that there was no difference in protein expression between the diverse lysis buffers utilised, but on the contrary it was determined that the lysis buffers used, in reality, slightly decreased the CroV PCNA protein expression, for yet unknown reasons. So for follow-up procedures the original buffer was used.

In the next part of the project, a CroV PCNA protein purification trial was undertaken, for similar reasons as stated above. The soluble fraction of the CroV PCNA was automatically filtered through a HiTrap Q (anion) column on the *AKTA* system; different buffers were utilised for lysing and eluting the protein samples. The protocol specifically carried out was a gradient of salt concentrations trying to determine exactly at what salt concentration

the CroV PCNA protein elutes from the column. The choice of the column itself was due to the fact that, first of all, there was no tag present so a column with an affinity towards a specific binding substrate would have been utterly useless, and secondly, the PCNA protein is by nature negatively charged so it should supposedly bind to such an anion column. The results of this purification trial were run on a SDS-PAGE gel (*Figure 3.15*); they illustrated that an apparently significant amount of CroV PCNA protein was expressed and this was clearly eluting from the column at approximately ~ 350mM of salt. Therefore, the HiTrap Q column seemed to have a high affinity for the protein. However, it was also thought that maybe much smaller amounts corresponding to the PCNA protein were in fact eluting throughout the elution protocol hence causing what could be considered as a significant loss of protein. The presence of the CroV PCNA protein was validated by mass spectrometry.

Subsequently, a relatively unusual CroV PCNA protein purification procedure was performed (*Section 3.4.1*); a soluble CroV PCNA protein fraction was attempted to be purified by another streptavidin pull-down protocol. For this particular purification procedure the most important component utilised was an intermediate biotinylated *Fen1* peptide containing a specific PIP-box motif (i.e. N--I--LL). This *Fen1* peptide sequence was particularly designed to match the last 20 aa of the CroV FEN1 protein, hence also containing the corresponding PIP-box domain. It was presumed that the presence of this oligopeptide would assist interaction and hence form a complex with the PCNA protein; this would supposedly be achieved by the PCNA protein through specific recognition of

and binding to the PIP-box motif of the *Fen1* peptide. If such an interaction between the *Fen1* peptide and the PCNA protein was accomplished, then hypothetically the PCNA protein could be pulled-down and purified from the soluble cell extract by specific streptavidin gel beads (i.e. EZview™ Red Streptavidin affinity gel beads), which are very competent in recognising and binding to biotinylated proteins (in this case the *Fen1* oligopeptide, which will expectantly be carrying along with it the PCNA protein). The resulting samples from this *Fen1*/PCNA co-purification trial were run on an SDS-PAGE gel (*Figure 3.17*). The outcome, however, was not as positive as expected; although a rather distinguishable band presumably representing the CroV PCNA protein was observed (this band could not be noticed in the equivalent control sample, thus the assumption that it must be the CroV PCNA protein), this was only present in fairly low amounts in the total protein extract (no protein of the anticipated molecular mass was visible in the soluble protein extract). Generally, very few protein bands were illustrated on the SDS-PAGE gel for the soluble protein extract, suggesting that the streptavidin beads characteristically had an unusually low affinity for biotinylated proteins. This overall outcome could have resulted because of the following reasons: a) The particular *Fen1* peptide did not have the capacity of interacting and forming a complex with the CroV PCNA protein, thus in this case, the PCNA protein would not have been pulled-down or purified. This fact could not be verified owing to the small molecular mass of the *Fen1* peptide; although it would have supposedly been pulled-down and purified by the technique performed, it would have remained undetected on the SDS-PAGE gel. b) Even if an extremely strong interaction between the two factors had indeed formed maybe it lacked the ability and affinity to bind tightly to the streptavidin beads. This would have essentially indicated that the biotin-tag of

the *Fen1* peptide, or even of the streptavidin beads, was problematic and did not have such a great affinity towards biotin; in this case, the *Fen1*/PCNA complex interaction if formed would have probably been present in the flow-through protein extract (not run on the gel) and hence it would have been lost. c) Finally, perhaps the whole *Fen1*/PCNA complex could not be properly solubilised in the lysis buffer used.

Taking into further consideration the third reason mentioned, the above procedure was repeated. The purpose this time was to test a variety of diverse lysis buffers, so as to evaluate if different buffers would have a varying effect on the formation of a *Fen1*/PCNA peptide-protein interaction complex; also to check whether they would enhance its affinity towards and allow its accurate binding to the streptavidin beads, as well as whether the various buffers could improve the general proteins' purification and solubility. Unfortunately, the results obtained were identical to the previous experiment; even though there was an extra band visible in the total protein extract, the equivalent band could not be observed in the soluble protein extract. Moreover, during this second CroV PCNA protein purification attempt numerous unspecific proteins seemed to have been pulled-down by the streptavidin beads, suggesting that for some reason (probably caused by the different lysis buffers used) the streptavidin beads had gained a greater, or most likely unspecific, affinity for biotinylated proteins. Whatever the actually cause, it was evident that the protocol required further optimisation.

Towards the end of the project, as there was some time left, a PCNA and FEN1 co-expression experiment was attempted taking advantage of the untagged '*Duet-1*' vectors' properties. The soluble fraction containing both the expressed PCNA and FEN1 CroV proteins was automatically filtered through a Heparin-Trap column on the *AKTA* system; once again different buffers were utilised for lysing and eluting the protein samples. The protocol created was a gradient of salt concentrations, testing at what salt concentration the PCNA/FEN1 protein-protein complex would elute at. The choice of the column itself was theoretically due to the fact that heparin is supposed to effectively bind any protein that has the capacity of interacting with DNA; therefore, it was assumed that the FEN1 protein, due to its direct action on the DNA template during replication, would consequently bind directly to the column. At the same time though, FEN1 can also form a strong interaction with PCNA itself, leading to a final two-way binding interaction (i.e. PCNA binds to FEN1 that binds to the column). So the overall hypothesis when planning this experiment was, that if the two proteins interact with each other then they could be purified and isolated by filtering them through such a column; eventually they would elute in the same fraction. This type of interaction is of major importance during DNA replication, as through it the PCNA can considerably enhance the activity of the FEN1. Therefore, technically purifying the two proteins together would further allow additional studies on the properties and nature of the original PCNA/FEN1 protein-protein interaction. The results of this PCNA/FEN1 co-expression purification trial were run on a SDS-PAGE gel (*Figure 3.16*); unfortunately, no PCNA or FEN1 protein expression could be visualised, probably suggesting one of three things: either that both proteins are gradually eluting at very low levels throughout the elution protocol, or that the PCNA/FEN1 complex is so tightly bound

to the heparin column that it does not elute with the specific buffer employed or that no PCNA/FEN1 interaction is taking place whatsoever. The second hypothesis could be re-tested.

To summarise, this study led to the following main conclusions: A) In the case of all Mimivirus, Marseillevirus and CroV PCNA proteins, as well as the CroV FEN1 protein, the presence of a tag in the selected expression vectors had an exceptionally negative effect on the proteins' expression levels and efficiencies, no matter how many different protein expression conditions were tested; thus protein purification techniques in this case were not possible and the biochemical analysis of the proteins could not be continued. B) On the contrary, the absence of any tag from other expression vectors used to clone both the CroV PCNA and FEN1 proteins had a surprisingly positive effect on the proteins' expression levels and efficiencies, hence allowing a certain amount of protein purification trials. It is assumed that the latter point should have also been the case for the Mimivirus and Marseillevirus PCNA proteins; it would have been of value proving this if more time had been available. As a final point, the experimental techniques performed during this study failed to provide evidence for a PCNA/FEN1 protein-protein interaction. However, this detail should not be considered of major significance, as it was possibly a result of the actual protocols not being fully optimised.

4.2 Technical and Experimental problems

The technical and experimental problems faced during this project have analytically been discussed throughout *Section 4.1*. In particular, these were: A) No protein being expressed; this problem was dealt with by altering the protein growth and expression conditions. B) Protein solubility problems; determined efforts to solve this problem were made by changing the protein expression conditions, as well as the lysis buffers used for resuspension of protein cells. C) Problems with the expression vectors of choice, and especially with the tags they had attached to them; the solution was to exchange them for vectors not contain any kind of tag. When this technical problem was encountered it was assumed that perhaps the tag was either being cleaved off from the protein structure, e.g. by proteases, or that it was being concealed within it, e.g. due to altered protein confirmation. In the former case, the tag would have been expressed normally, but because of its minute molecular mass it would have remained undetectable on the gel, while in accordance to the results obtained, the protein would not have been expressed at all. In the latter case, an atypical protein structure confirmation would have been caused by the tag folding into the PCNA structure, and therefore the protein would not have been able to express itself correctly. Nevertheless, the second case scenario appears more plausible; going by the protein expression experiments performed with the use of the un-tagged protein constructs, if the tag had supposedly been cleaved off from the protein structure then that should have resulted in a high degree of protein expression instead of no or little expression whatsoever. D) No protein being purified, when using the tagged protein constructs. The main reason behind this fact was thought to be that the protein of interest was actually not binding to the column, and as such, was directly flowing through. This

could be explained if the tag was cleaved off or concealed. So protein purification was quite impossible. E) Low levels of protein purity during the purification trials, when using the untagged protein constructs. This was a result of not being able to use purification columns that had a high specificity and affinity towards one particular substrate, meaning that more general columns had to be used due to the absence of a tagged-protein and these may have attracted a number of proteins with similar properties. F) No PCNA/FEN1 protein-protein interaction could be detected. This outcome could be due to a number of reasons: maybe the conditions chosen during the experimental procedure were not optimal for such an interaction to be accomplished; if the complex was formed then maybe it was not binding correctly to the purification column and thus, was directly filtering through; if the complex was binding optimally to the column then maybe it was not eluting.

During protein expression experiments there are generally many different obstacles causing difficulties which have to be solved. Proteins can unexpectedly be lost during the experimental studies performed or they can even degrade fairly dramatically over time, hence rendering them unusable. Another significant setback is whether the protein being produced is indeed stable or not; this could have as a consequence both structure confirmation and degradation problems. In addition, proteins may express rather poorly for reasons already described, while they may also do so due to host strain toxicity (maybe the protein is inconsistent with the expression strain of choice) or even due to the formation of secondary structures that influence and negatively affect the proteins interaction with other important components. Most importantly, uncontrolled protein expression can actually

affect the host cells' growth and therefore result in a decreased protein yield. At the same time, it has been reported that forceful protein induction can result in the formation of inclusion bodies. Finally, during this experimental study it was even considered that maybe the synthetic genes purchased were not as optimal as contemplated or maybe they were not the most satisfactory choice for these types of procedures. However, they were the easiest option as they are simple to use and manipulate, and most importantly lack any unsafe properties that the living viruses may have (i.e. the Mimivirus may be a causative agent of pneumonia).

4.3 Future work

In the near future, it would be extremely productive if some of the experimental plans discussed in this project were repeated just to confirm whether similar results would be achieved.

Concerning protein expression; a huge variety of other experimental processes could have been investigated and evaluated during the course of the project if there had been more time available. These would have been based on the procedures already performed, but always altering one or more of the parameters entailed. Some of these could have specifically been: A) Cloning the synthetic genes in such a way that the tag would be attached to their C-terminus, instead of their N-terminus as was employed during this study. This technique would validate whether or not attaching the tag on the opposite end

of the protein could solve any structural confirmation problems that may have been faced during this project. B) Cloning the synthetic genes in a selection of other tagged expression vectors, so as to examine if different tags would have the same effects as the ones utilised in this case. Moreover, it would have also been valuable if the Mimivirus and Marseillevirus PCNA proteins had been cloned into the pEHISGFPTEV, pETDuet-1 and pCDFDuet-1 vectors, as well as if the CroV PCNA protein had been cloned into the pEHISGFPTEV, for comparison reasons. D) The immense diversity of constructs generated could have been expressed in further host strains. Even though quite a significant number of *E.coli* host expression strains were assessed during this study and were determined not to have any sort of effect on the protein expression efficiencies, maybe other strains exist that would be more functional. Another proposal would have been to express the proteins of interest in other organisms, apart from bacteria, such as yeast or even native viruses. The results may have been of outstanding interesting. E) A final idea would have been to check a wide range of protein expression conditions, as well as lysis buffers. Generally, it is of utmost importance to optimise to the highest degree possible all the conditions employed during an experiment.

Concerning protein purification; towards the end of the project, and after having accomplished good-quality protein expression of the CroV PCNA protein cloned into the *Duet-1* vectors, a fairly successful and hopeful protein purification trial was also completed. This purification, which was carried out with the use of Hi-Trap Q (anion) column, however did not permit a high protein purity yield due to the nature of the column

used. Therefore, as a follow up experiment, it would be a key idea to use the samples obtained from the preceding experiment and perform a Hydrophobic Interaction Chromatography (HIC) purification technique, specifically using two types of columns the phenyl sepharose and the butyl sepharose. Both purification columns allow for the highest protein purity yield. HIC takes advantage of the presence of hydrophobic areas on the surface of proteins, which when filtered through such a HiTrap column will be attracted and bound to the hydrophobic areas of the solid support. Nonetheless, HIC is unique in that proteins bind to it at high salt concentrations and elute at low salt concentrations, in contrast to other columns (e.g. the HiTrap Q) for which the opposite method has to be applied. As a result, the protein of interest could be filtered through various HiTrap HIC columns following a reverse salt gradient protocol and this could be carried out multiple times until improved protein purity has been accomplished. At the same time, a simple gel filtration procedure could be undertaken. Gel filtration relies on the size and molecular weight of the proteins of interest and separates them on that basis. In a similar manner, concerning the FEN1 protein, equivalent experiments could be executed. As a general rule, in order to increase a proteins' purity it is always worth re-filtering the protein a number of times through a particular column. However, it is noteworthy that this routine would as a result cause a significant loss of protein levels.

Subsequent to successfully completing PCNA and/or FEN1 protein expression and purification, it would then be possible to continue and perform other biomolecular and

biochemical analysis on these proteins, so as to determine their molecular make-ups and evaluate their exact action during DNA replication.

In the broader context, further research is necessary in order to elucidate the functions and properties of the PCNA and FEN1 proteins in the Mimivirus, Marseillevirus and CroV, as well as other giant viruses. So far the majority of studies completed are based on computation analysis, mainly including bioinformatics and phylogenetic studies. Research focusing on the nature of these two proteins through a molecular and chemical perspective is very limited, if any at all. More virologists need to participate and get involved, directing and maybe focusing their attention towards *in vitro,* or even *in vivo*, experiments involving these key DNA replicatory proteins. In this way, it will become more likely to establish the actual role they play within these viruses, while at the same time it will be possible to evaluate precisely how they interact with each other in order to complete their action. By gaining more knowledge in the way these proteins function during DNA replication, will probably further allow the determination of the nature of giant virus' DNA replication; how they replicated inside their host and thus how do they propagate their genetic material through generations causing such a huge genomic expansion. As such, it will also be potential to determine whether or not these proteins have been evolutionary conserved in these organisms and hence specify their origin in terms of evolutionary history. This fact may further shed some light on information regarding the origins of the giant viruses themselves, and therefore, maybe it will become more evident whether or not they could be

considered as living organisms, based on current definition found in the literature, and included in a separate fourth domain of life.

4.4 Conclusions

In conclusion; in this study, even though initially it was in no way feasible to express the two proteins of interest, PCNA and FEN1, in the Mimivirus, Marseillevirus and CroV, eventually protein expression of these two proteins was accomplished only for the CroV. The detail that finally allowed protein expression to be achieved was the absence of a tag in the expression vectors used. This in itself was a very valuable and significant outcome, noteworthy for future references. Following this achievement, it was possible to attempt some protein purification trials but only for the CroV PCNA protein. However, with little experimental time remaining, the protein purification trials could not be fully finalised or properly evaluated; in general, it was possible to demonstrate a low level of PCNA protein purification, with the protein yield not being especially high and many unspecific proteins being pulled-down in the same fractions as the protein of interest. To improve these results it would have been essential to perform multiple other protein purification trials, during which the yield of the CroV PCNA protein would have had to be greatly increased and its purity significantly enhanced. As a consequence, by the end of the study it was not possible to obtain a high yield of pure CroV PCNA protein for use in further biomolecular and biochemical analysis, hence not contributing to our further understanding of the role of this specific gene in DNA replication of the CroV giant virus. There is still a great deal of

knowledge to be gained concerning these giant viruses, as well as their cellular and genetic components, and therefore further research is compulsory.

BIBLIOGRAPHY

Alberts BM (1987). Prokaryotic DNA replication mechanisms. *Philosophical transaction of the Royal Society of London, Series B, Biological Sciences* **317** (1187), 395-420.

Aves SJ (2009). DNA replication initiation. *Methods in Molecular Biology* **521**, 3-17.

Bambara RA, Murant RS, Henricksen LA (1997). Enzymes and reactions at the DNA replication fork. *The Journal of Biological Chemistry* **272** (8), 4647-4650.

Barry ER and Bell SD (2006). DNA replication in the Archaea. *Microbiology and Molecular Biology Reviews* **70** (4), 876-887.

Bell SP and Stillman B (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128-134.

Bell SP and Dutta A (2002). DNA replication in eukaryotic cells. *Annual Reviews in Biochemistry* **71**, 333-374.

Bochman ML and Schwacha A (2009). The MCM complex: Unwinding the mechanism of a replicative helicase. *Microbiology and Molecular Biology Reviews* **73** (4), 652-683.

Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, Robert C, Azza S, Sun S, Rossmann MG, Suzan-Monti M, La Scola B, Koonin EV, Raoult D (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *PNAS* **106** (51), 21848-21853.

Boyer M, Madoui MA, Gimenez G, La Scola B, Raoult D (2010). Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4th domain of life including giant viruses. *Plos One* **5** (12): e15530.

Bruck I and O'Donnell M (2001). The ring-type polymerase sliding clamp family. *Genome Biology* **2** (1): 3001.1-3001.3.

Bruning JB and ShamooY (2004). Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-δ p66 subunit and Flap Endonuclease-1. *Structure* **12**, 2209-2219.

Bryant JA, Moore K, Aves SJ (2001). Origins and complexes: the initiation of DNA replication. *Journal of Experimental Botany* **52** (355), 193-202.

Busso D, Peleg Y, Heidebrecht T, Romier C, Jacobovitch Y, Dantes A, Salim L, Troesch E, Schuetz A, Heinemann U, Folkers GE, Geerlof A, Wilmanns M, Polewacz A, Quedenau C, Büssow K, Adamson R, Blagova E, Walton J, Cartwright JL, Bird LE, Owens RJ, Berrow NS, Wilson KS, Sussman JL, Perrakis A, Celie PHN (2011). Expression of protein complexes using multiple *Escherichia coli* protein co-expression systems: A benchmarking study. *Journal of Structural Biology* **175**, 159-170.

Cann AJ (2001). Viral Genome. *Encyclopedia of Life Sciences.* 1-10.

Cann AJ (2003). Viral Replication. *Encyclopedia of Life Sciences.* 1-9.

Chapados BR, Hosfield DJ, Han S, Qiu J, Yelent B, Shen B, Tainer JA (2004). Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* **116**, 39-50.

Chilkova O, Stenlund P, Isoz I, Stith CM, Grabowski P, Lundstrom EB, Burgers PM, Johnasson E (2007). The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Research* **35** (19), 6588-6597.

Claverie JM (2006). Viruses take center stage in cellular evolution. *Genome Biology* **7**: 110.

Claverie JM, Ogata H, Audic S, Abergel C, Suhre K, Fournier PE (2006). Mimivirus and the emerging concept of "giant" virus. *Virus Research* **117**, 133-144.

Claverie JM and Abergel C (2009). Mimivirus and its virophage. *Annual Reviews in Genetics* **43**, 49-66.

Claverie JM, Abergel C, Ogata H (2009a). Mimivirus. *Current Topics in Microbiology and Immunology* **328**, 89-121.

Claverie JM, Grzela R, Lartigue A, Bernadac A, Nitsche S, Vacelet J, Ogata H, Abergel C (2009b). Mimivirus and Mimiviridae: Giant viruses with an increasing number of potential hosts, including corals and sponges. *Journal of Invertebrate Pathology* **101**, 172-180.

Claverie JM and Abergel C (2010). Mimivirus: the emerging paradox of quasi-autonomous viruses. *TRENDS in Genetics* **26**, 431-437.

Colson P and Raoult D (2010). Gene repertoire of amoeba-associated giant viruses. *Intervirology* **53**, 330-343.

Colson P, Gimenez G, Boyer MI, Fournous G, Raoult D (2011). The giant *Cafeteria roenbergensis* virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of life. *Plos One* **6** (4): e18935.

Dare RK, Chittaganpitch M, Erdman DD (2008). Screening pneumonia patients for Mimivirus. *Emerging Infectious Diseases* **14** (3), 465-467.

Da Silva M, Shen L, Tcherepanov V, Watson C, Upton C. (2006). Predicted function of the vaccinia virus G5R protein. *Bioinformatics* **22** (23), 2846-2850.

De Biasio A, Campos-Olivas R, Sanchez R, Lopez-Alonso JP, Pantoja-Uceda D, Merino N, Villate M, Martin-Garcia JM, Castillo F, Luque I, Blanco FJ (2012). Proliferating Cell Nuclear Antigen (PCNA) interactions in solution studied by NMR. *PLOS One* **7** (11), e48390.

DePamphilis ML (1993). Eukaryotic DNA replication: Anatomy of an origin. *Annual Reviews of Biochemistry* **62**, 29-63.

Deresinski S (2010). Marseillevirus: A DNA scavenger. *Clinical Infectious Diseases* **51** (3), iii-iv.

Desnues C and Raoult D (2010). Inside the lifestyle of the virophage. *Intervirology* **53** (5), 293-303.

Dickey Zakaib G (2011). Out on a limb: From giant viruses to unexplained marine DNA, there may be more to the tree of life than the three domains currently defined. *Nature* **476**, 20-21.

Domingo E (2007). Virus Evolution. *Encyclopedia of Life Sciences*. 1-9.

Filee J, Siguier P, Chandler M (2006). I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *TRENDS in Genetics* **23** (1), 10-15.

Filee J, Pouget N, Chandler M (2008). Phylogenetic evidence for extensive lateral acquisition of cellular genes by nucleocytoplasmic large DNA viruses. *BMC Evolutionary Biology* **8**: 320.

Filee J and Chandler M (2010). Gene exchange and the origin of giant viruses. *Intervirology* **53**, 354-361.

Fischer MG, Allen MJ, Wilson WH, Suttle CA (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. *PNAS* **107** (45), 19508-19513.

Fischer MG (2012). Sputnik and Mavirus: More than just satellite viruses. *Nature Reviews Microbiology* **10**, 78.

Flugel RM (2010). The precellular scenario of genovirions. *Virus Genes* **40**, 151-154.

Forterre P (2010). Giant Viruses: Conflicts in revisiting the virus concept. *Intervirology* **53**, 362-378.

Forterre P, Filee J, Myllykallio H (2000). Origin and evolution of DNA and DNA replication machineries. *Madame Curie Bioscience Database, Landes Bioscience.*

Gerbi SA, Strezoska Z, Waggener JM (2002). Initiation of DNA replication in multicellular eukaryotes. *Journal of Structural Biology* **140**, 17-30.

Gomes XV and Burgers PMJ (2000). Two modes of FEN1 binding to PCNA regulated by DNA. *The EMBO Journal* **19** (14), 3811-3821.

Grabowski B and Kelman Z (2003). Archaeal DNA replication: Eukaryal proteins in a Bacterial context. *Annual Reviews of Microbiology* **57**, 487-516.

Harper DR (2011). Viruses: Biology, applications, and control. First Edition. *Garland Science publications, New York.*

Harper DR (2012). Virus Replication. *Encyclopedia of Life Sciences*. 1-9.

Hickey RJ, Hoelz DJ, Malkas LH (2003). DNA replication: Mammalian. *Encyclopedia of Life Sciences.* 1-9.

Hosfield DJ, Mol CD, Chen B, Tainer JA (1998). Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: Coupling DNA and PCNA binding to FEN-1 activity. *Cell* **95**, 135-146.

Iyer LM, Balaji S, Koonin EV, Aravind L (2006). Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Research* **117**, 156-184.

Jonsson ZO, Hindges R, Hubscher U (1998). Regulation of DNA replication and repair proteins through interaction with the front side of proliferating cell nuclear antigen. *The EMBO Journal* **17** (8), 2412-2425.

Kaiser MW, Lyamicheva N, Ma W, Miller C, Neri B, Fors L, Lyamichev VI (1999). A comparison of Eubacterial and Archaeal structure-specific 5’-exonucleases. *The Journal of Biological Chemistry* **274**, 21387-21394.

Karp G (2009). Cell and Molecular biology: Concepts and experiments. 6th Edition. *John Wiley & Sons*, USA, 534-536.

Katsafanas GC and Moss B (2007). Linkage of transcription and translation within cytoplasmic poxvirus DNA factories provides a mechanism to coordinate viral and usurp host functions. *Cell Host Microbe* **2** (4), 221-228.

Kazlauskas D and Venclovas C (2011). Computational analysis of DNA replicases in double-stranded DNA viruses: relationship with the genome size. *Nucleic Acids Research*, **39** (19): 8291–8305.

Kelly TJ and Spillman B (2006). Duplication of DNA in eukaryotic cells. *DNA replication and Human Disease, Cold Spring Harbour Laboratory Press*, 1-30.

Kirchmaier AL (2011). Ub-family modifications at the replication fork: Regulating PCNA-interacting components. *FEBS Letters* **585**, 2920-2928.

Koonin EV and Yutin N (2010). Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. *Intervirology* **53**, 284-292.

Kuznetsov YG, Xiao C, Sun S, Raoult D, Rossmann M, McPherson A (2010). Atomic force microscopy investigation of giant Mimivirus. *Virology* **404**, 127-137.

La Scola B, Marrie TJ, Auffray JP, Raoult D (2005). Mimivirus in pneumonia patients. *Emerging Infectious Disease* **11** (3), 449-452.

Legendre M, Santini S, Rico A, Abergel C, Claverie JM (2011). Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. *Virology Journal* **8**:99.

Leipe DD, Aravind L, Koonin EV (1999). Did DNA replication evolve twice independently? *Nucleic Acids Research* **27** (17), 3389-3401.

Liu H and Naismith JH (2009). A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expression and Purification* **63,** 102-111.

Maga G and Hubscher U (2003). Proliferating Cell Nuclear Antigen (PCNA): A dancer with many partners. *Journal of Cell Science* **116**, 3051-3060.

Majka J and Burgers PMJ (2004). The PCNA–RFC families of DNA clamps and clamp loaders. *Progress in Nucleic Acid Research and Molecular Biology* **78**, 227-260.

Mailand N, Gibbs-Seymour I, Bekker-Jensen S (2013). Regulation of PCNA-protein interactions for genome stability. *Nature Reviews, Molecular Cell Biology, DNA Damage* **14**, 269-282.

McKenna R and Faulkner L (2001). Virus Structure. *Encyclopedia of Life Sciences.* 1-9.

McNally R, Bowman GD, Goedken ER, O'Donnell M, Kuriyan J (2010). Analysis of the role of PCNA-DNA contacts during clamp loading. *BMC Structural Biology* **10** (3), 1-14.

Mechali M (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. *Nature Reviews, Molecular Cell Biology* **11**, 728-738.

Minor PD (2007). Viruses. *Encyclopedia of Life Sciences.* 1-10.

Moldovan GL, Pfander B, Jentsch S (2007). PCNA, the maestro of the replication fork. *Cell* **129**, 665-679.

Moliner C, Fournier PE, Raoult D (2010). Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiology Review* **34**, 281-294.

Monier A, Claverie JM, Ogata H (2007). Horizontal gene transfer and nucleotide compositional anomaly in large DNA viruses. *BMC Genomics* **8**: 456.

Moreira D and Brochier-Armanet C (2008). Giant viruses, giant chimeras: The multiple evolutionary histories of Mimivirus genes. *BMC Evolutionary Biology* **8**: 12.

Mossi R, Keller RC, Ferrari E, Hubscher U (2000). DNA polymerase switching: II. Replication factor C abrogates primer synthesis by DNA polymerase α at a critical length. *Journal of Molecular Biology* **295** (4), 803-814.

Mrazek J and Karlin S (2007). Distinctive features of large complex virus genomes and proteomes. *PNAS* **104** (12), 5127-5132.

Mutsafi Y, Zauberman N, Sabanay I, Minsky A (2010). Vaccinia-like cytoplasmic replication of the giant Mimivirus. *PNAS* **107** (13), 5978-5982.

Nasir A, Kim KM, Caetano-Anolles G (2012). Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. *BMC Evolutionary Biology* **12**: 156.

Ogata H and Claverie JM (2007). Unique genes in giant viruses: Regular substitution pattern and anomalously short size. *Genome Research* **17**, 1353-1361.

Oh J and Broyles SS (2005). Host cell nuclear proteins are recruited to cytoplasmic vaccinia virus replication complexes. *Journal of Virology* **79** (20), 12852-12860.

O'Reilly DR, Crawford AM, Miller LK (1989). Viral proliferating cell nuclear antigen. *Nature* **337**: 606.

Pavlov YI and Shcherbakova PV (2010). DNA polymerases at the eukaryotic fork-20 years later. *Mutation Research* **685**, 45-53.

Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* **306**, 1344-1350.

Raoult D (2005). The journey from *Rickettsia* to Mimivirus. *ASM News* **71** (6), 278-284.

Raoult D, Renesto P, Brouqui P (2006). Laboratory infection of a technician by Mimivirus. *Annals of Internal Medicine* **144** (9), 702-703.

Raoult D and Forterre P (2008). Redefining viruses: Lessons from Mimivirus. *Nature Reviews, Microbiology* **6**: 315.

Raoult D (2010). Giant viruses from amoebae in a post-Darwinist viral world. *Intervirology* **53**, 251-253.

Raoult D and Boyer M (2010). Amoebae as genitors and reservoirs of giant viruses. *Intervirology* **53**, 321-329.

Remus D and Diffley JFX (2009). Eukaryotic DNA replication control: Lock and load, then fire. *Current Opinion in Cell Biology* **21**, 771-777.

Renesto P, Abergel C, Decloquement P, Moinier D, Azza S, Ogata H, Fourquet P, Gorvel JP, Claverie JM (2006). Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. *Journal of Virology* **80** (23), 11678-11685.

Robinson NP and Bell SD (2005). Origins of DNA replication in the three domains of life. *FEBS Journal* **272**, 3757-3766.

Ruiz-Saenz J and Rodas JP (2010). Viruses, virophages, and their living nature. *Acta Virologica* **54**, 85-90.

Rumbaugh JA, Henricken LA, DeMott MS, Bambara RA (1999). Cleavage of substrates with mismatched nucleotides by Flap Endonuclease 1: Implications for mammalian Okazaki fragment processing. *The Journal of Biological Chemistry* **274**, 14602-14608.

Sakurai S, Kitano K, Yamaguchi H, Hamada K, Okada K, Fukuda K, Uchida M, Ohtsuka E, Morioka H, Hakoshima T (2005). Structural basis for recruitment of human Flap Endonuclease 1 to PCNA. *The EMBO Journal* **24** (4), 683-693.

Samson T, Henricksen LA, Bambara RA (2000). Mechanism whereby Proliferating Cell Nuclear Antigen stimulates Flap Endonuclease 1. *The Journal of Biological Chemistry* **275** (14), 10498-10505.

Scorah J, Dong M-Q, Yates III JR, Scott M, Gillespie D, McGowan CH (2008). A conserved Proliferating Cell Nuclear Antigen-interacting protein sequence in Chk1 is required for checkpoint function. *The Journal of Biological Chemistry* **283** (25), 17250-17259.

Senkevich TG, Koonin EV, Moss B (2009). Predicted poxvirus FEN1-like nuclease required for homologous recombination, double-strand break repair and full-size genome formation. *PNAS Microbiology* **106** (42), 17921-6.

Shen Z and Prasanth SG (2012). Emerging players in the initiation of eukaryotic DNA replication. *Cell Division* **7** (22), 1-15.

Sinkovics JG (2011). Horizontal gene transfer with or without cell fusions in all categories of the living matter. *Advances in Experimental Medicine and Biology, Cell Fusion in Health and Disease* **950**, 5-89.

Stillman B (1996). Cell cycle control of DNA replication. *Science* **272** (5293), 1659-1663.

Suhre K (2005). Gene and genome duplications in *Acanthoamoeba polyphaga mimivirus*. *Journal of Virology* **79** (22), 14095-14101.

Suzan-Monti M, La Scola B, Barrassi L, Espinosa L, Raoult D (2007) Ultrastructural characterization of the giant volcano-like virus factory of Acanthamoeba polyphaga Mimivirus. *PLoS One* 2:e328.

Takisawa H, Mimura S, Kubota Y (2000). Eukaryotic DNA replication: From pre-replication complex to initiation complex. *Current Opinion in Cell Biology* **12**, 690-696.

Tsurimoto T (1999). PCNA binding proteins. *Frontiers in Bioscience* **4**, 849-858.

Ulrich HD and Takahashi T (2013). Readers of PCNA modifications. *Chromosoma*, 1-16.

Van Etten JL, Lane LC, Dunigan DD (2010). DNA Viruses: The really big ones (Giruses). *Annual Review of Microbiology* **64**, 83-99.

Van Etten JL (2011). Another really, really big virus. *Viruses* **3**, 32-46.

Vas A and Leatherwood J (2000). Where does DNA replication start in *Archaea*? *Genome Biology* **1** (3), reviews 1020.1–1020.4.

Vivona JB and Kelman Z (2003). The diverse spectrum of sliding clamp interacting proteins. *FEBS Letters* **546**, 167-172.

Wang TA and Li JJ (1995). Eukaryotic DNA replication. *Current Opinion in Cell Biology* **7**, 414-420.

Warbrick E (1998). PCNA binding through a conserved motif. *BioEssays* **20,** 195-199.

Warbrick E (2000). The Puzzle of PCNA's many partners. *BioEssays* **22**, 997-1006.

Wessner DR (2010). The origins of Viruses: How did viruses evolve? Are they a streamlined form of something that existed long ago, or an ultimate culmination of smaller genetic elements joined together? *Cell Origins and Metabolism. Nature Education* **3** (9), 37.

Williams TA, Embley TM, Heinz E (2011). Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. *Plos One* **6** (6): e21080.

Winter JA and Bunting KA. Rings in the extreme: PCNA interaction and adaptations in the Archaea. *Archaea*, 1-8.

Yoshida T, Claverie JM, Ogata H (2011). Mimivirus reveals Mre11/Rad50 fusion proteins with a sporadic distribution in eukaryotes, bacteria, viruses and plasmids. *Virology Journal* **8**: 427.

Yutin N and Koonin EV (2009). Evolution of DNA ligases of nucleo-cytoplasmic large DNA viruses of eukaryotes: A case of hidden complexity. *Biology Direct* **4**: 51.

Yutin N, Wolf YI, Raoult D, Koonin EV (2009). Eukaryotic large nucleocytoplasmic DNA viruses: Clusters of orthologous genes and reconstitution of viral genome evolution. *Virology Journal* **6**, 223.

Yutin N and Koonin EV (2012). Hidden evolutionary complexity of Nucleo-Cytoplasmic Large DNA viruses of eukaryotes. *Virology Journal* **9**: 161.

Zegerman P (2013). DNA replication: Polymerase epsilon as a non-catalytic converter of helicase. *Current Biology* **23** (7), 273-276.

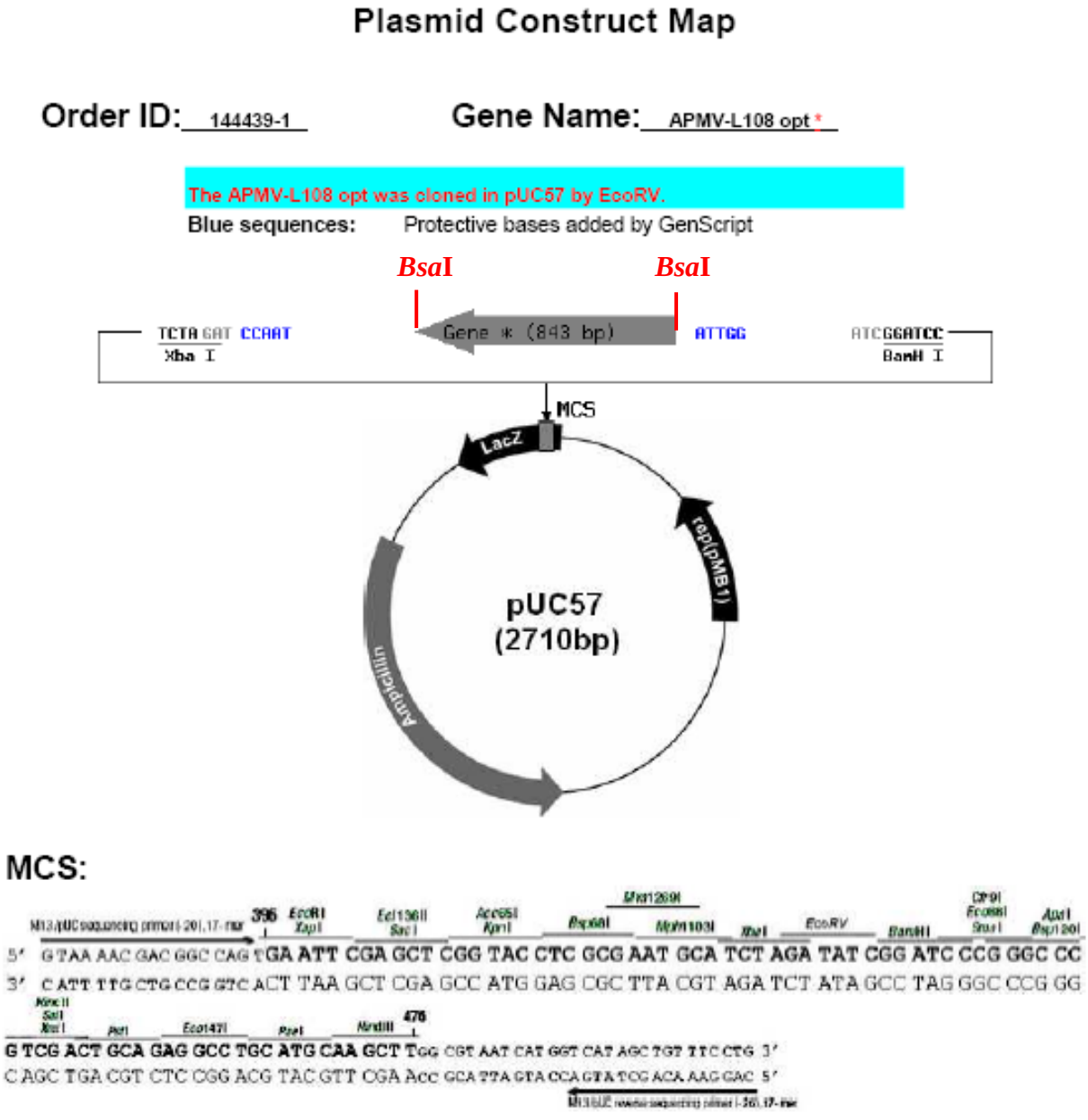
Zheng L, Dai H, Qiu J, Huang Q, Shen B (2007). Disruption of the FEN1/PCNA interaction results in DNA replication defects, pulmonary hypoplasia, pancytopenia, and newborn lethality in mice. *Molecular and Cellular Biology* **27** (8), 3176-3186.

Zheng L, Dai H, Hegde ML, Zhou M, Guo Z, Wu X, Su L, Zhong X, Mitra S, Huang Q, Kemstine KH, Pfeifer GP, Shen B (2011). Fen1 mutations that specifically disrupt its interaction with PCNA cause aneuploidy-associated cancer. *Cell Research* **21**, 1052-1067.

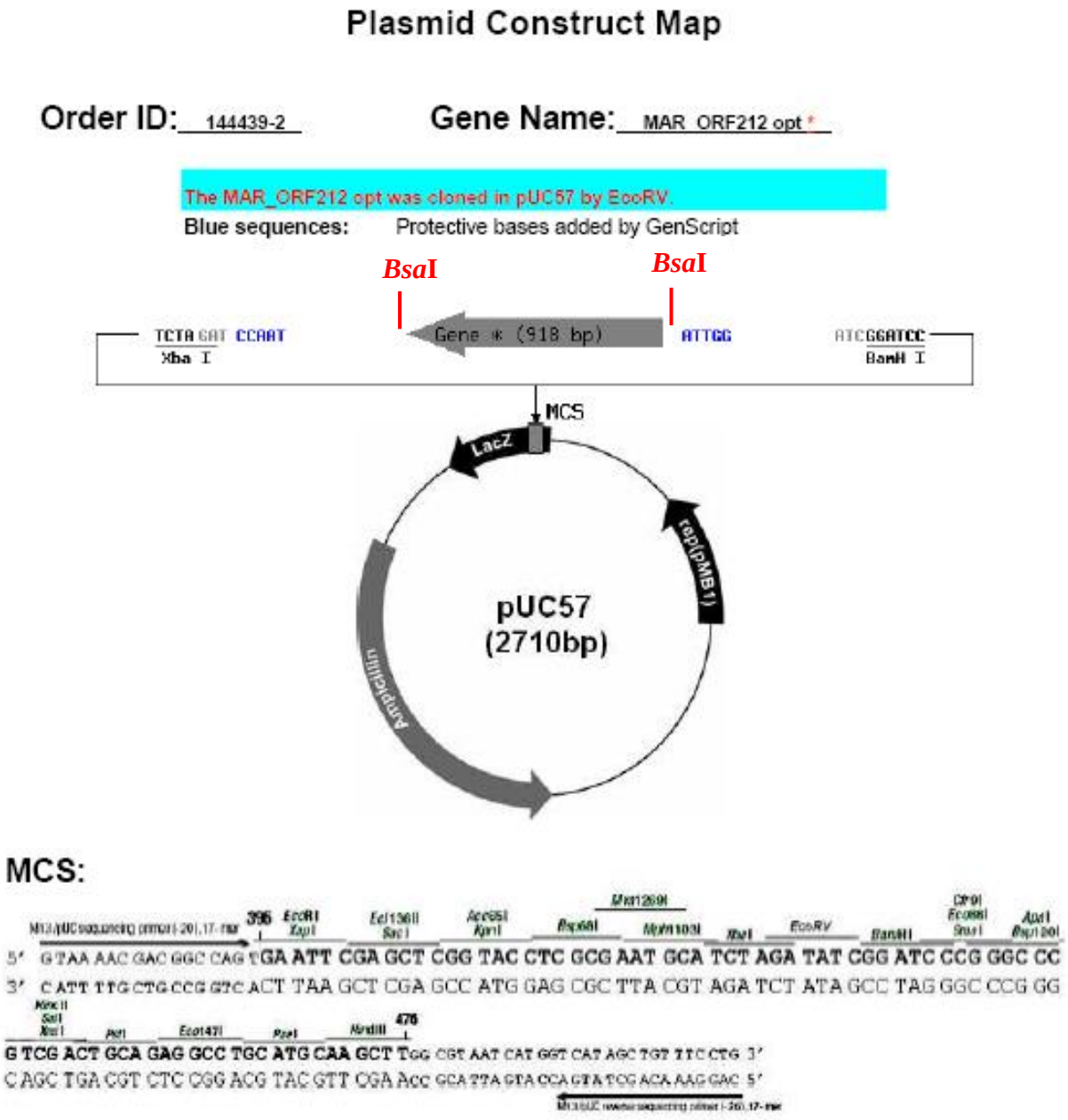
APPENDIX

A.1 Construct Maps of the Synthetic Genes (as provided by GenScript, UK)

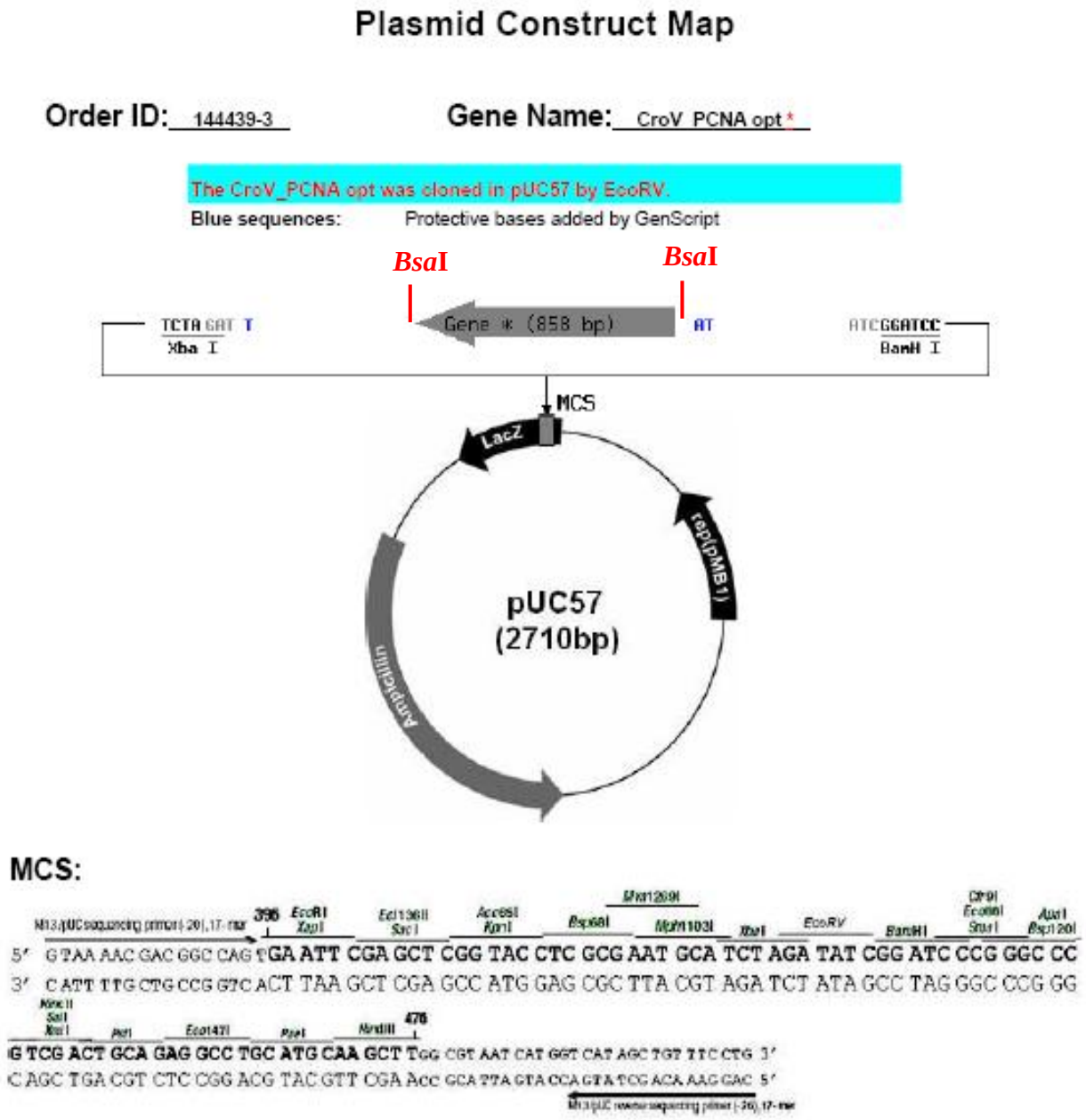
A.1.1 Mimivirus (APMV_L108) PCNA



A.1.2 Marseillevirus (MAR_ORF212) PCNA



A.1.3 CroV (CroV_219) PCNA



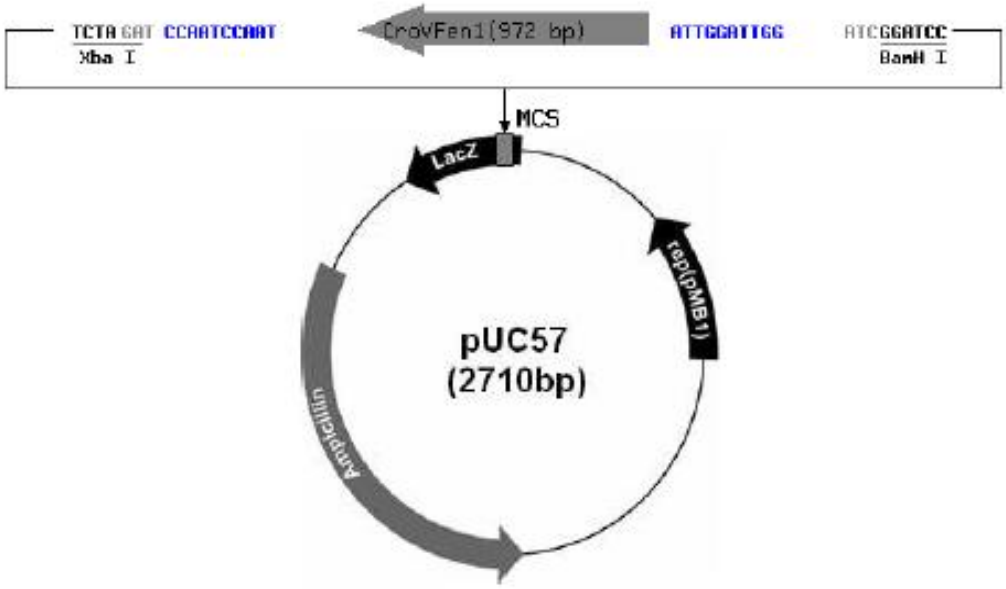
A.1.4 CroV (CroV_037) FEN1

Plasmid Construct Map

Order ID: 164105-1 **Gene Name:** CroVFen1

The CroVFen1 was cloned in pUC57 by EcoRV.

Blue sequences: Protective bases added by GenScript



MCS:

A.2 Maps of the pASK-IBA17plus PCNA constructs

(Modified from the original '*pASK-IBA17plus*' construct map, IBA, UK)

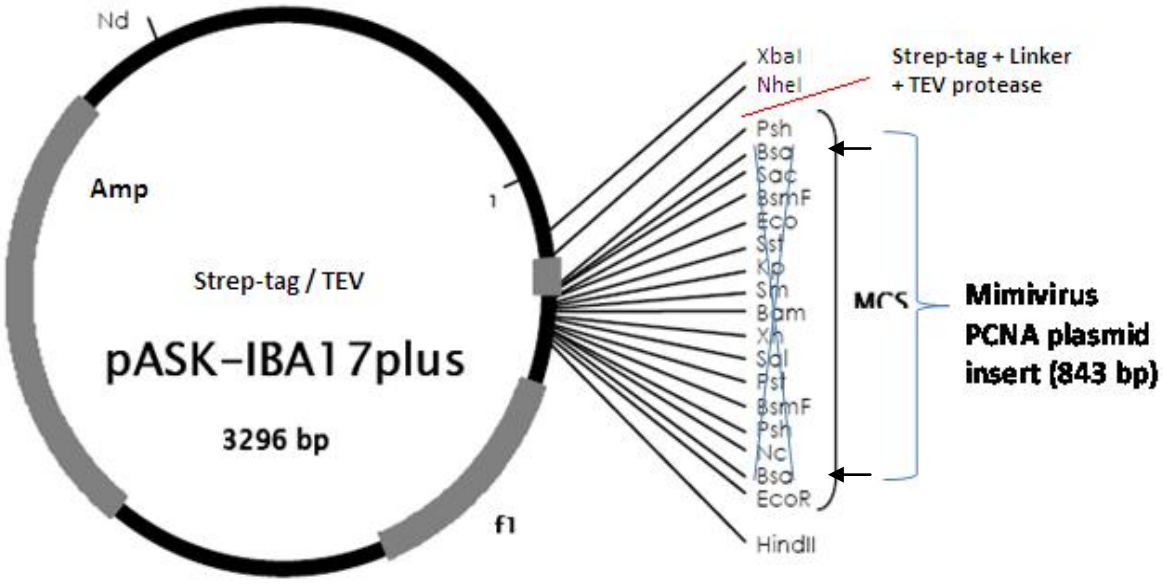
A.2.1 pASK-IBA17plus_Mimivirus (APMV_L108)_PCNA

Plasmid Name: pASK-IBA17plus_APMV_L108_PCNA_SENSE

Plasmid size: 4062 bp

Cloning Info: Digested *Bsa*I 843 bp fragment from PCNA plasmid construct APMV_L108 (Mimivirus) into digested *Bsa*I pASK-IBA17plus vector. Sense orientation.

Map:



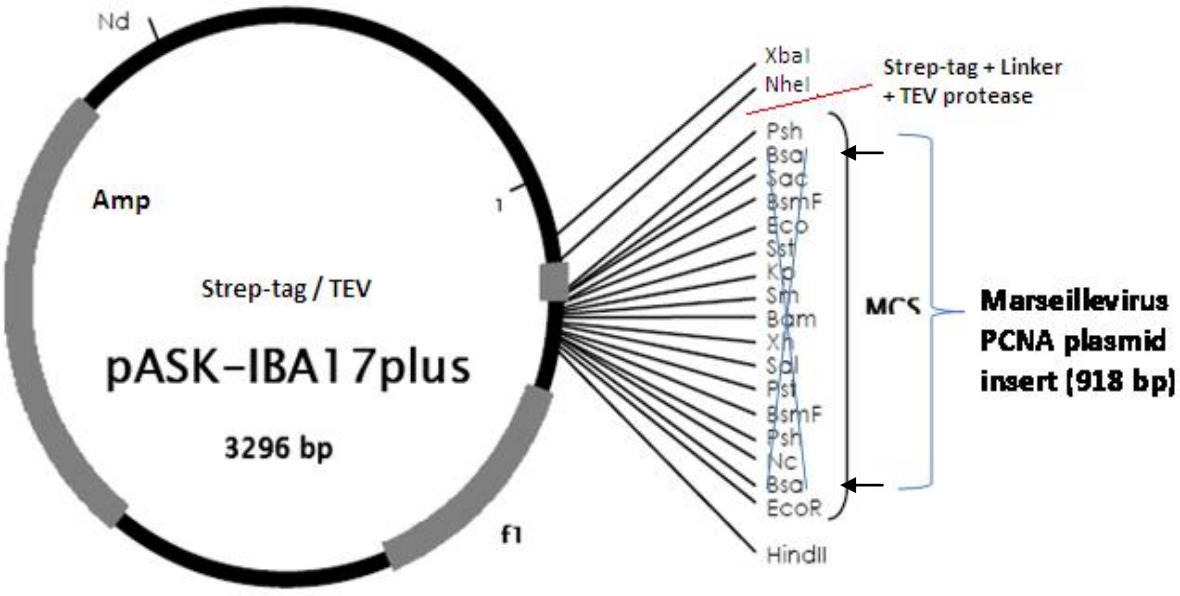
A.2.2 pASK-IBA17plus_Marseillevirus (MAR_ORF212)_PCNA

Plasmid Name: pASK-IBA17plus_ MAR_ORF212_PCNA_SENSE

Plasmid size: 4137 bp

Cloning Info: Digested *Bsa*I 918 bp fragment from PCNA plasmid construct MAR_ORF212 (Marseillevirus) into digested *Bsa*I pASK-IBA17plus vector. Sense orientation.

Map:



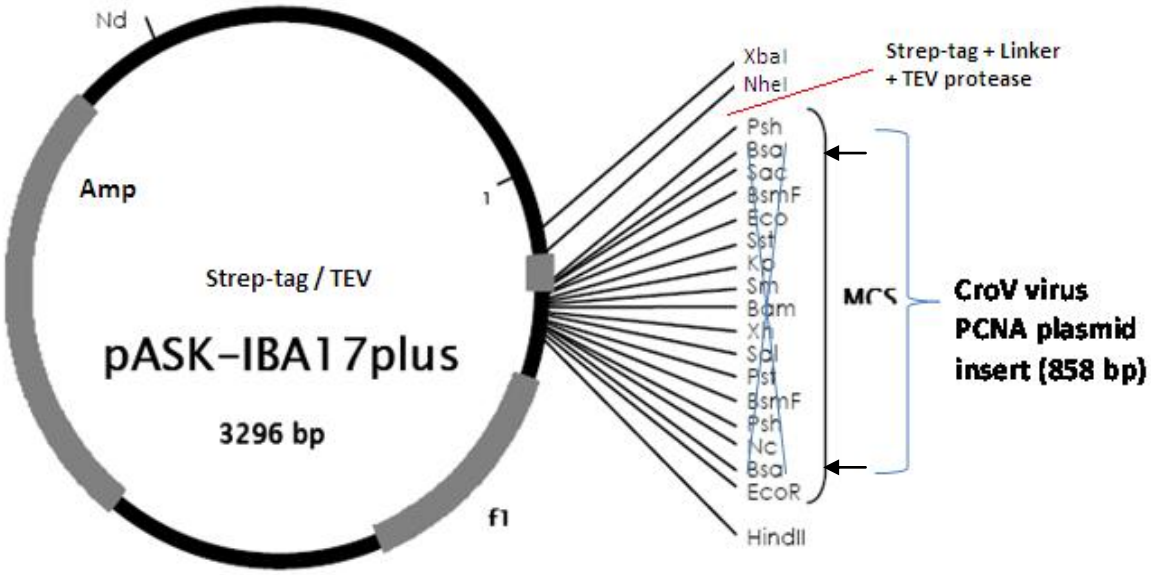
A.2.3 pASK-IBA17plus_CroV (CroV_219)_PCNA

Plasmid Name: pASK-IBA17plus_ CroV (CroV_219)_PCNA_SENSE

Plasmid size: 4077 bp

Cloning Info: Digested *Bsa*I 858 bp fragment from PCNA plasmid construct CroV (CroV virus) into digested *Bsa*I pASK-IBA17plus vector. Sense orientation.

Map:



A.3 Maps of the pEHISTEV PCNA constructs

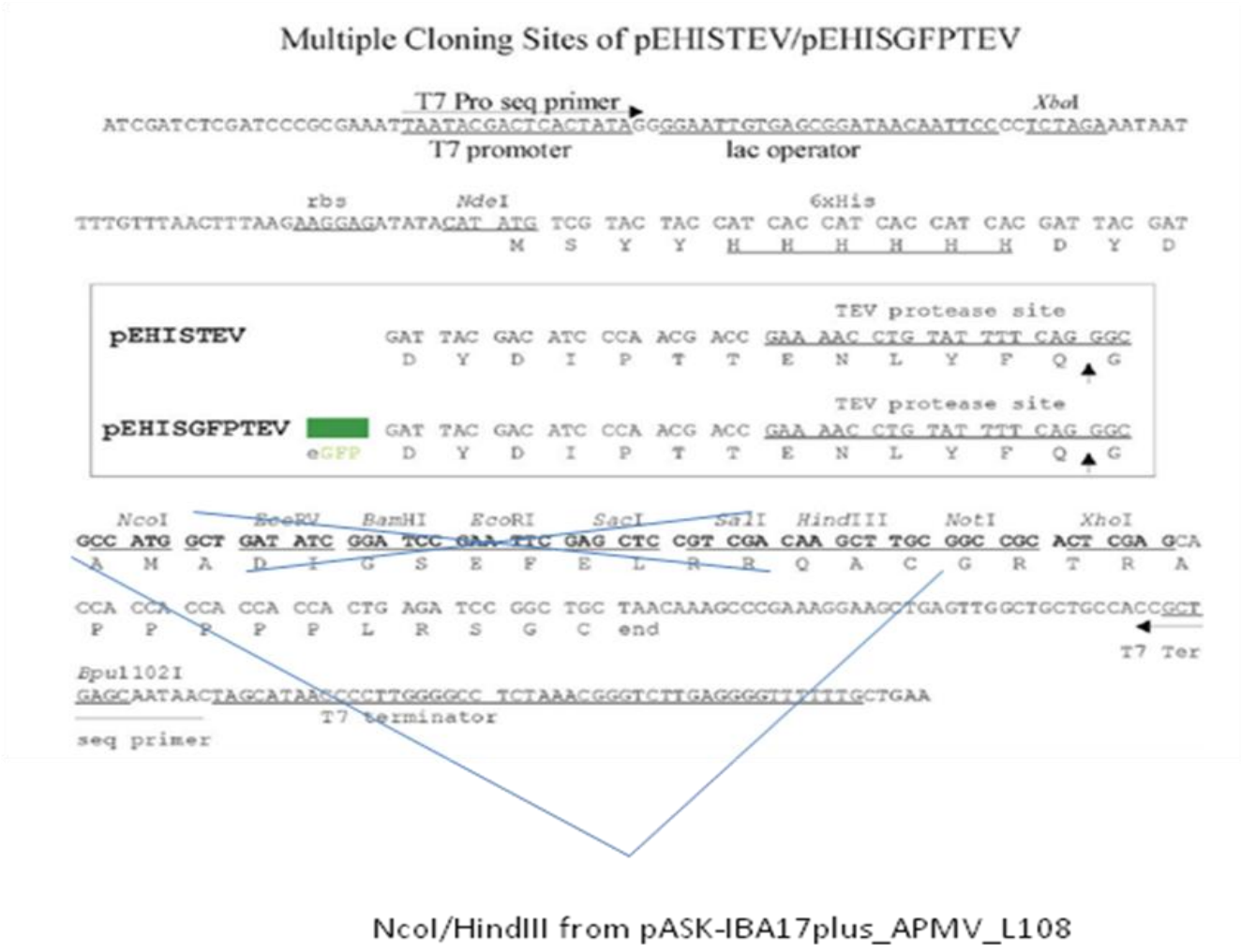
(Modified from Fig. 1: Multiple cloning sites of pEHISTEV/pEHISGFPTEV; *Liu H and Naismith JH, 2009*)

A.3.1 pEHISTEV_Mimivirus (APMV_L108)_PCNA

Plasmid Name: pEHISTEV_APMV_L108_SENSE

Cloning Info: Digested *Nco*I/*Hind*III 840 bp fragment from PCNA plasmid construct pASK-IBA17plus_APMV_L108 (Mimivirus) into digested *Nco*I/*Hind*III pEHISTEV vector. Sense orientation.

Map:

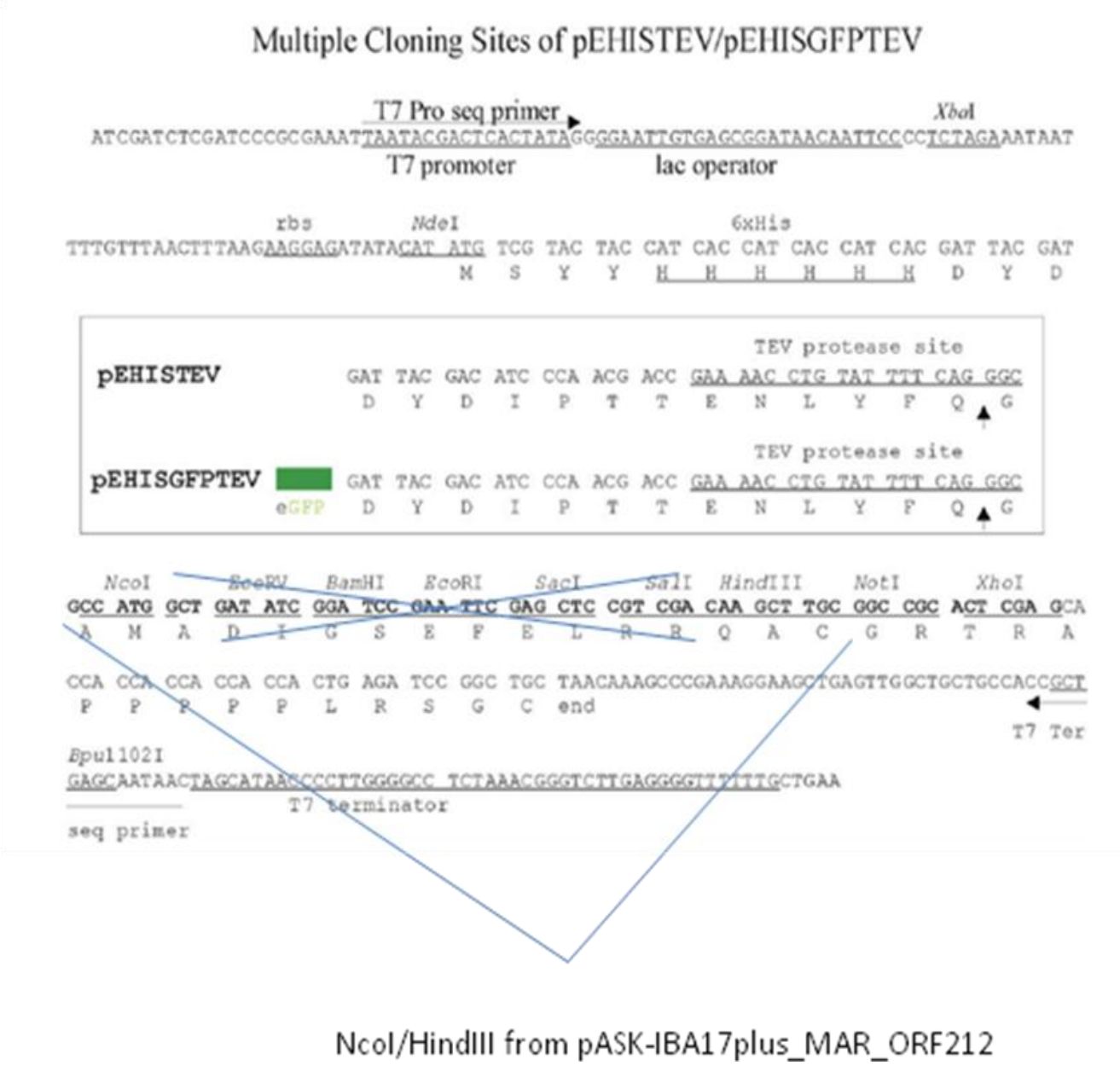


A.3.2 pEHISTEV _Marseillevirus (MAR_ORF212)_PCNA

Plasmid Name: pEHISTEV_MAR_ORF212_SENSE

Cloning Info: Digested *Nco*I/*Hind*III 920 bp fragment from PCNA plasmid construct pASK-IBA17plus_MAR_ORF212 (Marseillevirus) into digested *Nco*I/*Hind*III pEHISTEV vector. Sense orientation.

Map:



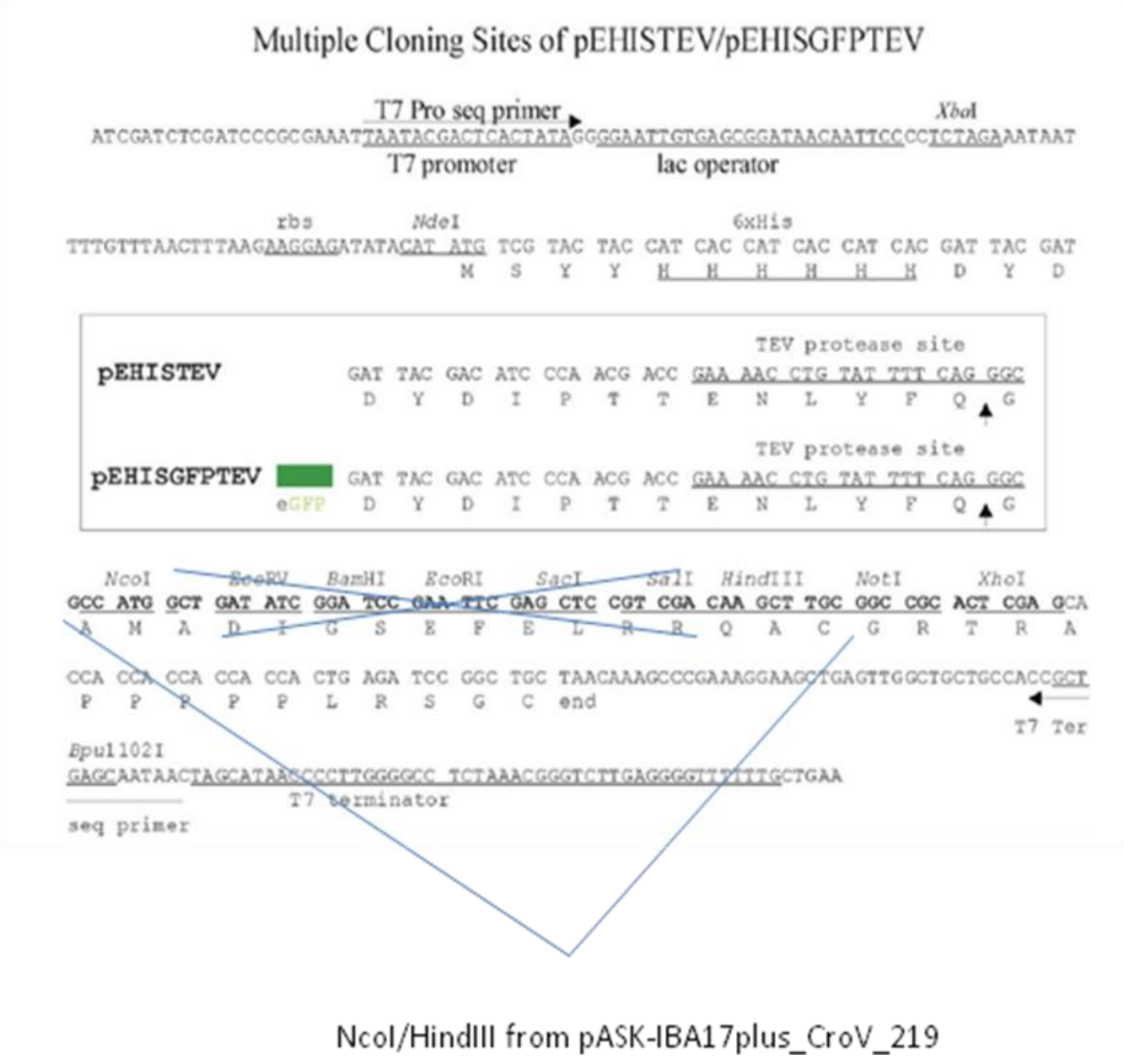
NcoI/HindIII from pASK-IBA17plus_MAR_ORF212

A.3.3 pEHISTEV _CroV(CroV_219)_PCNA

Plasmid Name: pEHISTEV_CroV_219_SENSE

Cloning Info: Digested *Nco*I/*Hind*III 860 bp fragment from PCNA plasmid construct pASK-IBA17plus_CroV (Cafeteria roenbergensis) into digested *Nco*I/*Hind*III pEHISTEV vector. Sense orientation.

Map:



NcoI/HindIII from pASK-IBA17plus_CroV_219

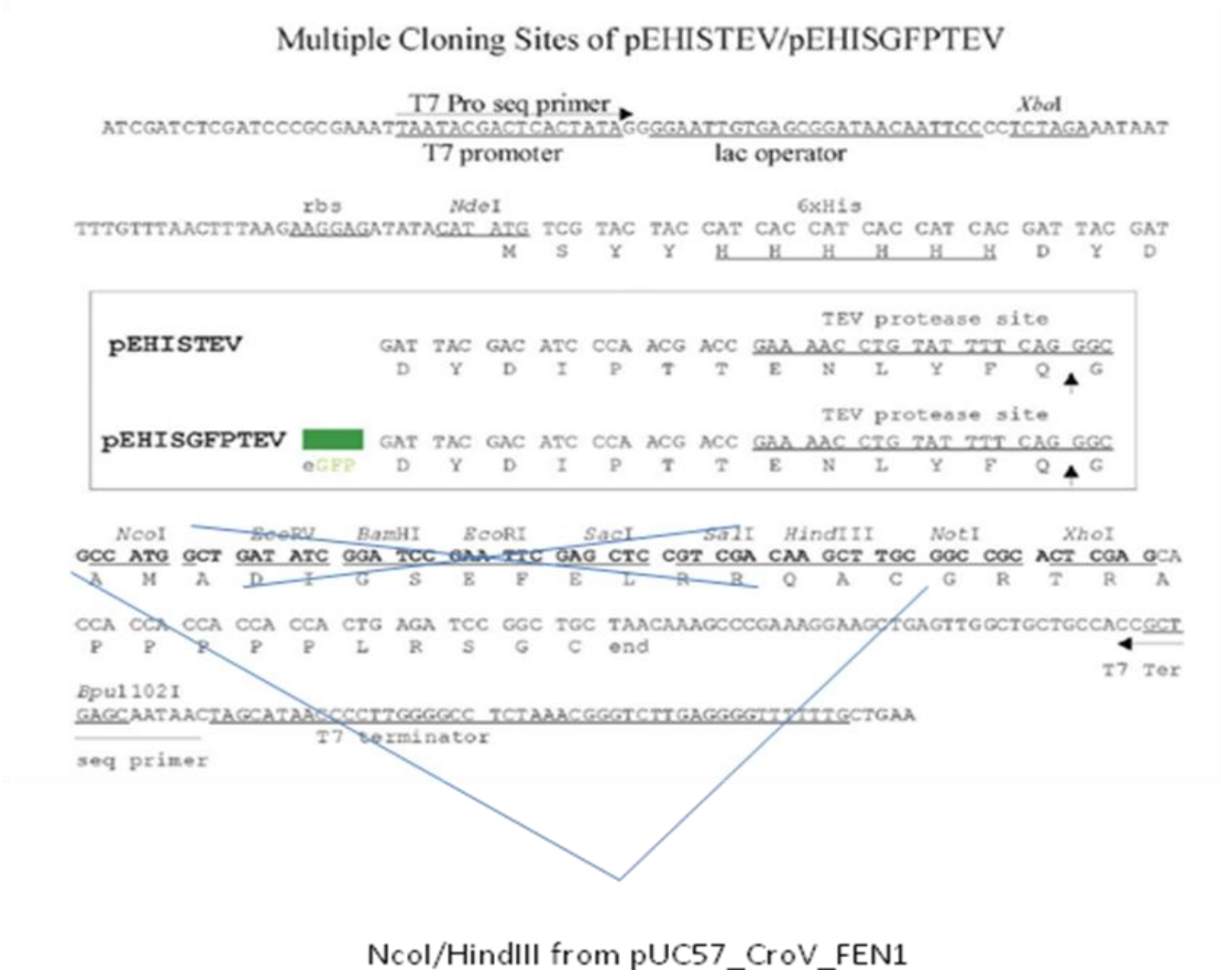
A.4 Construct Map of the pEHISTEV and pEHISGFPTEV CroV (CroV_037)_FEN1

(Modified from Fig. 1: Multiple cloning sites of pEHISTEV/pEHISGFPTEV; *Liu H and Naismith JH, 2009*)

Plasmid Name: pEHISTEV_CroV_037_SENSE & pEHISGFPTEV_CroV_037_SENSE

Cloning Info: Digested *Nco*I/*Hind*III 970 bp fragment from FEN1 plasmid construct pUC57_CroV (Cafeteria roenbergensis) into digested *Nco*I/*Hind*III pEHISTEV and pEHISGFPTEV vectors. Sense orientation.

Map:



NcoI/HindIII from pUC57_CroV_FEN1

A.5 Maps of the pETDuet-1 and pCDFDuet-1 constructs

(Modified from the original '*pETDuet-1*' and '*pCDFDuet-1*' construct map, Novagen, UK)

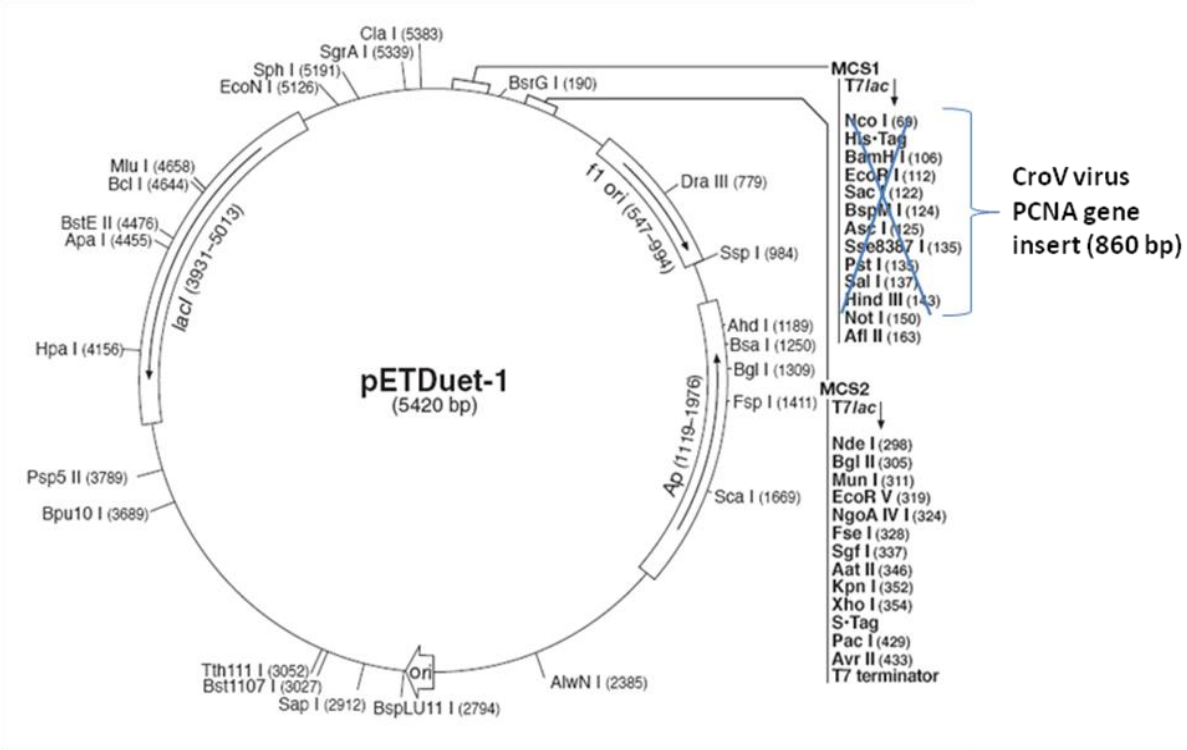
A.5.1 pETDuet-1_CroV_219_PCNA

Plasmid Name: pETDuet-1_ CroV (CroV_219)_PCNA_SENSE

Plasmid size: 6,200 bp

Cloning Info: Digested *Nco*I/*Hind*III 860 bp fragment from CroV PCNA construct (pEHISTEV_CroV_219_PCNA) into digested *Nco*I/*Hind*III pETDuet-1 vector. Sense orientation.

Map:



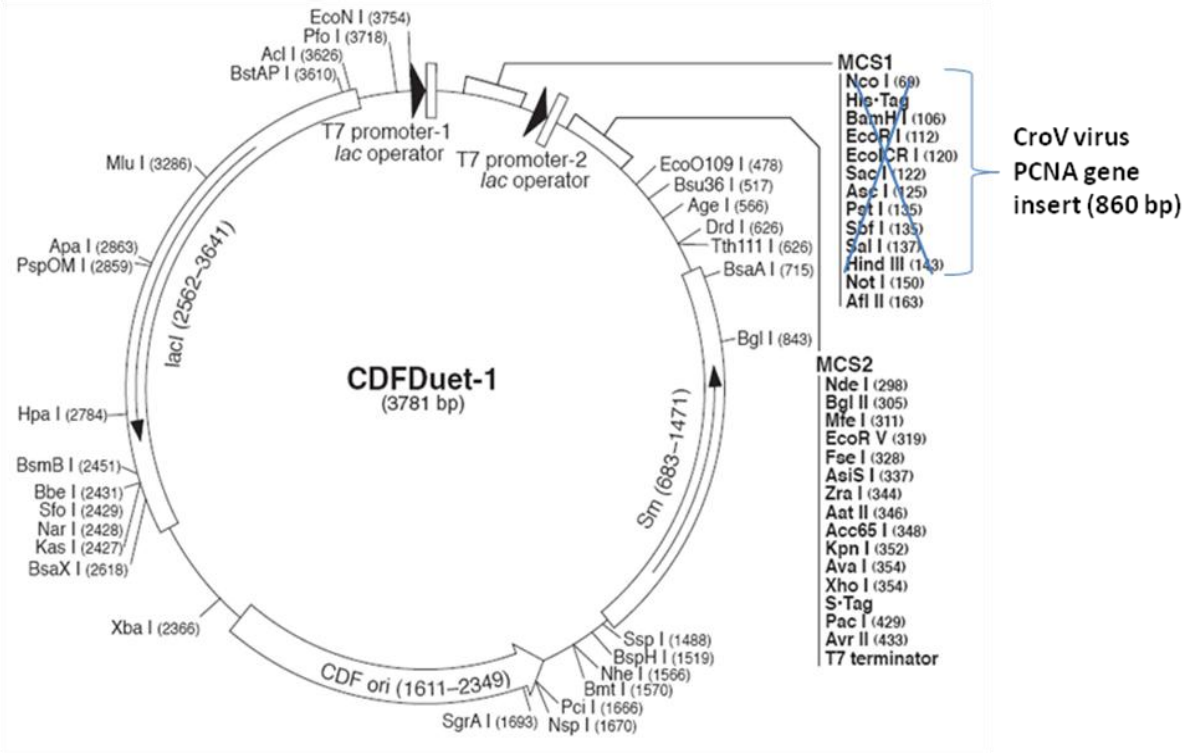
A.5.2 pCDFDuet-1_CroV_219_PCNA

Plasmid Name: pCDFDuet-1_ CroV (CroV_219)_PCNA_SENSE

Plasmid size: 4,567 bp

Cloning Info: Digested *Nco*I/*Hind*III 860 bp fragment from CroV PCNA construct (pEHISTEV_CroV_219_PCNA) into digested *Nco*I/*Hind*III pCDFDuet-1 vector. Sense orientation.

Map:



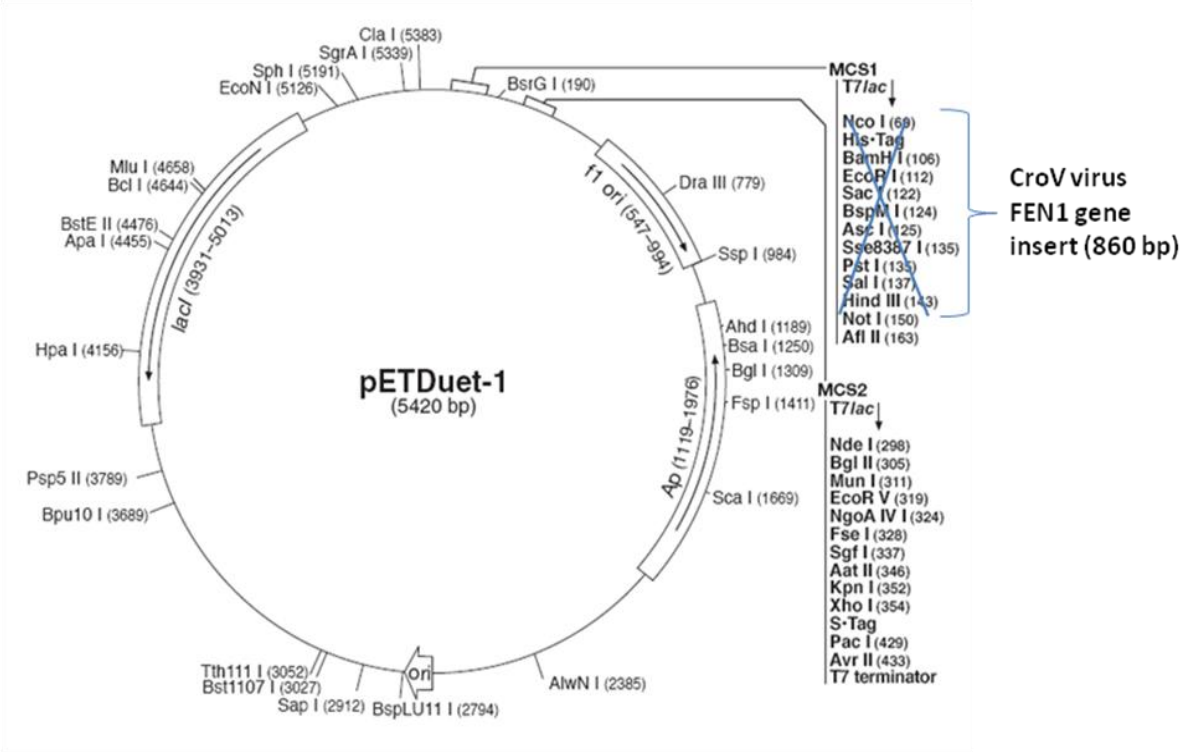
A.5.3 pETDuet-1_CroV_037_FEN1

Plasmid Name: pETDuet-1_ CroV_037_FEN1_SENSE

Plasmid size: 6,320 bp

Cloning Info: Digested *Nco*I/*Hind*III 970 bp fragment from CroV FEN1 construct (pEHISTEV_CroV_037_FEN1) into digested *Nco*I/*Hind*III pETDuet-1 vector. Sense orientation.

Map:



A.5.4 pCDFDuet-1_CroV_037_FEN1

Plasmid Name: pCDFDuet-1_ CroV_037_FEN1_SENSE

Plasmid size: 4,677 bp

Cloning Info: Digested *Nco*I/*Hind*III 970 bp fragment from CroV FEN1 construct (pEHISTEV_CroV_037_FEN1) into digested *Nco*I/*Hind*III pCDFDuet-1 vector. Sense orientation.

Map: