A Bioinformatic Exploration of Poxviruses

by

Melissa Elizabeth Da Silva
B.Sc., University of Victoria, 2002

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

Supervisory Committee

A Bioinformatic Exploration of Poxviruses

by

Melissa Elizabeth Da Silva
B.SC., University of Victoria, 2002

Supervisory Committee

Dr. Christopher Upton, (Department of Biochemistry and Microbiology)
Supervisor

Dr. Caren Helbing, (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Francis Nano, (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Ben Koop, (Department of Biology)
Outside Member

Dr. David Evans, (Department of Medical Microbiology and Immunology)
External Examiner

Abstract

Supervisory Committee
Dr. Christopher Upton, (Department of Biochemistry and Microbiology)
Supervisor
Dr. Caren Helbing, (Department of Biochemistry and Microbiology)
Departmental Member
Dr. Francis Nano, (Department of Biochemistry and Microbiology)
Departmental Member
Dr. Ben Koop, (Department of Biology)
Outside Member
Dr. David Evans, (Department of Medical Microbiology and Immunology)
External examiner

The overall theme of this dissertation is the genomic analysis of poxviruses using bioinformatics. The first analysis presented in this dissertation (Chapter 2) focuses on a new method for predicting which open reading frames (ORFs) in poxviruses are likely to be expressed. A measure that takes into account the amino acid and purine content of all predicted open reading frames (ORFs) in the genome was developed and when used on the vaccinia virus (VACV) strain Copenhagen genome (training case), the measure had a success rate of 94%. Using the measure on an extremely adenine and thymine rich entomopoxvirus (test case), 241 ORFs were found to be potentially expressed and 51 ORFs were likely not expressed although further biochemical experiments will be required to confirm this result.

The second analysis of this dissertation (Chapter 3) focuses on determining the nature of an interesting background pattern similar to a set of stripes that was observed while analyzing a self-dotplot of the molluscum contagiosum virus genome. These stripe regions were further analyzed and were found to have a nucleotide composition and amino acid usage that was different to the remainder of the genome. Given this differing nucleotide and amino acid usage, the genes contained in these stripe regions are thought to have been recently acquired from the host or another virus, making these regions similar to bacterial pathogenicity islands.

The third analysis of this dissertation (Chapter 4) focuses on predicting the function of "unknown" poxvirus proteins by using a hidden Markov model (HMM) comparison search tool to scan all "unknown" proteins in the VACV genome looking for any database matches that may have been missed by

conventional approaches (BLASTp and PSI-BLAST). One protein, the VACV G5R protein, in this scan showed a promising hit (96% probability) to an archaeal flap endonuclease (FEN-1) protein. A structural model of the G5R protein was created and subsequently compared to the crystal structure of the human FEN-1 protein and was found to be highly conserved in both secondary and tertiary structure and with three of the five main features of the FEN-1 protein including the active site suggesting that the G5R protein should be classified as a flap endonuclease protein.

Related to the analysis in Chapter 4, are the results presented in Chapter 5 of this dissertation that focus on locating a protein encoded by the VACV genome that is similar to proliferating cell nuclear antigen (PCNA). Knowing that the FEN-1 protein requires PCNA as an intermediary to contact DNA, the genome of VACV was scanned using InterProScan in order to identify any potential proteins that were similar to PCNA. One protein (VACV G8R) was identified and subsequently modeled and compared to the crystal structure of the human PCNA protein. The secondary and tertiary structure was highly conserved between the two proteins suggesting that the G8R protein should be classified as a sliding clamp similar to human PCNA.

Table of Contents

Supervisory Committee ii
Abstract iii
Table of Contents v
List of Tables vii
List of Figures viii
List of Abbreviations ix
Acknowledgments xi
Dedication xiii
1.0 Introduction 1
1.1 Poxvirus classification 1
1.2 History of variola virus (smallpox) 2
1.3 Variola virus disease progression 7
1.3.1 Strains of variola virus 8
1.3.2 "Ordinary"-type smallpox disease progression 8
1.3.3 Disease progression of variola major clinical types 10
1.3.4 Complications relating to variola infection 12
1.4 Vaccination and its complications 12
1.5 Virion Structure 16
1.6 Virus life-cycle 20
1.6.1 Entry 20
1.6.2 Uncoating 22
1.6.3 Gene expression 22
1.6.4 Replication 34
1.6.5 Assembly and release 39
1.7 Virus-host interactions 41
1.7.1 Inhibition of host macromolecular synthesis 42
1.7.2 Stimulation of host cell growth and prevention of apoptosis 43
1.7.3 Modulation of the host immune response 44
1.8 Bioinformatics and poxvirus genomes 47
1.8.1 Bioinformatics tools used in this dissertation 48
1.9 Dissertation outline 51
2.0 Using purine skews to predict genes in AT-rich poxviruses 53
2.1 Introduction 53
2.2 Methods 56
2.2.1 Purine skews 56
2.2.2 Purine/pyrimidine ratio comparison 56
2.3 Results and discussion 57
2.4 Conclusions 70
3.0 Host-derived pathogenicity islands in poxviruses 72
3.1 Introduction 72
3.2 Methods 75
3.2.1 Creation of dotplots 75
3.2.2 Codon usage 76
3.3 Results and discussion 77
3.4 Conclusions 91

4.0 Predicted function of the vaccinia virus G5R protein....93
4.1 Introduction....93
4.2 Methods....97
4.3 Results and discussion....98
4.3.1 HHsearch results....98
4.3.2 Structural modeling of the VACV G5R protein....100
4.4 Conclusions....110
5.0 VACV-G8R: A comparison to human proliferating cell nuclear antigen...112
5.1 Introduction....112
5.2 Methods....115
5.3 Results and discussion....116
5.3.1 Secondary and tertiary structure conservation....117
5.3.2 Functional domain comparison....123
5.4 Conclusions....127
6.0 Discussion....130
Bibliography....136

List of Tables

Table 1. List of incorrectly classified VACV-COP ORFs.....64
Table 2. List of potentially incorrectly classified AMEV ORFs. 67
Table 3. List of 6 AMEV ORFs classified as minor that do not fit the definition of a minor ORF.....67
Table 4. Mean purine to pyrimidine ratios for each codon position of vaccinia virus Copenhagen major and minor ORFs. 69
Table 5. Description of genes in regions 1 and 2.....82
Table 6. Codon usage differences between regions 1 and 2 and 49 MOCV-1 genes conserved in all poxviruses.....88
Table 7. Codon usage differences between regions 1 and 2 and 50 human genes.....89
Table 8. Codon usage differences between 49 MOCV-1 genes that are conserved in all poxviruses and 50 human genes.....90
Table 9. Number of positively charged amino acids in the helical regions of various PCNA proteins.....126

List of Figures

Figure 1. The poxvirus promoter. 25
Figure 2. Structure of six intermediate promoters. 28
Figure 3. Correlation between transcription direction and purine content on the mRNA synonymous strand....54
Figure 4. Correlation between the purine skew and the direction of transcription of the VACV-COP genome, excluding the non-coding terminal inverted repeats. 60
Figure 5. Results of the "quality" measure for VACV-COP. 62
Figure 6. Results of the "quality" measure for *Amsacta moorei* virus (AMEV)....65
Figure 7. An example of a typical dotplot. 74
Figure 8. Dotplot depicting a comparison of the molluscum contagiosum virus genome to itself....79
Figure 9. G+C composition plots created using viral genome organizer (VGO) (Upton *et al.*, 2000). 81
Figure 10. Dotplot comparing molluscum contagiosum virus genome to a random sequence of different nucleotide content. 86
Figure 11. Optimal double-flap DNA substrate used by human FEN-1. 95
Figure 12. The five important regions of the FEN-1 protein. 96
Figure 13. Multiple alignment between vaccinia virus G5R, human FEN-1 and *A. fulgidus* protein sequences. 100
Figure 14. Tertiary structure comparisons between human FEN-1 and vaccinia G5R....102
Figure 15. The hydrophobic wedge region of the G5R structural model. 106
Figure 16. Comparison between two alignments of the same upstream DNA binding region....108
Figure 17. Comparative alignment between VACV G8R, human and yeast PCNA....119
Figure 18. Structures of the VACV G8R, human and yeast PCNA proteins....120
Figure 19. Three major differences between the VACV G8R and human PCNA structures. 122
Figure 20. Complete trimer of the VACV G8R and human PCNA protein structures....123
Figure 21. Electrostatic surface diagrams of the VACV G8R protein and 3 other PCNA proteins. 127

List of Abbreviations

aa	Amino acid
AG	Adenine + Guanine
AMEV	Amsacta moorei entomopoxvirus
APBS	Adaptive Poisson-Boltzmann solver
AT-content	Adenine + Thymine content
ATI	A-type inclusion
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp	Base pairs
CEV	Cell-associated enveloped virus
CPU	Central processing unit
C-terminus	Carboxyl terminus
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded deoxyribonucleic acid
EEV	Extracellular enveloped virus
ER	Endoplasmic reticulum
GC-content	Guanine + Cytosine content
H3TH	Helix-3 Turn-Helix
hFEN-1	Human flap endonuclease
HLH	Helix-Loop-Helix
HMM	Hidden Markov Model
IEV	Intracellular enveloped virus
IFN	Interferon
IFN-γ	Interferon-gamma
IFN-α/β	Interferon-alpha/beta
IL	Interleukin
IMV	Intracellular mature virus
ITRs	Inverted terminal repeats
MOCV-1	Molluscum contagiosum virus
mRNA	Messenger ribonucleic acid
nt	Nucleotide

N-terminus	Amino terminus
ORF	Open reading frame
PAI	Pathogenicity island
PCNA	Proliferating cell nuclear antigen
PDB	Protein data bank
PKR	Protein kinase
PSI-BLAST	Position specific iterative basic local alignment search tool
R	Purine
rER	Rough endoplasmic reticulum
RFC	Replication factor C
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RSCU	Relative synonymous codon usage
SCOP	Structural classification of proteins
SCR	Short consensus repeat
Ser	Serine
SFV	Shope fibroma virus
SPI	Serine protease inhibitor
ssDNA	Single-stranded deoxyribonucleic acid
SUMO	Small ubiquitin-related modifier
TBP	TATA box binding protein
TMP	Thymidine monophosphate
TNF	Tumour necrosis factor
UDG	Uracil DNA glycosylase
VACV	Vaccinia virus
VETF	Viral early transcription factor
VITF	Viral intermediate transcription factor
VLTF	Viral late transcription factor
VOCs	Virus Orthologous Clusters
WHO	World Health Organization
XPG	Xeroderma pigmentosum
Y	Pyrimidine

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Chris Upton, for giving me the freedom to grow as a scientist and for never allowing me to give up. Your constant support –from not saying "no" when I started teaching, to allowing me to try wet-lab work, to all of the amazing opportunities you gave me to stretch my mind at scientific conferences across Canada –will never be forgotten.

To the members of my supervisory committee, Drs. Caren Helbing, Fran Nano and Ben Koop, thank you for supporting and encouraging me for the last five years. Thank you to Dr. Caren Helbing and Dr. Ed Ishiguro for all of the letters of support you wrote on my behalf for various funding applications, the time you both took to support me did not go unappreciated.

To past and present members of the Upton lab, especially Angelika Ehlers, Gord Brown, and David Esteban, thank you so much for your support. To Angelika and Gord, I'll always cherish our daily lunch hours together, you made my time in the lab that much more enjoyable. To the members of the Misra lab, especially Mariana Vetrici and Teresa Francescutti, thank you both so much for our meaningful chat sessions, and for all of your advice during my brief stint at the wet-lab bench. Biochemistry seminars will not be the same without you Teresa. Thank you also to John Hall, Deb Penner, Melinda Powell and Sandra Boudewyn, the support staff in the Biochemistry office, for making all of the administrative tasks that go along with writing a dissertation and being a graduate student in general such a breeze.

Finally, to my entire family who has been behind me since the day I decided to attend UVic as an undergraduate, words cannot express how much your support has meant to me. To my mom who taught me at an early age that you can only eat an elephant one piece at a time which has been my mantra throughout grad school, to my dad for never wavering in his support and to my sisters who were both always there to listen even if it meant having to put up with deep science talk, I thank you all. To my second set of parents, my parents-in-law, Ted and Lou Stroomer, thank you so much for your love and support. And to my loving husband, Chad, who has been there for me since I was in my first year at UVic, your constant reassurance that life would not end if I did not get the next grant or if I messed up during a talk is what kept me going these last 5 years. I could not have made it this far without you!

Dedication

To my four beautiful nieces –Payton, Maddyn, Kieryn and Karissa –may you endeavour to achieve all that you dream.

1.0 Introduction

1.1 Poxvirus classification

The family *Poxviridae* comprises a set of complex, double-stranded DNA viruses that replicate in the cytoplasm of the host cell and can be divided into two subfamilies: *Chordopoxvirinae* (infecting vertebrates) and *Entomopoxvirinae* (infecting insects) (Moss, 2001). The *Chordopoxvirinae* subfamily is further subdivided into eight genera (*Avipoxvirus*, *Capripoxvirus*, *Leporipoxvirus*, *Molluscipoxvirus*, *Orthopoxvirus*, *Parapoxvirus*, *Suipoxvirus*, and *Yatapoxvirus),* the *Entomopoxvirinae* into 3 (*Entomopoxvirus* A through C), with viruses belonging to each genus being related in both virion morphology and host range (Moss, 2001). The most studied poxviruses belong to the *Orthopoxvirus* genus, which includes poxviruses such as cowpox virus, ectromelia virus (mousepox), vaccinia virus (the prototypical poxvirus) and variola virus (the causative agent of smallpox) (Moss, 2001).

With the advent of sequencing techniques in the late 1970's that allowed Sanger to sequence the complete bacteriophage phi X174 genome, only 5375 nucleotides (Sanger *et al.*, 1978), came the eventual sequencing of the first complete poxvirus genome (vaccinia virus strain Copenhagen; VACV-COP) in 1990 (Goebel *et al.*, 1990a). In 1993, an isolate of variola major virus was sequenced (Shchelkunov *et al.*, 1994, Shchelkunov *et al.*, 1996) which permitted the comparison between variola and vaccinia virus (the virus used to vaccinate against smallpox) using bioinformatics techniques (Shchelkunov *et al.*, 1993). Recent advances in DNA sequencing techniques and improvements to bioinformatics approaches has facilitated the analysis of greater than 100 poxvirus genomes and has allowed for the computational comparison of variola

virus to several other orthopoxviruses and most recently, the comparison of 45 variola virus isolates to each other in attempts to pinpoint the origins of the virus and characterizing virulence determinants (Esposito *et al.*, 2006). Perhaps due to the relative simplicity of the poxvirus genome (i.e. no transcript splicing and no overlapping genes), the use of bioinformatics to analyse these genomes has been extensive, and continues to prove a valid technique in the analysis of the poxvirus proteome. The research presented in this dissertation uses bioinformatics to analyse poxvirus genomes, so it is fitting that although the remainder of this introduction will focus on the overall infection and life cycle of poxviruses, it will, wherever possible, highlight the bioinformatics approaches and evidence that was used to contribute to our overall understanding of poxviruses.

1.2 History of variola virus (smallpox)

The first known cases of smallpox infection were thought to have occurred in ancient Egypt, India and China (Fenner *et al.*, 1988, Behbehani, 1983). One of the first clues that smallpox occurred during these times was the unearthing of the mummy of Pharaoh Ramses V in 1898 (Behbehani, 1983, Fenner *et al.*, 1988). Ramses V ruled Egypt for no more than 4 years and died of an "acute" illness in 1157 B.C. (Fenner *et al.*, 1988, Radetsky, 1999). Further examination of his body revealed yellow coloured pustules that covered his face, neck, shoulders and arms and it was later speculated that these pustules were caused by variola virus infection (Radetsky, 1999, Fenner *et al.*, 1988). In China, the first documented cases of smallpox infection occurred in 1122 B.C. where the first attempts to protect individuals by intentional infection with less virulent variola virus

isolates, known as variolation, took place (Fenner *et al.*, 1988, Gross and Sepkowitz, 1998).

Smallpox reached Central America in the sixteenth century, brought by ships carrying slaves from Africa (Fenner *et al.*, 1988, Behbehani, 1983). From Central America, smallpox spread to the native Mexican population and it is thought that more than 3.5 million Aztecs succumbed to the disease after a Spanish Conquistador introduced it through an infected slave he acquired after his conquests in Cuba (Behbehani, 1983). Whole tribes in Brazil, as well as the entire Peruvian population, were also decimated by the introduction of smallpox to South America in the sixteenth century (Behbehani, 1983).

Throughout Europe in the seventeenth and eighteenth centuries, epidemic smallpox infections occurred with a mortality rate of 1 in 10 people in France and in London, England (Behbehani, 1983). In fact, Queen Mary II became infected and later died of smallpox in 1694 and five other monarchs also died of smallpox infection in the eighteenth century (Fenner *et al.*, 1988, Behbehani, 1983). Smallpox epidemics were also frequent in the United States during the eighteenth century (Radetsky, 1999, Behbehani, 1983).

Efforts to prevent smallpox infection are thought to have started with the Chinese in 1122 B.C. who nasally inoculated susceptible patients with variola acquired from the pustules of infected individuals; this method only had minimal success (Gross and Sepkowitz, 1998). Variolation in different forms was performed throughout the world in the seventeenth century but it wasn't until Lady Mary Montagu, who had been horribly disfigured by a smallpox infection years earlier, had her children inoculated with variola in the early eighteenth century (1721) to protect them from a smallpox epidemic that the practice began

to be tested for efficacy (Radetsky, 1999). In order to prove the efficacy of variolation, doctors in London inoculated six prisoners who were condemned to death, and subsequently released them into areas where smallpox epidemics were occurring (Radetsky, 1999). One of these prisoners slept for six weeks in the same bed as a 10-year-old boy who had smallpox and did not catch the disease (Radetsky, 1999). Despite proving the efficacy of variolation in Britain during this time, deaths due to the inoculation still occurred frequently enough for the practice to be only mildly accepted by the common people (Gross and Sepkowitz, 1998).

In the United States during the eighteenth century, a Reverend named Cotton Mather began promoting the use of variolation to prevent smallpox disease (Gross and Sepkowitz, 1998, Radetsky, 1999). Mather was quite successful in decreasing the mortality rate of a smallpox epidemic occurring in Boston at the time, with records showing a mortality rate of 1 in 47 inoculated individuals compared to a 1 in 6 mortality rate in non-variolated people (Radetsky, 1999, Gross and Sepkowitz, 1998). Upon learning the results of Mather's inoculations, doctors in England began to variolate more susceptible individuals, mainly rich people who could afford to be isolated in a hospital, but attempts at large scale smallpox protection did not catch on until Edward Jenner began his experiments of inoculating individuals with cowpox in 1789 (Gross and Sepkowitz, 1998, Radetsky, 1999).

The notion that cowpox could be used to protect against smallpox infection was suggested to Jenner in 1770 after speaking with a milkmaid who explained that she could not catch smallpox since she had already been sick with cowpox (Radetsky, 1999). Jenner subsequently began experiments to test this hypothesis

first by inoculating his son with virus taken from a nurse who had contracted cowpox (Radetsky, 1999). Observing only mild disease in his son and thus assuming his son was now protected from smallpox, Jenner then inoculated him with variola virus and noted that his son did not contract the disease (Radetsky, 1999). Jenner continued to test his hypothesis over the next six years coming to the conclusion that using fresh inoculums directly from cowpox infected cattle was less effective at protecting from smallpox infection than using inoculums of cowpox spread between inoculated humans (Radetsky, 1999). It was Jenner and his continued work with cowpox that eventually lead to the term vaccination being used to describe the administration of antigen in order to protect from disease (Radetsky, 1999).

In only 10 years, the practice of vaccination was adopted worldwide and with most countries mandating required vaccination of the entire population, the incidences of smallpox infection progressively decreased during the remainder of the nineteenth century (Radetsky, 1999). Smallpox infections were nearly nonexistent by the mid-twentieth century in developed nations, and in 1967 the World Health Organization (WHO) put forth an initiative to globally eradicate smallpox (Gross and Sepkowitz, 1998). Smallpox was officially declared eradicated two years after the last naturally occurring case of smallpox in 1979 (Gross and Sepkowitz, 1998).

Global eradication would not have been made possible had it not been for two important discoveries. The first was the development of a stable vaccine that could easily be shipped to developing nations and could be stored for years before use (Radetsky, 1999, Behbehani, 1983). The second was the development of the bifurcated needle to administer the vaccine, which not only decreased the

amount of vaccine that needed to be administered to the patient, but offered portability, consistency in administration and a cost effective way of delivering the vaccine (Radetsky, 1999). Prior to the development of the bifurcated needle, administering the vaccine relied upon the scratch method where the vaccinator would scratch the patient's skin, potentially leaving deep scars, and administer four times more vaccine than the bifurcated needle (Radetsky, 1999). The bifurcated needle, shaped like a horseshoe with two needles at the ends, allowed vaccinators to dip the needles into the vaccine picking up one fourth the amount of vaccine as the scratch method and apply approximately 15 punctures to the skin (Radetsky, 1999). Its portability and relatively low cost were the keys to ensuring success at globally eradicating smallpox (Radetsky, 1999). Other factors that contributed to global eradication included the fact that humans were the only known reservoir for variola virus and that infection was sustained in densely populated locations (Belongia and Naleway, 2005).

There remains, however, to this day, a mystery regarding the origins of the vaccine used to eradicate smallpox. The exact virus that Jenner used to vaccinate individuals in the eighteenth century may have originated from horses in the form of horsepox virus (Radetsky, 1999). Since it was commonly accepted during that time that horsepox could infect both horses and cows, what Jenner thought was cowpox infecting cattle could have been horsepox infecting cattle (Radetsky, 1999). Unfortunately, comparisons with the horsepox virus seen in the eighteenth century with the vaccine strain used during eradication efforts of the twentieth century cannot be made since horsepox had essentially become extinct at the end of the nineteenth century (Radetsky, 1999). The recent sequencing and phylogenetic analysis of a horsepox virus genome that was

isolated in an infected horse in Mongolia in 1976, revealed that in certain regions of the genome, it was similar to other vaccinia virus strains but in other regions, it was similar to cowpox virus (Tulman *et al.*, 2006). Although the exact origins of this virus cannot be pinpointed without the sequences of horsepox virus derived from the eighteenth century when Jenner first began vaccination, the authors speculate that this 1976 strain of horsepox may have been in the process of evolving into a strain similar to that of vaccinia virus but had not yet reached that stage given its similarities in certain regions to both cowpox and vaccinia virus (Tulman *et al.*, 2006). It is also possible that human to horse spread of vaccinia virus could have taken place given that vaccination efforts were still taking place in 1976 (Tulman *et al.*, 2006). This is not entirely unprecedented considering recently published reports of a vaccinia-like virus infecting individuals in Brasil, named Cantagalo virus, that may have been derived from a vaccinia virus strain used in Brasil during smallpox eradication efforts (Damaso *et al.*, 2000, Nagasse-Sugahara *et al.*, 2004). It is thought that this vaccinia infection persisted by infecting local animal populations where it accrued many mutations and led to the recent infection seen in farm workers (Damaso *et al.*, 2000, Nagasse-Sugahara *et al.*, 2004).

1.3 Variola virus disease progression

Since the eradication of smallpox over 25 years ago, there have been no new outbreaks and as such, the majority of original sources were published in the 1980's. This section will summarize the information found in the defining source of the time, a book entitled *Smallpox and its Eradication* written by Frank Fenner and published by the World Health Organization in 1988 (Fenner *et al.*, 1988),

and where other sources were cited, they are referenced within the text as appropriate.

1.3.1 Strains of variola virus

There are two primary variants of variola virus that are capable of causing a smallpox infection that can be distinguished by observing clinical symptoms in infected individuals or through the use of biochemical and bioinformatics techniques (Moore *et al.*, 2006, Cowley and Greenaway, 1990). The first variant, officially named variola minor virus but also known as variola alastrim, has a 1% mortality rate in unvaccinated individuals and is associated with a milder form of the disease. The second variola variant, variola major virus, is associated with the "classical" or "ordinary" smallpox disease and has a mortality rate of up to 30% in unvaccinated individuals (Moore *et al.*, 2006). Variola major can be further subdivided into five clinical types of disease, which cause different clinical manifestations in infected individuals and include, "ordinary" smallpox, "modified" smallpox, variola sine eruptione, hemorrhagic smallpox and flat smallpox. The following section (Section 1.3.2) will describe the progression of an "ordinary"-type smallpox infection.

1.3.2 "Ordinary"-type smallpox disease progression

The normal course of disease progression of "ordinary" smallpox occurs in several stages. The incubation stage marks the start of infection for a person afflicted with smallpox and begins by initial exposure to the virus, usually through inhaled droplets. The virus can also gain entry to the body via pre-existing cuts or lesions contacting virus filled pustules on an infected individual; such infections tend to have a shorter incubation period than respiratory acquired smallpox. A third, less common way that the virus could gain entry

into the body was through the eyes or conjunctiva. The incubation stage can be as short as 7 days and as long as 19 days with an average length of 12 days; this stage is asymptomatic and the patient is not contagious (Moore *et al.*, 2006). A short-lived viraemia follows and coincides with the onset of the prodromal stage of infection (Moore *et al.*, 2006).

The prodromal stage or pre-eruptive stage of infection is marked by a high fever that lasts for 2 to 4 days. Other symptoms experienced during this stage include feelings of general malaise, headache, backache and vomiting. Following this stage is the eruptive stage, which begins with the onset of lesions within the mouth and on the tongue of the patient (known as enanthem) and progresses to a rash involving lesions called macules on the face and particularly the forehead (Moore *et al.*, 2006, Lofquist *et al.*, 2003). The appearance of the rash on the patient's body marks the time when viral load in the respiratory tract is at its highest and thus the patient is the most contagious (Moore *et al.*, 2006). The macules on the patients face spread to the shoulders and upper legs, the trunk and finally the forearms, hands and feet within 24 hours after the initial rash is observed. The macules progress into papules (slightly raised) within 2 days of developing the initial rash and then into vesicles 2 days after the papules are formed (Lofquist *et al.*, 2003). Although the papules were only minimally raised off the surface of the skin, they could be rolled between two fingers and felt as though there was a foreign body embedded in the skin. Approximately 3 days after the progression of papules to vesicles, the vesicles progress into pustules and by the fourteenth day after rash development, the pustules scab over and the brunt of the infection has been weathered (Lofquist *et al.*, 2003).

Two of the main defining characteristics of smallpox disease that can be used to distinguish it from chickenpox (varicella zoster virus), a herpes virus that causes a rash often confused with smallpox, are the presence of lesions on the palms and soles and the localization of lesions to the extremities and only mildly on the trunk in smallpox infected individuals (Lofquist *et al.*, 2003). Since the scabs can be infectious, an infected individual was not declared free of the disease until all of the scabs had fallen off the body, a process that could take up to 27 days after the initial onset of symptoms (Moore *et al.*, 2006). As the scabs fall off, the skin is left with the characteristic scars or "pocks" (Lofquist *et al.*, 2003). Patients who succumbed to smallpox often did so in the second week of infection and it is thought that death was due to toxaemia and hypotension caused by the circulation of immune complexes and viral antigens, despite the fact that the patients organs were heavily infiltrated with virus (Henderson *et al.*, 1999).

1.3.3 Disease progression of variola major clinical types

The disease progression described in section 1.3.2 focuses on symptoms associated with "ordinary"-type smallpox infection. This clinical type of smallpox can be further subdivided into three subtypes: *confluent* which manifested as confluent lesions on the face and arms and had a mortality rate of 62% in unvaccinated individuals; *semi-confluent* where lesions only appeared confluent on the face and had a mortality rate of 27% in the unvaccinated; and *discrete* which involved lesions that were separated by regions of normal skin and had a mortality rate of 9% in the unvaccinated. The remainder of this section will focus on the 4 other clinical types of variola major infection: modified, variola sine eruptione, flat and hemorrhagic.

The modified variola major infection had a more rapid disease progression than ordinary-type infection, with resolution of scabs seen only 10 days after rash formation. Variola sine eruptione, Latin for variola without lesions, is characterized by the sudden onset of fever, headache and backache that lasts for only 2 days with no appearance of lesions.

Although both flat-type and hemorrhagic clinical types rarely occurred, when they did, their outcomes were almost always fatal. Flat-type smallpox has a pre-eruptive stage that is identical in symptoms to that of ordinary smallpox however, the symptoms of backache and headache are much more severe with flat-type smallpox and last beyond the end of the pre-eruptive stage. The lesions in patients with flat-type smallpox were always flat and level with the skin surface and looked as though they were embedded into the skin. Most of the lesions also exhibited signs of haemorrhage at the base and rarely filled with pus, suggesting either a lack of an innate immune response in the infected individual or the ability of this strain of variola to block the innate immune response. Patients with flat-type smallpox often suffered from complications of the disease including pneumonia, stomach dilation and widespread sloughing of lesions.

Hemorrhagic smallpox can be subdivided into two variations of the disease, early and late hemorrhagic smallpox. Early hemorrhagic smallpox acted rapidly in infected individuals, with fever and death occurring within 6 days of becoming infected, leaving no time for the characteristic rash or lesions of "ordinary"-type smallpox to develop. As the name would suggest, infected individuals bled from the eyes, the gums, and the skin and succumbed not from haemorrhage but from heart failure and pulmonary oedema. Late hemorrhagic smallpox causes haemorrhage late in the course of infection after the onset of rash. In contrast to

early hemorrhagic smallpox where lesions don't have time to form, haemorrhaging during late hemorrhagic smallpox is observed at both the site of the lesion and at the mucous membranes with death occurring within 10 days after the onset of rash.

1.3.4 Complications relating to variola infection

Given the extent of infection in individuals infected with ordinary-type smallpox, it's no surprise that a variety of complications related to the virus occurred. The skin of patients infected with smallpox was prone to secondary bacterial infections, as well, complications affecting the central nervous system, the joints, bones and the eyes were also observed. Arthritis affected almost 2% of the people infected with ordinary-type variola major and during the scabbing stage of infection, individuals could experience a complication that caused the bones of the elbow to bow out and become permanently misshapen.

Secondary bacterial infections were observed in the skin of infected individuals as well as the patient's lungs where bronchopneumonia and in some cases pulmonary oedema was seen. Long-term effects of the skin due to smallpox infection included massive scarring, hypopigmentation in darker skinned individuals and hyperpigmentation in fairer skinned individuals, and deep pits at the site of each lesion, all of which occurred in 65% of infected individuals.

1.4 Vaccination and its complications

As was mentioned in Section 1.2, the origins of Jenner's vaccine strain remain a mystery. During the eradication of smallpox in the twentieth century, propagation of the vaccine involved infecting the skin of cows with vaccinia virus and collecting the vaccinia containing scabs so that they could be homogenized and eventually used as a vaccine (Moore *et al.*, 2006). Harvested scabs were

often referred to as vaccine pulp and contained contaminants such as hair, dead skin cells, bacteria, immune cells and plasma (Fenner *et al.*, 1988). Following homogenization and clarification, the remaining liquid, which was usually still contaminated with bacteria, was referred to as vaccine lymph (Fenner *et al.*, 1988). The vaccine lymph was added to glycerol and phenol to eliminate any bacteria and to help improve the stability of the vaccine (Fenner *et al.*, 1988). In 1948, Leslie Collier developed a method to freeze-dry vaccine lymph which vastly improved vaccine stability at higher temperatures and offered an alternative to the liquid vaccine in tropical climates (Fenner *et al.*, 1988).

It is important to note that although it is commonly thought that Jenner's vaccine consisted of either cowpox or horsepox, the vaccine that was used during eradication and is currently used today consists of vaccinia virus (Fenner *et al.*, 1988). In order to explain the origins of vaccinia virus, it was originally speculated that vaccinia virus was a hybrid between cowpox and variola virus (Fenner *et al.*, 1988). The sequencing of cowpox, variola and vaccinia virus and the subsequent computational comparison of these three genomes revealed that despite all three genomes having over 90% similarity, there still exists unique regions in the cowpox genome not seen in vaccinia or variola, and deletions of genes in vaccinia virus compared to variola virus which suggest that not only is vaccinia a distinct poxvirus species that belongs to the orthopoxvirus genus but that it evolved from a common ancestor to the cowpox and variola viruses (Shchelkunov *et al.*, 1993, Shchelkunov *et al.*, 1998). These types of comparisons are also further complicated by the fact that there are three cowpox virus strains, and one has several differences compared to the other two.

Although there were several different strains of vaccinia virus that were used during eradication efforts, the licensed vaccine of the United States was known as the New York Board of Health strain even though it consisted of a mixture of several different individual vaccine strains (Belongia and Naleway, 2005). Given the recent potential for a bioterrorist attack involving smallpox, the US government has contracted the Acambis Corporation to manufacture a new clonal vaccine, consisting of just one virus strain rather than a mixture, that eliminates the need for propagation in cattle (Artenstein *et al.*, 2005). The vaccine strain currently in clinical trials, named Acam2000, has been shown to be equally as effective as the New York Board of Health vaccine although one case of myopericarditis was observed in a patient receiving Acam2000, prompting further study into the effects of the vaccine on the heart (Artenstein *et al.*, 2005).

Severe, albeit rare (1 in approximately 26,000 people vaccinated in the US in 1968), complications in immunocompromised individuals are associated with smallpox vaccination (Belongia and Naleway, 2005, Fenner *et al.*, 1988). In people who are not immunocompromised, the most severe complication, again very rare, is post-vaccinial encephalitis that causes paralysis and the inability to speak and results in death in 35% of patients (Fenner *et al.*, 1988, Belongia and Naleway, 2005). Skin complications at the site of vaccination are common and can be subdivided into three groups: eczema vaccinatum, progressive vaccinia and generalized vaccinia (Belongia and Naleway, 2005, Fenner *et al.*, 1988). Progressive vaccinia occurs in immunocompromised individuals and is characterized by an inability of the lesion at the site of vaccination to heal, as well as a spread of lesions to other parts of the body, causing death within months of vaccination (Fenner *et al.*, 1988, Belongia and Naleway, 2005). Generalized

vaccinia occurs in healthy individuals and is characterized by a systemic vaccinia infection, causing rash with a low mortality rate (Belongia and Naleway, 2005, Fenner *et al.*, 1988). Individuals suffering from eczema or atopic dermatitis who receive the vaccine can develop eczema vaccinatum as a complication, which is characterized by the immediate formation of lesions where previous outbreaks of eczema have occurred (Fenner *et al.*, 1988, Belongia and Naleway, 2005). Treatment of eczema vaccinatum involves intravenously administering vaccinia immunoglobulin (Belongia and Naleway, 2005, Fenner *et al.*, 1988).

Accidental infection with vaccinia virus is the most common, non-fatal complication of vaccination and occurs when an unvaccinated individual with cuts or lesions on their skin comes into contact with the lesion of a vaccinated individual (Fenner *et al.*, 1988, Belongia and Naleway, 2005). The most common routes of accidental inoculation are through the eyes, vulva and perineum, through cuts on the skin in these regions (Belongia and Naleway, 2005, Fenner *et al.*, 1988). Two examples of accidental infection have recently been published; the first involved a recently vaccinated US soldier in Alaska who accidentally inoculated a woman's vulva during sexual intercourse (Centers for Disease Control and Prevention (CDC), 2007a) and the second involved a recently vaccinated US soldier in Illinois who accidentally inoculated his son who had a history of severe eczema. The boy contracted severe eczema vaccinatum and recovered after the administration of vaccinia immunoglobulin after a 48-day hospitalization (Centers for Disease Control and Prevention (CDC), 2007b). Given that most military personnel in the United States are currently being vaccinated for smallpox with the potential one day for more people to become vaccinated, it is important for health care providers to stress proper hand-

washing after touching the vaccination site to avoid these types of accidental infections.

1.5 Virion Structure

Poxvirus genomes consist of a single linear double-stranded DNA molecule; they range in size from 130 kbp to 360 kbp and exhibit a range of AT-content, from the extremely AT-rich entomopoxviruses (~80% A+T) to the AT-poor parapoxviruses and molluscum contagiosum virus (~35% A+T) with the orthopoxvirus genomes being somewhat AT-rich (~65% A+T) (Moss, 2001, Bawden *et al.*, 2000, Goebel *et al.*, 1990a, Darai *et al.*, 1986). The ends of the genome are sealed by a hairpin nucleotide loop and are flanked by inverted terminal repeats (ITRs) that are identical in sequence and are oriented in the opposite direction at either end of the genome (Moss, 2001). Poxvirus genes do not overlap except, occasionally for a few nucleotides, and are present on both strands of the genome (Moss, 2001). Genes that are highly conserved amongst most poxviruses and essential for poxvirus replication are located toward the centre of the genome, and the non-essential usually virulence factor genes tend to be located towards the ends of the genome (Moss, 2001, Gubser *et al.*, 2004). One possibility for this observed genome organization could be that since ITRs at the ends of the genomes are AT-rich and would be capable of interacting with poly-(dT) cDNA in the host cell, these regions would be prone to recombination and gene transfer events and thus are more likely to contain genes that provide an advantage to the virus life cycle (perhaps acquired from an external source) but that are non-essential to virus survival in the host (Yao and Evans, 2001). There are 49 orthologous genes that are found in all sequenced genomes and these represent genes that are either known or are expected to be essential and

have previously been labelled as the minimum essential genome of the poxvirus (Upton *et al.*, 2003). Throughout the following three sections (Sections 1.5 through 1.7), the term "open reading frame" (ORF) will be used when little or no experimental evidence exists to indicate that the ORF has a gene product associated with it, and the term "gene" will be used when an ORF has been experimentally shown to be expressed. The following three sections will also focus on the life cycle and virulence factors of vaccinia virus strain Copenhagen (the prototypal poxvirus) and thus will use the naming scheme first described by Goebel, unless otherwise specified (Goebel *et al.*, 1990a).

Although it is known that the viral genome must be compacted in some way to fit into the viral core, no proteins have been shown to be essential in the compaction process (Condit *et al.*, 2006). The compacted genome along with proteins required for early-stage transcription, including RNA polymerase and viral early transcription factors, form the viral core of the virus particle (Moss, 2001). The core membrane surrounds the viral core and contains at least 12 non-glycosylated proteins, some of which are likely involved in making the core membrane appear to be studded with spikes that extend from the surface of the core (Moss, 2006, Fenner *et al.*, 1988, Dubochet *et al.*, 1994). The proteins that are known to make up the outer part of the core membrane are named 4a, 4b and p39 and correspond to genes A10L (Heljasvaara *et al.*, 2001, Rodriguez *et al.*, 2006), A3L (Kato *et al.*, 2004) and A4L (Cudmore *et al.*, 1996) in vaccinia virus, respectively (Pedersen *et al.*, 2000, Moss, 2001). The p25 protein (L4R gene in vaccinia) is found on the inner part of the core and has been classified as a DNA and RNA binding protein, which is fitting given its proximity to the DNA, housed within the core (Moss, 2001, Pedersen *et al.*, 2000, Bayliss and Smith,

1997). The p11 protein (F17R in vaccinia) was also thought to interact with the DNA within the core (Ichihashi *et al.*, 1984) although more recent evidence shows that p11 does not localize with the DNA of the core, rather it localizes in the space around the outer part of the core membrane (Pedersen *et al.*, 2000) and has been shown to be essential to virion formation (Zhang and Moss, 1991, Condit *et al.*, 2006). The p11 protein has also been shown to interact with actin although the exact reason for this interaction remains to be determined (Reckmann *et al.*, 1997). The viral cores take on a characteristic dumbbell shape with two protein aggregates in the shape of ovals known as lateral bodies situated directly above and below the sites of concavity of the cores (Moss, 2006, Pogo and Dales, 1969, Moss, 2001).

An outer membrane that is ribbed with surface tubule proteins surrounds the core and the lateral bodies and altogether they make up the intracellular mature virion (IMV) particle (Moss, 2001, Fenner *et al.*, 1988). Whether the IMV gets wrapped in a single or double membrane and the origins of this membrane remain controversial (Condit *et al.*, 2006), with some researchers believing that the IMV is surrounded by a single membrane that is synthesized *de novo* and others believing that the IMV is wrapped by a double membrane that forms from the endoplasmic reticulum or the trans-Golgi network (Condit *et al.*, 2006, Sodeik and Krijnse-Locker, 2002). Electron microscopy studies tend to favour the former hypothesis (Sodeik and Krijnse-Locker, 2002) although how the membrane forms *de novo* still remains to be investigated. Proteins that make up the outer membrane of an IMV particle include the vaccinia virus A17L (Wallengren *et al.*, 2001), A27L (Vázquez and Esteban, 1999) and A14L (Traktman *et al.*, 2000, Mercer and Traktman, 2003) proteins, which have been shown to

interact with each other (Rodríguez *et al.*, 1997), and the L1R protein (Aldaz-Carroll *et al.*, 2005) (myristilated) which is required for production of completely formed infectious IMV particles (Ravanello and Hruby, 1994). A brick-shaped fully formed IMV particle is infectious and measures approximately 350 nm long and 250 nm wide (Moss, 2001). Two mass spectrometry analyses performed in 2006 showed that there were approximately 65 vaccinia virus proteins comprising the IMV particle, of which, 22 were shown to be membrane-associated although the exact role that some of these identified proteins play in virion morphogenesis and virion entry into the host cell are yet to be determined (Chung *et al.*, 2006, Yoder *et al.*, 2006).

Two lipid membranes, thought to be derived from the trans-golgi network or from early endosomes, enwrap the IMV particle and the resulting particle is then known as an intracellular enveloped virion (IEV) and has a total of 3 lipid membranes wrapped around it (Smith *et al.*, 2002). There are 7 viral proteins embedded in this envelope, all but one (F13L in vaccinia) are glycosylated (Smith *et al.*, 2002). The IEV particle buds from the cell and remains associated with the surface of the host cell at this stage losing 1 of its membranes to have only 2 membranes wrapped around it; the particle is now referred to as a cell-associated enveloped virion (CEV) (Smith *et al.*, 2002). Interestingly, this transition from IEV particle to CEV particle results in the loss of one protein (F12L) from the envelope (Smith *et al.*, 2002). The CEV particles can detach from the host-cell surface and are then referred to as extracellular enveloped virions (EEV) (Smith *et al.*, 2002). The envelope of the EEV particle consists of 5 of the original 7 proteins seen on the surface of the IEV particle, having lost the A36R protein in the transition from a CEV particle to an EEV particle (Smith *et al.*, 2002).

1.6 Virus life-cycle

1.6.1 Entry

Given the number of years that have been spent researching poxviruses, it is somewhat surprising that very little is known about host-cell surface receptors that facilitate viral entry and the actual mechanism by which the virus enters the cell. Typically, enveloped viruses gain entry to the host cell by fusing their lipid envelopes with the host-cell surface or with the membrane of a vesicle formed during endocytosis (Sieczkarski and Whittaker, 2005). In the case of poxviruses, the end result of this stage of the viral life cycle is that the viral core is successfully delivered to the host-cell cytoplasm, free of any surrounding membranes. Several possible mechanisms have been proposed to explain how vaccinia virus enters the host cell, depending on the type of virus particle (IMV or EEV) (Moss, 2006). Since EEV particles are simply IMV particles with an extra lipid membrane, it would seem unlikely that simple fusion with the plasma membrane surrounding the host cell would occur since the end result of this type of fusion would be an IMV particle located in the cytoplasm of the host rather than the viral core. Immunofluorescence data supports an endocytosis entry hypothesis for EEV particles, which involves a two-step process (Vanderplasschen *et al.*, 1998). First, the EEV particle is endocytosed and the low pH environment of the endosome causes the outer EEV membrane to degrade. This degradation results in the formation of an IMV-like particle whose outer membrane fuses with the endosomal membrane causing the release of the viral core into the cytoplasm of the cell (Vanderplasschen *et al.*, 1998).

The entry of IMV particles is better characterized than EEV entry because the latter are exceptionally fragile and very difficult to isolate (Vanderplasschen *et al.*,

1998, Moss, 2001). During the initial stages of infection *in vivo*, entry of EEV particles is more likely to occur since IMV particles are only released when the cell is lysed, an event that occurs later in the infection cycle when the virus has completely taken over the cell (Moss, 2001). In contrast to the low pH requirement seen with the entry of EEV particles, IMV particles do not require a low pH environment to undergo membrane fusion (Vanderplasschen *et al.*, 1998). It remains to be determined, however, whether the IMV particle enters the host-cell via fusion to the plasma membrane or to an endosomal membrane although it is possible that both modes of entry could be used depending on the strain of virus and host-cell type (Moss, 2006).

Before any type of fusion mechanism can occur, the virus must first attach to the host-cell and then fusion proteins must become activated (Moss, 2006). Three IMV proteins (D8L, A27L and H3L) are thought to facilitate virus-host cell attachment through binding of glycosaminoglycans on the surface of the host-cell, although they are not individually essential for virus entry (Hsiao *et al.*, 1999, Hsiao *et al.*, 1998, Chung *et al.*, 1998, Lin *et al.*, 2000, Carter *et al.*, 2005). In 2005, a group of eight IMV associated proteins (A16L, A21L, A28L, G3L, G9R, H2R, H5R and L5R) all of which are present in all poxviruses sequenced to date, were found to play a role in membrane attachment and fusion since knocking each gene out produced IMV particles that were no longer able to penetrate the host cell membrane and were therefore no longer infectious, suggesting that these proteins may make up the viral entry/fusion complex (Moss, 2006, Senkevich *et al.*, 2005).

Although it is thought that host-cell receptors that facilitate viral attachment must be involved in IMV and EEV entry into the host cell, any definitive

receptors have yet to be determined (Moss, 2006, Moss, 2001). As mentioned above, it is thought that IMV entry into the host cell is initiated by the binding of 3 proteins (D8L, A27L and H3L) in the vaccinia IMV particle (Carter *et al.*, 2005). Another possibility is that lipid rafts play a role in virus entry since the removal of cholesterol from the host cell membrane prevented IMV particles from entering the cell (Chung *et al.*, 2005). Regardless of which receptor is used, binding of the IMV particle to the surface of the host cell triggers cell-signalling events in the host that include the phosphorylation of protein kinase C and the formation of actin filaments, both of which help the IMV particle enter the cell (Locker *et al.*, 2000). Interestingly, the triggering of cell-signalling events has not been observed in EEV particles, which could be due to the different surface proteins of EEV and IMV particles (Moss, 2006).

1.6.2 Uncoating

Once the viral core enters the cytoplasm of the host cell, it is transported via microtubules to just outside the nucleus where viral early gene transcription begins inside the viral core (Moss, 2001, Smith *et al.*, 2003). The mRNA exits the core and is translated into early proteins, some of which are required for the release of viral DNA from the core (uncoating) (Mallardo *et al.*, 2002).

1.6.3 Gene expression

Since poxviruses replicate in the cytoplasm and not the nucleus where cellular transcription takes place, they must encode a functioning gene transcription system, including proteins that function as transcription factors, RNA polymerases, and mRNA capping and polyadenylating enzymes (Moss, 2001). Gene expression in poxviruses is temporally regulated and has three stages: early, intermediate and late (Moss, 2001). The viral core is packaged with a

complete set of proteins capable of initiating early-stage gene transcription, the proteins made during the early-stage subsequently are used to transcribe the intermediate stage genes and the proteins resulting from the intermediate stage are used to transcribe the late stage genes (Moss, 2001).

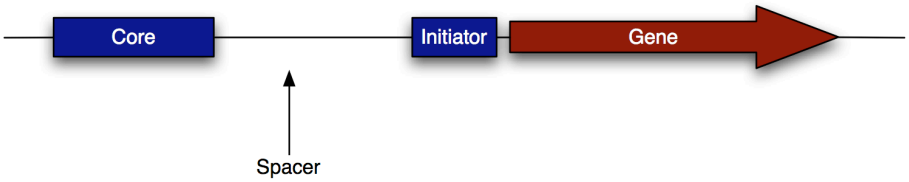
The virally encoded RNA polymerase, which consists of 9 subunits, carries out transcription of all three classes of poxvirus genes (Broyles, 2003). Two of the large subunits and one of the small, 147 kDa, 132 kDa and 7 kDa corresponding to vaccinia virus genes J6R, A24R and G5.5R respectively, show sequence similarity to eukaryotic and prokaryotic RNA polymerase subunits (Davis *et al.*, 2002, Patel and Pickup, 1989, Murakami *et al.*, 2002). Although the crystal structures of these vaccinia proteins have not been solved, the crystal structure of the two large subunits of yeast RNA polymerase II shows that these large subunits form a complex that resembles a claw that is capable of clamping onto the DNA and it is likely, given the high similarity of the vaccinia proteins to yeast RNA polymerase, that the vaccina RNA polymerase large subunits also clamp DNA in this way (Murakami *et al.*, 2002, Patel and Pickup, 1989).

1.6.3.1 Early gene expression

Transcripts of early-stage poxviral genes are made within minutes of infection and reach their maximum abundance approximately 1.5 hours post-infection (Broyles, 2003, Moss, 2001). Cryoelectron tomography of the vaccinia virus core shows the presence of pores in the viral core where early-stage transcripts are likely released into the cytoplasm (Cyrklaff *et al.*, 2005) and it is thought that the transcripts then move away from the core via the microtubule network where they form granular structures that then recruit host translational machinery (Mallardo *et al.*, 2002).

Over half of the genes encoded by a poxvirus genome are thought to be early stage genes (Oda and Joklik, 1967) yet surprisingly, only one poxvirus early promoter has been characterized (Davison and Moss, 1989a). The lack of characterization of other early promoters is likely due to the fact that early mRNA transcripts are synthesized within the viral core and thus any attempts to express a reporter gene under the control of a viral early promoter would fail because the construct could not enter the viral core to be transcribed. In his characterization of early promoters, Davison (Davison and Moss, 1989a) undertook the mammoth task of creating a recombinant virus for each of the mutations of the putative promoter.

All poxvirus promoters, independent of class, consist of two main regions: the core and initiator regions (Figure 1A) (Moss, 2001). The initiator region contains the transcriptional start site, which is labelled as position +1. The core region is always located upstream of the initiator region and is labelled relative to the transcriptional start with a position that is negative (e.g. -30). The sequence separating the core and initiator regions is known as the spacer region and depending on the class of promoter may play an important role in poxvirus gene expression (Moss, 2001).



A

	Core	Initiator
Early	AAAxTxGAAAxxTA	A/G
Intermediate	xxTxxxTTxAAAxAA	TAAA
Late	A/T rich	TAAATG/A

B

Figure 1. The poxvirus promoter.
(A) Diagram of a typical poxvirus promoter with core, spacer and initiator regions indicated. (B) Important nucleotides of the poxvirus promoter. Nucleotides in the core regions of early and intermediate promoters represent experimentally determined consensus sequences with x representing any nucleotide at that position.

The results of the early promoter characterization performed by Davison showed that the core region of the promoter was located between position -12 and -29 nucleotides (nt) upstream of the transcription start site (Moss, 2001, Broyles, 2003, Davison and Moss, 1989a). The nucleotide sequence of the this core region is highly variable but is AT-rich and comparison with other predicted early promoters shows one highly conserved guanine residue at either position -21 or -22 (Figure 1B) (Davison and Moss, 1989a). The initiator site of the poxvirus early promoter contains a key purine residue that is between 12 and 17 nucleotides downstream of the core (Moss, 2001). The spacer region of the early promoter may also play a role in transcription (Davison and Moss, 1989a). Attempting to locate the AT-rich core and initiator regions of all three types of poxvirus promoters within an AT-rich poxvirus genome using bioinformatics techniques is very challenging. However, the promoters of GC-rich poxviruses

(parapoxviruses and molluscum contagiosum virus), which are also AT-rich, can be useful in the building of consensus sequences.

Early-stage transcription involves one early transcription factor (ETF), and the RNA polymerase complex (Broyles, 2003). The ETF is a heterodimer consisting of the products of the D6R (Broyles and Fesler, 1990, Gershon and Moss, 1990) and A7L (Gershon and Moss, 1990) genes which binds to the DNA at the core region of the promoter and approximately 7 nt downstream of the initiator region, leaving the transcription start site available for RNA polymerase binding (Cassetti and Moss, 1996, Broyles, 2003). The ETF then recruits the RNA polymerase complex to the DNA template where it must wait for the ETF to dissociate from the template DNA since it blocks the region 7 nt downstream of the initiation site (Li and Broyles, 1993). This dissociation of the ETF requires the hydrolysis of ATP (Broyles, 1991). The 95 kDa subunit of the RNA polymerase (H4L) (Ahn and Moss, 1992, Kane and Shuman, 1992) is required for docking of the RNA polymerase to the ETF complex and also plays a role in early-stage transcription elongation and termination (Mohamed and Niles, 2001).

The next step in early gene transcription is elongation of the viral transcript. The early-stage elongation complex consists of the RNA polymerase complex, which must contain the 95 kDa subunit (H4L), a capping enzyme encoded by the D1R and D12L vaccinia virus genes, and the NPH I (nucleoside phosphohydrolase I) protein (encoded by the D11L gene in vaccinia) (Broyles, 2003). The NPH I protein has been shown to directly interact with the H4L protein and it is possible that the H4L protein also interacts with other proteins of the elongation complex that have yet to be identified (Mohamed and Niles, 2000).

Termination of early-stage transcription occurs when the RNA polymerase encounters a U_5NU termination sequence (where N is any nucleotide) in the transcribed mRNA (Broyles, 2003). Transcription does not stop immediately after encountering the termination signal, rather it continues for approximately 50 nucleotides (Deng and Shuman, 1997). The proposed model for early-stage termination is as follows. The termination sequence, UUUUUNU, on the growing mRNA transcript comes into contact with the capping enzyme that is associated with the RNA polymerase complex, which in turn triggers the RNA polymerase to stop elongation (Deng and Shuman, 1997). The NPH I protein acts as a motor to slow transcription once the termination signal is encountered, and through the hydrolysis of ATP, causes the RNA polymerase to release itself from the template DNA (Deng and Shuman, 1998, Christen *et al.*, 1998).

1.6.3.2 Intermediate gene expression

Both intermediate and late stage gene expression takes place only after the viral genome has exited the viral core and has been replicated at least once (Broyles, 2003, Moss, 2001). Intermediate gene expression begins approximately 1.5 hours post-infection and begins to decline 2 hours post infection (Moss, 2001). There have been only 6 experimentally identified intermediate genes, although several more have been predicted as being intermediate by repressing a viral late transcription factor and observing expression levels of these putative intermediate genes (Zhang *et al.*, 1992a). Of these 6 intermediate genes, three are late-stage transcription factors (Moss, 2001). The intermediate stage promoter consists of an AT-rich 14 nucleotide core, which is similar to the early promoter core region, a 10-12 nucleotide spacer region, which is important for transcription initiation, and an initiator region that closely resembles the late

promoter and has the sequence TAAA (Figure 1B) (Moss, 2001, Broyles, 2003). To highlight the variability of the core region and to show the conservation of the initiator region of intermediate promoters, a multiple alignment and LOGO consensus diagram of the promoter regions of the 6 intermediate genes is shown in Figure 2. Locating intermediate promoters using bioinformatics is especially difficult because not only does the AT-rich core need to be located amidst an AT-rich genome, but the initiator region must be distinguished from a late-stage initiator region which shares a nearly identical sequence to the intermediate initiator region (Section 1.6.3.3).

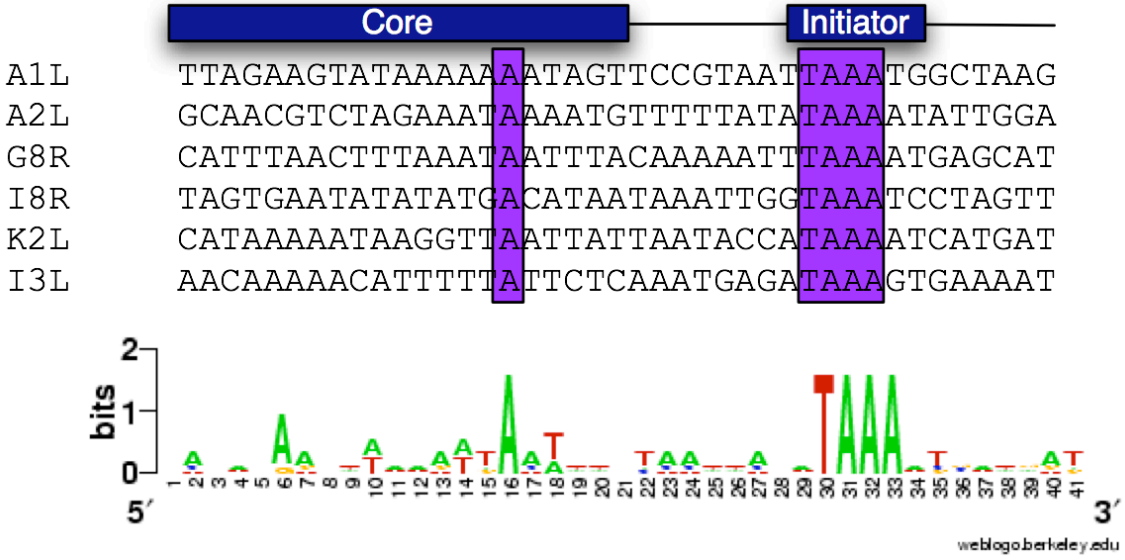


Figure 2. Structure of six intermediate promoters.
Top panel shows an alignment of the 6 intermediate promoters with core and initiator regions marked. Regions highlighted in purple represent nucleotides that are conserved in all promoters. Bottom panel shows a LOGO sequence diagram, visually highlighting the nucleotides that are best conserved over the length of the promoter.

Intermediate transcription begins at one of the three adenine nucleotides in the TAAA sequence although pinpointing the exact adenine nucleotide is difficult due to the slippage that occurs when RNA polymerase attempts to initiate

transcription at poly-A sequences (Broyles, 2003, Moss, 2001). This slippage results in the addition of approximately 30 adenine nucleotides to the 5'-end of the transcript (Bertholet *et al.*, 1987, Schwer *et al.*, 1987). As in early gene transcription, there is evidence that intermediate transcription elongation requires the capping enzyme as well as three intermediate transcription factors (VITF 1-3) (Broyles, 2003) . VITF-1 is actually the 30 kDa subunit of RNA polymerase (encoded by the E4L gene) (Rosales *et al.*, 1994) and VITF-3 is a heterodimer of the products of the A8L and A23R genes (Gershon and Moss, 1990, Hu *et al.*, 1998, Sanz and Moss, 1999). VITF-2 was the first identified host-encoded protein that has been shown to play a role in poxvirus transcription, and it has recently been identified as being a complex of two proteins, G3BP (Ras-GTPase-activating protein) and p137 (cytoplasmic activation/proliferation associate protein) (Katsafanas and Moss, 2004). Although little is known about the function of p137 in transcription, G3BP has been shown to play a role in RNA metabolism in the host cell (Katsafanas and Moss, 2004). Another host-encoded protein that has been shown to bind to the initiator region of intermediate promoters is the YinYang1 protein, which is a host transcription factor (Broyles *et al.*, 1999). Immunofluorescence microscopy has shown that it is recruited to the cytoplasm in poxvirus-infected cells and that it can bind to the TAAATGG initiator region of a known intermediate vaccinia virus gene (Broyles *et al.*, 1999).

The termination of intermediate and late transcription happens using similar mechanisms and as such, this section will cover termination for both stages of transcription. Unlike early stage transcription termination, intermediate and late termination does not utilize a termination sequence and often results in transcripts that are variable in length (Xiang *et al.*, 2000). There are three

proteins that are thought to play a role in termination of intermediate and late transcripts (Broyles, 2003). The first, A18R, has DNA helicase activity and mutants produce much longer transcripts (Simpson and Condit, 1995, Xiang *et al.*, 1998). The A18R protein is also unable to terminate transcription on its own, requiring cell extract from uninfected cells in order to be functional, and thus implicating an as yet unidentified host protein as part of the viral transcription termination process (Lackner and Condit, 2000). The second gene, G2R, produces an opposite phenotype when deleted, making shorter than normal transcripts (Black and Condit, 1996). This observation implies that G2R may play a role in the elongation step of intermediate and late transcription (Broyles, 2003). The third protein, J3R, may function similar to the G2R protein since shorter transcripts are also seen in virus knockouts (Xiang *et al.*, 2000, Latner *et al.*, 2000). Interestingly, the J3R protein was not expected to play a role in transcription termination because it had previously been experimentally determined to be a 2′-O-methyltransferase that functions both independently during the mRNA 5′-capping process, and together as a heterodimer with the poly(A) polymerase protein during the addition of the 3′-poly(A) tail (Latner *et al.*, 2002, Latner *et al.*, 2000, Schnierle *et al.*, 1992).

1.6.3.3 Late gene expression

Transcription of late genes reaches its peak at 4 hours post infection and can continue for 48 hours post infection, which is thought to be a result of the shorter (less than 30 minute) half-life of late transcripts (Moss, 2001). Most of the products of late genes are structural in nature and include the majority of the proteins that comprise the virion, however, some are non-structural (enzymatic)

proteins that are packaged with the virion such as the VETF protein and the H4L gene product as well as other important enzymes (Moss, 2001).

The poxvirus late promoter is similar in nature to the intermediate promoter and consists of an AT-rich core region that is 20 nucleotides long, a short spacer region of only 6 nucleotides and an initiator region that is highly conserved in all late promoters and consists of the TAAAT sequence (Figure 1B) (Moss, 2001, Broyles, 2003, Davison and Moss, 1989b). A purine residue usually follows the initiator sequence and quite often when a guanine residue follows the initiator sequence to form the sequence TAAATG, the ATG is actually the translational start codon (Broyles, 2003). The main difference between late and intermediate promoter regions is the longer spacer region that is observed in intermediate promoters. Deletion of nucleotides in the spacer region of an intermediate promoter and conversely the addition of nucleotides to a late promoter can cause these promoters to switch from intermediate to late or vice versa (Knutson *et al.*, 2006). A model describing the possible role that the spacer plays in switching from intermediate to late-stage transcription is described at the end of this section. As in intermediate transcription initiation, the viral RNA polymerase slips at the poly(A) sequence of the initiator region, causing the 5'-end of the transcript to be polyadenylated with approximately thirty adenine residues, in effect, creating a 5'-untranslated region that would not have existed had transcription started immediately at the translational start codon (Moss, 2001, Broyles, 2003). This purpose of this 5'-untranslated region is not known, although it is possible that it provides stability to the transcript (Broyles, 2003).

Transcription of late genes does not begin until after the viral genome has been replicated once, and requires *de novo* synthesized RNA polymerase in order

to begin (Hooda-Dhingra *et al.*, 1989). Three late transcription factors have been identified that play a role in the initiation of late transcription: A1L, A2L and G8R (Carpenter and DeLange, 1992, Passarelli *et al.*, 1996, Keck *et al.*, 1993b, Keck *et al.*, 1993a, Zhang *et al.*, 1992a, Keck *et al.*, 1990, Hubbs and Wright, 1996). Although the function of these proteins is currently unknown, yeast two hybrid studies have shown that the G8R protein (discussed further in Chapter 5 of this dissertation) interacts with itself and the A1L protein (Dellis *et al.*, 2004). The H5R protein has also been shown to play a stimulatory role in late transcription and since it has been identified as a substrate for the product of the B1R kinase, it is likely that protein phosphorylation may regulate transcription on some level (Cresawn and Condit, 2007, Brown *et al.*, 2000, Kovacs and Moss, 1996). A host-encoded late transcription factor (VLTF-X) that is known to bind tracts of thymine nucleotides and stimulate transcription has also been identified (Gunasinghe *et al.*, 1998). Thymine tracts were shown by Davison in 1989 to act as an upstream element to the late promoter (Davison and Moss, 1989b, Broyles, 2003). Two candidate proteins encoded by the host and identified through mass spectrometry, A2/B1 and RBM3, both of which are known nuclear ribonucleoproteins and can interact with mRNA, have been identified as being VLTF-X (Wright *et al.*, 2001). Since none of the three virally encoded late transcription factors have ever been shown to directly bind DNA, the VLTF-X protein may be an intermediary between the virally encoded transcription factors and DNA (Wright *et al.*, 2001).

Recently, the host-encoded TATA-box binding protein (TBP) has been shown to bind the core region of both intermediate and late promoters, making it a second host-encoded transcription factor that may act as an intermediary

between the promoter DNA and the viral late transcription factors (Knutson *et al.*, 2006). Results showed that TBP was able to interact with the core regions of intermediate and late promoters and when knocked out via RNA silencing decreased transcription significantly (Knutson *et al.*, 2006). Despite being a nuclear protein, TBP has also been shown to be present in the cytoplasm of vaccinia virus infected cells thus further supporting the hypothesis that it is a late transcription factor (Knutson *et al.*, 2006). Since TBP acts on both intermediate and late promoters, a model has been proposed to show how viral transcription could switch from intermediate to late (Knutson *et al.*, 2006). TBP would bind to newly synthesized viral DNA at all intermediate and late promoter core regions. The intermediate transcription factors (made during early-stage protein synthesis) would then come into contact with TBP at both intermediate and late promoter regions, however, since the spacer region of the late promoter is shorter than that of the intermediate promoter, a fully functioning pre-initiation complex could not form at late promoters and could only form at intermediate promoters. The intermediate genes would then be transcribed (some of which are late transcription factors) and this would cause an accumulation of late transcription factors that would come into contact with the TBP bound to both intermediate and late promoters. As the concentration of late transcription factors increased, they would displace the intermediate transcription factors from the pre-initiation complex at both promoter types and due to the short spacer region of the late promoter, transcription would only occur with late transcription factors at late promoters, thus marking a switch between intermediate and late transcription (Knutson *et al.*, 2006).

1.6.4 Replication

Poxvirus DNA replication takes place in distinct cytoplasmic regions, known as viral factories, that are surrounded by a membrane made from the rough endoplasmic reticulum (rER) (Moss, 2001, Schramm and Locker, 2005, Tolonen *et al.*, 2001). Time course electron microscopy shows that this membrane takes shape approximately 45 minutes after replication begins and disintegrates after virion assembly starts (Tolonen *et al.*, 2001). Several viral proteins are thought to facilitate in this wrapping step, two of which have been characterized: the gene products of the E8R (Doglio *et al.*, 2002a) and A40R (Wilcock *et al.*, 1999) vaccinia genes (Schramm and Locker, 2005). The E8R protein has been predicted to have two transmembrane domains although an actual function has yet to be attributed to it (Schramm and Locker, 2005, Doglio *et al.*, 2002a). The N- and C-termini of the protein have been predicted to be located in the inner part of the viral factory and have previously been shown to be able to interact with DNA, thus potentially making this protein able to initiate rER wrapping while it holds the viral genome in place (Tolonen *et al.*, 2001, Schramm and Locker, 2005). At late times of infection, this protein has been shown to be phosphorylated by the F10L kinase (Schramm and Locker, 2005, Doglio *et al.*, 2002a). Phosphorylation is thought to make the protein less able to stabilize the genome and would coincide with the unwrapping of the viral factory in order for the newly synthesized genome to be released and packaged into viral cores for release (Schramm and Locker, 2005). A recent mutational analysis of the E8R protein, however, has shown that the E8R protein while being located at viral factories, plays less of a role in ER-membrane wrapping and more of a role in transcription at the viral core, better fitting its previously determined ability to interact with DNA (Kato *et*

al., 2007). Recently, the A40R protein was shown to be modified at several sites by a SUMO (small ubiquitin-related modifier) group, which in eukaryotic proteins is a method of targeting proteins to the nucleus (Palacios *et al.*, 2005, Johnson, 2004). SUMOylation is used in DNA viruses that replicate in the nucleus as a way of targeting viral proteins back to the nucleus and preventing these proteins from becoming SUMOylated *in vitro* interferes with a wide range of viral functions, most commonly transcription (Wilson and Rangasamy, 2001). In the case of the A40R protein, SUMOylation targets it to the viral factory where it likely helps in rER enwrapping; any further roles in replication have yet to be determined (Palacios *et al.*, 2005).

Orthopoxviruses encode several proteins that function in the deoxyribonucleotide synthesis pathways, most likely to enhance DNA replication in quiescent cells (Moss, 2001). Vaccinia virus encodes a thymidine and thymidylate kinase, ribonucleotide reductase and a dUTPase none of which are essential to virus survival in the host cell (Moss, 2001). Thymidine kinase (J2R) has been shown to function as a tetramer that can bind ATP and magnesium ions catalyzing the conversion of 2'deoxy-thymidine to deoxythymidine monophosphate (dTMP), and is regulated via feedback inhibition by dTDP and dTTP (El Omari *et al.*, 2006, Black and Hruby, 1992). Thymidylate kinase (A48R) catalyzes the second step of the reaction forming deoxythymidine diphosphate (TDP) from TMP (Hughes *et al.*, 1991) and the two-subunit ribonucleotide reductase enzyme, encoded by the F4L and I4L genes in vaccinia virus, catalyzes the formation of deoxyribonucleoside diphosphates from ribonucleoside diphosphates (Howell *et al.*, 1992, Slabaugh *et al.*, 1993). The dUTPase enzyme, encoded by the F2L gene, is responsible for converting dUTP

into dUMP, which is subsequently used in the synthesis of thymidine triphosphate (TTP) (Broyles, 1993).

The replication of the DNA genome begins between 1 and 2 hours post infection and ends after the formation of approximately 10,000 copies of the genome, 5,000 of which get packaged into new virions (Moss, 2001). Although several viral encoded proteins have been implicated in replication, still very little is known about the process of viral replication. One working model that has been proposed to explain how poxvirus replication occurs is described below. It is thought that a nick in the DNA at the hairpin terminus or at both termini is the triggering event for poxvirus replication. The two DNA strands are then displaced allowing replication to take place. The replication process causes the formation of concatemers, which must then be resolved to create two copies of the genome. Sedimentation and labelling studies support the hypothesis that replication begins at a nick at the terminus of the genome, however the site of this nick has yet to be determined (Pogo *et al.*, 1981, Pogo, 1980). Further to this, sedimentation studies performed in 1978 showed that small DNA fragments were linked covalently to RNA primers, which would support the notion of lagging strand DNA synthesis, and most recently, the vaccinia virus D5R protein has been shown to not only play a role in DNA replication, but is also capable of synthesizing RNA primers suggesting that lagging strand DNA replication is a possibility (De Silva *et al.*, 2007, Pogo and O'Shea, 1978).

DNA replication would not be possible without DNA polymerase (E9L), which has high sequence identity to eukaryotic and other viral polymerases and has also been shown to have 3'-exonuclease activity (Challberg and Englund, 1979, Earl *et al.*, 1986). Experiments knocking out the product of the D5R gene in

vaccinia virus, a nucleoside triphosphatase which acts in a nucleic acid independent fashion, resulted in a termination of replication, though what role it plays in replication is yet to be determined (Evans *et al.*, 1995, Evans and Traktman, 1987). Another protein important in DNA replication is the B1R gene product that has been shown to be a protein kinase whose substrate is the H5R protein that is involved in transcription (Kovacs and Moss, 1996) but is not known to play a role in replication (Beaud *et al.*, 1995, Brown *et al.*, 2000, Lin *et al.*, 1992). The B1R protein has also been recently shown to phosphorylate the host BAF (barrier to autointegration factor) protein, which decreases BAF association with DNA and other nuclear components thus playing a role in the disassembly of the nucleus (Nichols *et al.*, 2006). The D4R gene product also plays a role in replication although its initial classification was as a uracil DNA glycosylase (UDG) which was first identified using bioinformatics database searches (Upton *et al.*, 1993, Stuart *et al.*, 1993), and was shown to be responsible for removing uracil residues that have been introduced into DNA (Stuart *et al.*, 1993). More recently, this dual role of replication and repair for the UDG protein has further been solidified as it has been found that it interacts with the A20R protein (see below) acting as a processivity factor for the viral DNA polymerase (Stanitsa *et al.*, 2006). The A20R gene product is another processivity factor of viral DNA polymerase, which has been shown to interact with the UDG, D5R and H5R proteins (Ishii and Moss, 2002). Interestingly, knocking out the vaccinia virus DNA ligase (A50R), results in a normal replication cycle suggesting that a cellular DNA ligase can functionally replace it (Beaud, 1995). The observation that ligase deficient vaccinia virus is more susceptible to UV light irradiation also points to a role of the ligase in DNA repair (Beaud, 1995).

Since the DNA genome of poxviruses contains two hairpin-loop termini, the resolution of these termini after replication is complete becomes a challenge. It has been shown that the replicated DNA genome forms concatemers that must be resolved into individual units through the cleavage of the concatemer junctions that consist of two identical copies of the hairpin loop region of the genome and resemble a Holliday junction (Moss, 2001). Three viral late-stage proteins (H6R, K4L and A22R) have been identified as possibly playing a role in concatemer resolution (Moss, 2001). The first is a sequence specific topoisomerase, encoded by the H6R gene, that cleaves DNA at the sequence [C/T]CCTT and has been shown to be capable of resolving Holliday junctions typically seen in replicative concatemers of vaccinia virus (Shuman, 1991). The protein encoded by the vaccinia virus K4L gene has been shown to play a role in replication and has DNase activity with the ability to nick and also join DNA (Moss, 2001, Eckert *et al.*, 2005). The K4L protein was previously shown through bioinformatics to belong to the phospholipase D superfamily (Cao *et al.*, 1997). The recent results by Eckert showing it has DNA nick and joining activity is actually not contradictory to its classification in the phospholipase D superfamily since this superfamily is defined by a protein motif that is seen in both nucleases as well as phospholipases (Stuckey and Dixon, 1999). The A22R protein, a Holliday junction resolvase that has sequence identity to bacterial resolvases, has been shown to resolve Holliday junctions *in vitro* and is able to resolve concatemers during viral DNA replication (Garcia and Moss, 2001).

Based on the data presented above, the model for DNA replication begins with a nick at either one end or both ends of the genome and results in a free 3'-hydoxyl group that is then available for DNA replication (Moss, 2001).

Replication of the nicked strand continues with the newly replicated DNA strand folding back on itself and driving DNA synthesis in the forward direction (Moss, 2001). Replication continues until concatemer junctions are formed, which occurs when replication proceeds through the hairpin terminus of the genome (Moss, 2001). In their simplest form, concatemers consist of two copies of the genome, but if replication begins again before concatemer junctions are resolved, the concatemers can consist of multiple genomes all of which would need to be resolved (Moss, 2001). Ultimately, the concatemers are resolved into their unit-length linear genomes (Moss, 2001).

1.6.5 Assembly and release

Assembly of the virus particle begins approximately 5 hours post-infection and occurs at the viral factories where DNA replication takes place (Moss, 2001, Schramm and Locker, 2005). At roughly the same time as the rER membrane surrounding the viral factories begins to degrade, crescent shaped membranes are formed at these viral factories (Tolonen *et al.*, 2001, Schramm and Locker, 2005). There is some controversy surrounding these crescent shaped membranes as one proposed model for their formation involves the *de novo* synthesis of the crescent which consists of a single lipid bilayer that is not continuous with any other cellular membrane, two phenomena that have never before been seen in nature (Hollinshead *et al.*, 1999, Dales and Mosbach, 1968, Heuser, 2005, Sodeik and Krijnse-Locker, 2002). Despite there being evidence to support this single-membrane hypothesis, other researchers propose a second, double-membrane hypothesis (Sodeik and Krijnse-Locker, 2002). The double-membrane hypothesis proposes that as the rER surrounding the viral factories dissolves, crescent membranes embedded with necessary viral proteins, form

from the tightly apposed double membrane of the smooth ER (Griffiths *et al.*, 2001a, Griffiths *et al.*, 2001b). As these crescents associate with newly synthesized genomes and begin to mature into nearly formed virus particles, the resulting IMV particle will have a double membrane (Schramm and Locker, 2005).

There are several proteins that may play a role in virion assembly. First, the F10L kinase has been shown to phosphorylate the gene products of the A14L and A17L genes (Betakova *et al.*, 1999). Virus knockouts for the A14L and A17L genes produce small vesicles rather than a continuous membrane, suggesting these two proteins also play a role in virion morphogenesis (Rodríguez *et al.*, 1998, Rodriguez *et al.*, 1995, Traktman *et al.*, 2000, Wolffe *et al.*, 1996). Interestingly, the H5R gene that has already been shown to play a role in RNA transcription has also been shown to play a role in virion morphogenesis (DeMasi and Traktman, 2000). In addition, three proteins (A9L, L1R and H3L) have been implicated in crescent formation and IMV maturation, of which the H3L protein also played a role in facilitating virion entry into the host cell (Yeh *et al.*, 2000, Ravanello and Hruby, 1994, da Fonseca *et al.*, 2000a, da Fonseca *et al.*, 2000b, Lin *et al.*, 2000). During virion assembly, it is thought that the early transcription machinery becomes encapsidated into the viral core because the RNA polymerase, capping enzyme and several other proteins important in early transcription form a complex that associates with the VETF protein which is attached to the newly synthesized and packaged DNA genome (Moss, 2001).

Prior to being exported from the cell, the IMV particles can become occluded in some poxviruses (Moss, 2001). Cowpox, ectromelia and fowlpox virus all encode a full-length, fully functioning protein that forms a dense matrix known as an A-type inclusion body(ATI) (Funahashi *et al.*, 1988a). Camelpox and vaccinia

viruses both encode a truncated version of the A-type inclusion body gene and therefore do not form occluded IMV particles (Meyer and Rziha, 1993). These A-type inclusions are important in IMV stability after infected cells lyse, and can aid in the spread of IMV particles between hosts (Smith *et al.*, 2003).

Once IMV particles are fully formed, they travel with the aid of the A27L membrane protein via the microtubule network from the area of assembly to the cellular membrane (Smith *et al.*, 2003, Schramm and Locker, 2005). The IMV particle becomes wrapped with membrane derived from either the trans-Golgi or the endosomal network and requires the A27L protein as well as the F13L and B5R proteins found in the EEV particle (Smith *et al.*, 2003). The wrapped IMV particles then fuse to the cellular membrane, which pushes the particle outside of the cell and in the process results in the loss of the Golgi- or endosomally-derived membrane (Smith *et al.*, 2003). The majority of released virus particles remain associated with the cell membrane as CEV particles but EEV particles result when an actin-containing microvillus forms with the CEV particle located at the tip (Smith *et al.*, 2003, Schramm and Locker, 2005). This microvillus will only form when the A33R, A34R and A36R proteins are expressed and can only be functional when the A36R protein is phosphorylated (Wolffe *et al.*, 1998, Roper *et al.*, 1998, Sanderson *et al.*, 1998, Wolffe *et al.*, 1997).

1.7 Virus-host interactions

Poxviruses have evolved several mechanisms that increase their survival in the host cell. Some of these mechanisms include inhibiting protein and DNA synthesis, stimulating cell growth, and modulating the host immune response. Bioinformatics has played a key role in the identification of most of the poxvirus proteins responsible for modulating the host immune response. A few that will

be mentioned in section 1.7.3 are the interferon-gamma (IFN-γ) and interferon-alpha/beta (IFN-α/β) binding proteins, the tumour necrosis factor-alpha (TNF-α) binding protein, and the interleukin-1 (IL-1) binding protein. This section will focus on the ways in which vaccinia virus interacts with the host, however, it should be noted that there are several other proteins encoded by other poxviruses and not vaccinia that are important in viral-host interactions in those viruses.

1.7.1 Inhibition of host macromolecular synthesis

Poxviruses are known to shut down host protein, DNA and RNA synthesis in order to further their own macromolecular synthesis. The observation that host protein synthesis could be inhibited prior to the start of viral gene expression pointed to a protein located in the viral particle as a potential host protein synthesis inhibitor (Moss, 2001). Two candidate proteins (F17R and B1R) in the vaccinia virion have been proposed as playing a role in protein synthesis inhibition (Person-Fernandez and Beaud, 1986, Beaud *et al.*, 1994). Within a few hours after infection, the cytoplasm is filled with viral mRNAs and this alone may be enough to switch translation from host proteins to viral ones (Moss, 2001). Recently, the mutT domain of the vaccinia D10R gene product has been shown to be involved in mRNA decapping, which would increase the degradation of both host and viral mRNA, however since viral mRNAs are in great abundance, subsequent viral protein synthesis would likely not be as affected as host protein synthesis (Parrish *et al.*, 2007). Host DNA replication is also inhibited although the mechanism of this inhibition is still unknown (Moss, 2001).

1.7.2 Stimulation of host cell growth and prevention of apoptosis

Orthopoxviruses have been shown to stimulate host cell growth through the synthesis and secretion of a viral growth factor with similarity to the epidermal growth factor protein encoded by the host (Twardzik *et al.*, 1985, da Fonseca *et al.*, 1999). This stimulation of host cell growth results, for some viruses such as fowlpox virus, molluscum contagiosum virus, shope fibroma virus and yaba virus, in the formation of massive clumps of infected cells, resembling tumours, in the skin of infected hosts (Moss, 2001).

Somewhat related to promoting cell growth is the inhibition of apoptosis that occurs in some poxviruses. Apoptosis is a "suicide"-type mechanism employed by infected cells that allows the cells to kill themselves before the virus has a chance to replicate (Everett and McFadden, 2002). The first protein identified in apoptosis inhibition in poxviruses was the product of the E3L gene in vaccinia virus, which had previously been shown to be a double-stranded RNA binding protein (Haga and Bowie, 2005, Lee and Esteban, 1994). The E3L protein blocks the action of the interferon (IFN) induced double-stranded RNA dependent protein kinase (PKR), which in turn blocks the apoptotic effects of IFN (Kibler *et al.*, 1997, Haga and Bowie, 2005). The serine protease inhibitor (SPI-2), encoded by the B13R gene of vaccinia virus strain Western Reserve, blocks apoptosis via two mechanisms (Haga and Bowie, 2005). First, it blocks apoptosis that is induced by the tumour necrosis factor (TNF) protein and second, it blocks the caspase cascade, possibly by binding to caspase-8, that is commonly seen in cells that are apoptotic (Haga and Bowie, 2005, Kettle *et al.*, 1997, Dobbelstein and Shenk, 1996). Another SPI, SPI-1, that is encoded by the B22R gene in vaccinia virus strain Western Reserve, is also thought to block apoptosis but must work

in conjunction with the SPI-2 protein in order to be successful (Haga and Bowie, 2005). It is important to note that the Copenhagen strain of vaccinia virus encodes fragmented versions of these SPI genes that are thought to be non-functional.

1.7.3 Modulation of the host immune response

From the moment a host becomes infected with a poxvirus, it begins to mount an immune response to the virus using interferons, complement, cytokines, chemokines, natural killer cells and eventually cytotoxic T-cells and antibodies (Haga and Bowie, 2005). Poxviruses employ several different immune evasion tactics to prevent these immune molecules and cells from functioning correctly including proteins that block cytokines, interferons and complement (Haga and Bowie, 2005).

The secreted vaccinia C3L protein is able to inactivate the classical and alternative complement pathways by binding to the C4B and C3B complement proteins and inactivating them (McFadden and Murphy, 2000, Mullick *et al.*, 2005, Sahu *et al.*, 1998). This protein is secreted from infected cells and contains a short consensus repeat (SCR), which consists of four tandem copies of a 60 amino acid sequence (McFadden and Murphy, 2000). This SCR is also seen in the protein encoded by the vaccinia virus B5R gene although evidence has yet to be found that implicates it in inactivating complement (Moss, 2001).

Orthopoxviruses encode two different secreted proteins that are able to inactivate both types of interferon (IFN-γ and IFN-α/β) (Moss, 2001). The first is a secreted IFN-γ receptor encoded by the B8R gene in vaccinia virus. This receptor was originally discovered in 1992, in myxoma virus using a global alignment algorithm for mouse and human protein database searches which

detected weak similarity to IFN-γ receptors; the viral protein was subsequently shown to bind rabbit IFN-γ and, in effect, block its activity (Upton *et al.*, 1992, Mossman *et al.*, 1995). The vaccinia virus B19R gene product acts as a secreted IFN-α/β receptor and can bind to IFN-α albeit with lower specificity than the host encoded IFN-α/β receptor (Alcamí *et al.*, 2000, Colamonici *et al.*, 1995). The B19R protein is also a virulence factor since knockout viruses are attenuated (Alcamí *et al.*, 2000). As was mentioned in section 1.7.2, not only does the E3L protein play a role in apoptosis, but it also functions to block IFN activity (Beattie *et al.*, 1995). The vaccinia E3L protein functions in conjunction with the K3L protein, which has also been shown to block IFN by blocking the effects of the PKR protein (Beattie *et al.*, 1995, Davies *et al.*, 1993, Shors *et al.*, 1998).

Several different cytokines and chemokines are blocked by orthopoxviruses during the course of an infection. First, ectromelia virus (mousepox) encodes an interleukin-18 (IL-18) binding protein, which is similar in amino acid sequence to the host IL-18 binding protein as well as other poxvirus IL-18 binding proteins and prevents IL-18 from inducing IFN-γ, and working together with IL-12 to activate natural killer and T-cells (Smith *et al.*, 2000, Calderara *et al.*, 2001). Although this protein in not present in vaccinia virus strain Copenhagen, it is present and apparently functional in vaccinia virus strains Lister and Western Reserve and corresponds to the gene product of the Western Reserve C12L gene (Smith *et al.*, 2000). The Lister strain of vaccinia virus encodes two TNF receptor proteins (A53R and B28R) only one of which is functional (A53R) and can bind TNF preventing it from triggering apoptosis (see section 1.7.2) and blocking the inflammatory response typically seen when TNF is fully functional (Moss, 2001, Haga and Bowie, 2005, Alcamí *et al.*, 1999). The TNF protein was also first

identified using bioinformatics in the genome of shope fibroma virus (SFV) (Smith *et al.*, 1990a). Comparisons between the different vaccinia virus genomes using bioinformatics showed the variability of the TNF genes in different vaccinia strains, with the B28R and A53R genes in strain Copenhagen being significantly mutated to the point that the TNF genes are likely not functional (Alcamí *et al.*, 1999). Further phylogenetic and comparative analyses of the differences between the TNF genes in different vaccinia strains may provide some clues as to the hypothesis of immune-based selection (McFadden and Murphy, 2000). The vaccinia virus strain Western Reserve B15R protein, shown through protein database comparison to be similar to the human IL-1 receptor, is a soluble IL-1 receptor mimic that binds IL-1β blocking associated signal transduction pathways (Smith and Chan, 1991, Moss, 2001). The Copenhagen strain of vaccinia virus also encodes this protein, although it is shortened by 36 amino acids at its N-terminus compared to the Western Reserve protein, and no studies have been performed to determine the effects of this deletion. In uninfected cells, the mature forms of IL-1β and IL-18 are synthesized from cleavage of their "pro" forms by caspase-1 (Haga and Bowie, 2005). In addition to its role in apoptosis (see section 1.7.2), the SPI-2 protein in vaccinia virus blocks caspase-1 and in turn prevents IL-1β and IL-18 activation (Haga and Bowie, 2005). Vaccinia virus also encodes two identical copies of a secreted chemokine binding protein that binds to several different chemokines and helps prevent the chemotaxis of pro-inflammatory immune cells to the site of infection (Mahalingam and Karupiah, 2000).

1.8 Bioinformatics and poxvirus genomes

This introduction has touched upon the many ways that bioinformatics has been used thus far in the analysis of poxvirus genomes. From the genomic comparisons between variola and vaccinia viruses to the protein database searches used to identify several viral immune modulating proteins, bioinformatics has played and continues to play an important role in poxvirus analyses. With the development of modern, more streamlined techniques to perform bioinformatics analyses as well as the development of faster genome sequencing and annotation techniques, we are already seeing an increase in the amount of poxvirus genomic data that is available for analysis. One of the biggest challenges facing bioinformaticians today is the lack of easily accessible and easy to use databases that store the large quantity of genomic data that is currently available. Our lab has taken steps to improve this by creating a database that can be easily queried and houses all sequenced and complete poxvirus genomes (Ehlers *et al.*, 2002). This database allows for a quick comparison of the gene complement (i.e. which genes are present in variola virus but not in vaccinia) and other tools developed in our lab allow researchers to take analyses one step further by providing detailed nucleotide or amino acid level multiple alignments (Brodie *et al.*, 2004b) or by providing a visual genome versus genome comparison and alignment in the form of dotplots (Brodie *et al.*, 2004a). At the analysis level, there remain several challenges with which poxvirus researchers are faced and each of these challenges cannot be investigated without the used of a variety of bioinformatics tools. The following section will briefly introduce each of the tools that were used in this dissertation.

1.8.1 Bioinformatics tools used in this dissertation

1.8.1.1 Dotplots

The comparison of two genomes usually begins with a pairwise alignment of these genomes. Aligning two or more large genomes, however, can be difficult since current algorithms to align large genomes are computationally intense and are not always accurate, requiring manual editing of the alignment after the fact. Dotplots provide a visual method to compare two genomes that is not only less computationally intense than other alignment algorithms, but can be more accurate, giving a visual overview of how the two sequences align (Sonnhammer and Durbin, 1995) (Chapter 3).

1.8.1.2 Multiple alignments

The comparison of multiple DNA or protein sequences using a multiple alignment is the most common method to find conserved regions in two or more sequences. Several different multiple alignment algorithms exist for aligning several sequences, three of which are used in this dissertation and will be described briefly. The first algorithm used in this dissertation is the Basic Local Alignment Search Tool (BLAST), which uses a substitution matrix to score each pair of amino acids or nucleotides that align correctly between two sequences (Altschul *et al.*, 1997). The overall goal with BLAST is to find segments of the two sequences being compared that have equal lengths and a maximal score that cannot be improved by extending the sequence in either direction (Altschul *et al.*, 1997). In order to have two sequences to compare, BLAST relies on the presence of DNA and protein sequences of many different organisms in a central database housed at the National Center for Biotechnology Information (NCBI) and the BLAST algorithm is used to match a query sequence to these

sequences in the database. Depending on what type of sequence is being used as the query (DNA or protein), different BLAST searches can be performed. BLASTn and BLASTp take DNA and protein sequences as queries respectively, and match them to DNA and protein sequences in the database. BLASTx takes a nucleotide sequence as a query and then translates it to a protein sequence and returns hits that are from the protein sequence database. The tBLASTn search takes a protein sequence as a query and searches translations of all nucleotide sequences in the database, and tBLASTx takes a nucleotide sequence as a query, translates it and then returns hits from a translated nucleotide sequence database. Related to BLAST is the Position-Specific Iterative-BLAST (PSI-BLAST), which creates a position-specific matrix from all significant hits of an initial BLASTp search and then uses this matrix to perform a second iteration of BLASTp (Altschul and Koonin, 1998). After the second iteration is performed, a second position-specific matrix is created which includes all of the significant hits found during the second iteration and then with this matrix a third iteration of BLASTp is performed. Iterations continue until no new significant hits are found. Due to its iterative nature, PSI-BLAST is often used as a means of identifying distantly related proteins and can give clues as to the function of your query protein.

The next alignment algorithm used in this dissertation is the ClustalW alignment algorithm, which begins by taking each of the sequences to be aligned and creates a matrix based on the results of each possible pairwise alignment between each of the sequences (Thompson *et al.*, 1994). This matrix is then used to create an unrooted neighbour joining tree, which is subsequently used to create a rooted tree (guide tree) where each branch length is known and each

sequence is given a weight. This rooted tree is then used to progressively create the multiple alignment. Multiple iterations of this algorithm are then performed until the alignment and the guide tree converge, representing the best possible alignment of the sequences of interest.

The other multiple alignment algorithm used in this dissertation works in a slightly different fashion to ClustalW and is called T-Coffee (Notredame *et al.*, 2000). The main difference between T-Coffee and ClustalW is that T-Coffee does not use a guide tree to create the final alignment. T- Coffee begins with the set of sequences to be aligned and creates a library of all possible pairwise alignments of the sequences using both the ClustalW and FASTA alignments (Notredame *et al.*, 2000). Each of the pairwise alignments has a given weight associated with it, which depends on the percent sequence identity between the two sequences in the alignment. Each of the alignments in the library are then extended using a complex heuristic algorithm and these extended alignments make up a second sequence alignment library. This extended sequence alignment library is then used to progressively align the sequences to each other.

1.8.1.3 Profile-based homology searches

In general, profile-based searches rely on comparison of different profiles of a given sequence rather than sequence alone to identify homologs in a database and often yield more accurate results when compared to simple sequence-based approaches like PSI-BLAST (Söding *et al.*, 2005). The profile-based search used in this dissertation is called HHSearch and uses a profile based on a Hidden Markov Model (HMM) to perform the search (Söding *et al.*, 2005). HHSearch begins by taking the query protein sequence and running several iterations of PSI-BLAST in order to create a multiple alignment of the query plus related protein

sequences. This alignment is then turned into an HMM which is then used to search HMMs created from the protein databank (PDB), structural classification of proteins (SCOP) and CATH databases. The HMMs of both the query and of the proteins in each of the databases are created based on secondary structure either from a prediction of secondary structure using PSIPRED or from the tertiary structure of the protein if a crystal structure is in the database. The profile HMM of the query is then compared to the HMM of each sequence or structure in the database and hits are displayed based on the probability that the query sequence is a true match to the hit sequence.

1.9 Dissertation outline

The overall theme of this dissertation is the analysis of poxvirus genomes using bioinformatics. Given that there remain several challenges with which poxvirus researchers are currently faced, each of the remaining chapters of this dissertation aims to address these challenges. Chapter 2 of this dissertation addresses the challenge of predicting genes that are expressed and functional in genomes that have been minimally characterized. The use of dotplots as a means for comparison of complete poxvirus genomes lead to the interesting finding of unusual regions in the background of these dotplots which is further investigated in Chapter 3 of this dissertation. Another challenge that faces poxvirus virologists is the determination of the function of all proteins that are currently still unknown. Through the use of structural bioinformatics and comparative alignments, the function of the unknown G5R protein in vaccinia virus strain Copenhagen has been predicted in Chapter 4 of this dissertation. Chapter 5 of this dissertation is a continuation of the work presented in Chapter 4 that attempts to structurally characterize a late transcription factor that may

interact with the G5R protein and represents a work in progress. Although each of these chapters is related through the theme of genomic analysis, each chapter will be treated as distinct and as such will each have a brief introduction and discussion section, with an overall discussion section tying all 4 chapters of research together, presented in Chapter 6 of this dissertation.

2.0 Using purine skews to predict genes in AT-rich poxviruses

Published in *BMC Genomics* in 2005 (Da Silva and Upton, 2005a).

2.1 Introduction

The overall goal when predicting genes in poxviruses is to not only identify potential open reading frames (ORFs) but to ensure that these open reading frames are expressed as genes and therefore functional in the poxvirus. Predicting which of these ORFs is likely to be expressed is not a trivial task and relies on the use of other methods and criteria beyond the standard minimum ORF length cut-off. Some of these methods, including locating promoter regions upstream of the predicted ORF to confirm its likelihood of expression, can be difficult to employ in an AT-rich genome since, in the case of promoter prediction, the promoters are also AT-rich and difficult to resolve. Often comparison to other well characterized strains in a genus can help to determine which ORFs are likely to be expressed, but in genera where there are no relatives of a given strain, this method cannot be employed. Previously, amino acid composition was used as an additional indicator of which predicted ORFs were likely to be expressed (Upton, 2000). We wanted to add to this method by incorporating the analysis of ORF nucleotide composition because of work done by Szybalski (Szybalski *et al.*, 1966).

In 1966, Szybalski first discovered that the mRNA synonymous strand of DNA contained a predominance of purine-rich clusters (Szybalski *et al.*, 1966); by convention, the top strand of a linear dsDNA molecule is viewed 5′→3′, therefore when transcription of a gene is to the right, the top strand is considered the mRNA synonymous strand and if transcription is to the left, the top strand is the template strand (Figure 3). Chargaff's second parity rule states

that for single-stranded DNA %A ≈ %T and %C ≈ %G (Rudner *et al.*, 1968, Karkas *et al.*, 1968) and implies that for regions with clusters of purines there must be local deviations from Chargaff's second parity rule favouring purines (Bell and Forsdyke, 1999). These local deviations from Chargaff's second parity rule, also known as Chargaff differences, have been seen in a variety of organisms including vaccinia virus. Previous work by Bell *et al.* determined that Chargaff differences do correlate with direction of transcription and that the number of A nucleotides was greater than the number of T nucleotides in 83 of 92 vaccinia genes tested (Bell and Forsdyke, 1999).

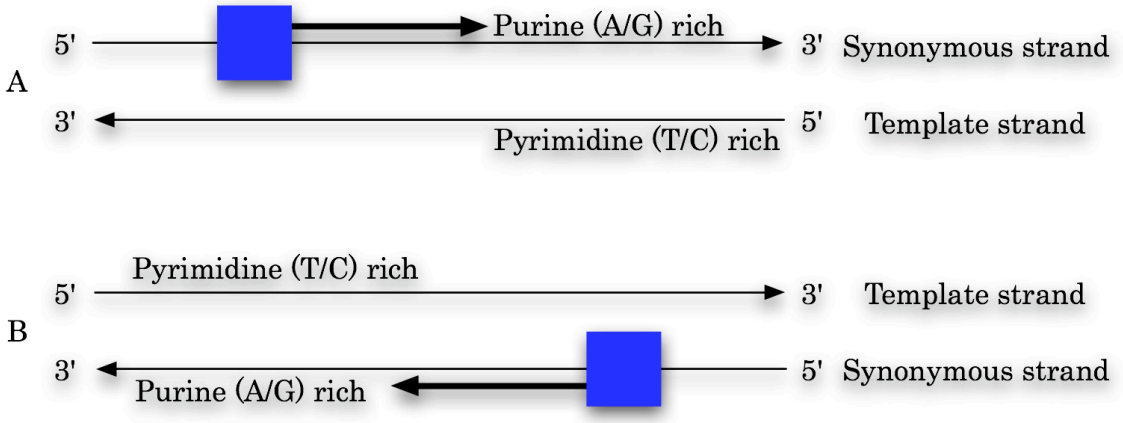


Figure 3. Correlation between transcription direction and purine content on the mRNA synonymous strand.
(A) Rightward transcription direction (B) Leftward transcription direction.

Many programs have been designed to predict genes, but few actually rate the "quality" or significance of the prediction and leave researchers to evaluate this themselves. In poxviruses, predicting which ORFs are likely to be expressed (genes) without the use of biochemical analysis usually involves simply choosing a minimum ORF length cut-off and excluding all ORFs that are smaller than the cut-off. Analysis may be extended to include manual inspection of each

predicted ORF for the presence of promoter consensus sequences, although identifying AT-rich promoters in an AT-rich genome may be difficult. However, excluding ORFs that are smaller in size than the cut-off risks missing genes that are unusually short. This occurred during annotation of vaccinia virus strain Copenhagen (VACV-COP) where at least three recently verified genes (ranging from 162 bp-231 bp) were not included in the initial annotation of the complete genome; these genes, VACV-COP A2.5L (Senkevich *et al.*, 2002a, Senkevich *et al.*, 2002b), A14.5L (Betakova *et al.*, 2000) and G5.5R (Amegadzie *et al.*, 1992) have now been included in the poxvirus section of Virus Orthologous Clusters (VOCs) (Unknown, Ehlers *et al.*, 2002).

Poxvirus genes are transcribed from both DNA strands and so far have never been shown to overlap more than a few nucleotides. Despite this knowledge, some poxvirus genomes have been liberally annotated so as to include all ORFs above a certain size, irrespective of whether they overlap larger well-characterized genes. Thus, the current GenBank file for VACV-COP contains 202 major (large and likely to be real genes) ORFs and 64 minor (small and unlikely to be real genes) ORFs (Goebel *et al.*, 1990a, Upton *et al.*, 2003). The majority of these minor ORFs in VACV-COP overlap larger, major ORFs on the opposite DNA strand.

The results of this chapter will show that for the AT-rich poxviruses, the purine skews can be used to help predict the synonymous (coding) strand, particularly in regions where smaller ORFs overlap each other on opposite strands of the genome and neither have orthologs in other poxvirus genomes. Furthermore, it is shown that the majority of minor ORFs found in VACV-COP are unlikely to be functional genes and that based on purine content, two of the three genes

initially excluded from the annotation of the vaccinia virus genome due to their small size, fit our definition of a major ORF.

2.2 Methods

2.2.1 Purine skews

Purine skews were created using the GraphDNA program (Thomas *et al.*, 2007), which implements the simple algorithm originally developed by Lobry (Lobry, 1996) that assigns a direction to each base encountered in the sequence. In the case of purine skews, the graph begins at position (0,0) and moves upwards one unit if the base encountered is a purine (A or G) and moves downwards one unit if the base encountered is a pyrimidine, (C or T). The plot continues along the x-axis until the end of the sequence is reached. A variable window size can also be set. In this case, the plot trend will be either upwards or downwards, depending on the average number of purines or pyrimidines in the window. The window then slides over the number of bases defined by the window size. For example, if the window size was defined as 10 bp, the window will slide over to the eleventh base and then count the average. The GraphDNA program is accessible as a Java WebStart program at www.virology.ca (Thomas *et al.*, 2007).

2.2.2 Purine/pyrimidine ratio comparison

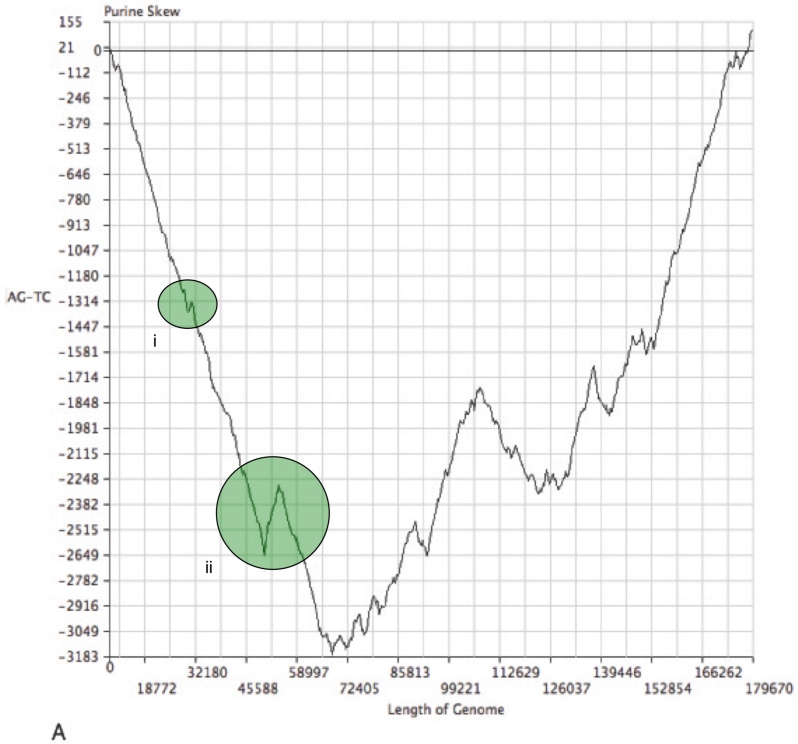
To analyse the ratio of purines to pyrimidines at each codon position, the total number of each nucleotide at each codon position was first calculated using the codontree program with the BC=A option (calculate the base composition at all 3 codon positions) selected (Pesole *et al.*, 1988, Codontree). Once the base composition at each codon position was calculated, the purine to pyrimidine ratio

(R/Y) was calculated for each ORF of the dataset. The mean values of the R/Y ratio for each dataset were compared using Student's T-Test to determine if the mean R/Y ratio for each dataset was statistically different. The null hypothesis for the Student's T-test was that the means were equal and the null hypothesis was rejected if the p-value was < 0.05. The two datasets used for this portion of the paper consisted of (1) all ORFs classified as major in VACV-COP and (2) all ORFs classified as minor in VACV-COP.

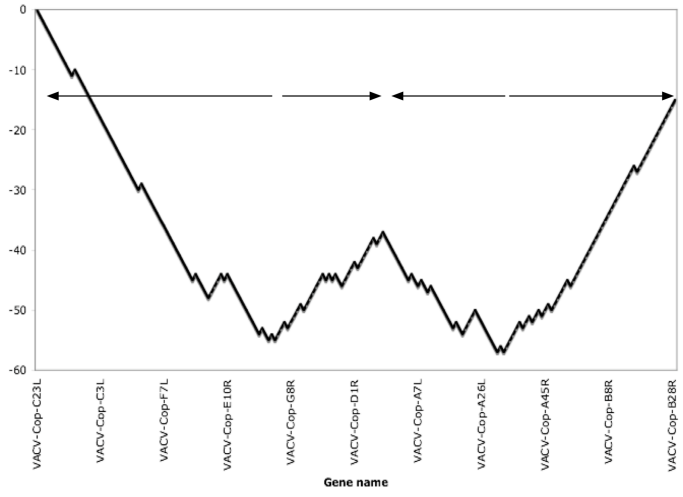
2.3 Results and discussion

The purine content for all ORFs in the VACV-COP genome was determined and it was found that of the 76 ORFs with a purine content under 50%, 52 were classified as minor ORFs. Comparing the purine content between major and minor ORFs shows that 89% of major ORFs had a purine content over 50% and 81% of all minor ORFs had a purine content under 50%. These differences in purine content can be effectively visualized by looking at a purine skew of the VACV-COP genome. Figure 4 shows the genomic purine skew (Figure 4A) and the direction of transcription (Figure 4B) for the major ORFs (genes) in VACV-COP. The direction of transcription (Figure 4B) of the VACV-COP major ORFs was drawn by ordering all major ORFs according to their start position on the genome and then assigning a numerical value of +1 or -1 depending on if the ORF is located on the top or bottom strand respectively. Since the major ORFs of VACV-COP are spread out evenly across the genome, and Figure 4B was created using only the major VACV-COP ORFs, the two figures (Figure 4A and B) follow very similar trends. A characteristic "W" shaped plot can be seen for both graphs; in Figure 4B, this is the result of a trend for large blocks of genes to be transcribed in the same direction (see arrows in Figure 4B). These data

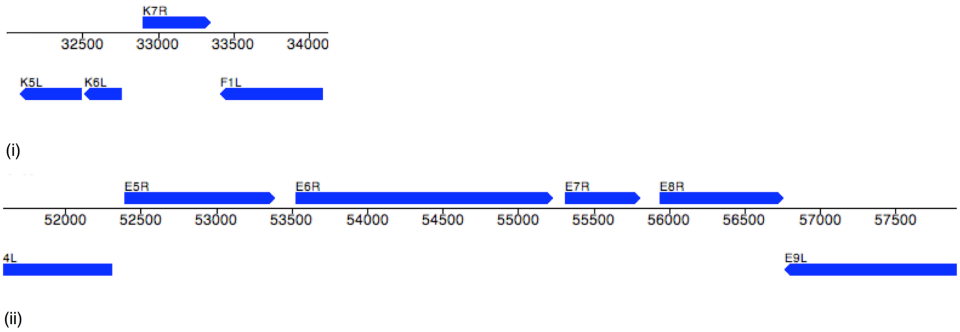
indicate a good correlation between the purine content of the genomic DNA and the direction of transcription; for example, for genes that are transcribed in the leftward direction, the bottom/synonymous strand is purine rich and the opposite is true for genes that are transcribed to the right. The correlation between purine content and the likelihood that an ORF is major is further supported by the fact that 180 of the 202 major ORFs of VACV-COP have a purine content greater than or equal to 50%. In this way, purine skews can be used to help annotate newly sequenced genomes by aiding in the determination of the mRNA synonymous strand.


Purine Skew
AG–TC
i
ii
Length of Genome

A


VACV-Cop-C23L
VACV-Cop-C3L
VACV-Cop-F7L
VACV-Cop-E10R
VACV-Cop-G8R
VACV-Cop-D1R
VACV-Cop-A7L
VACV-Cop-A26L
VACV-Cop-A45R
VACV-Cop-B8R
VACV-Cop-B28R
Gene name

B


K7R
K5L
K6L
F1L
(i)
E5R
E6R
E7R
E8R
4L
E9L
(ii)

C

Figure 4. Correlation between the purine skew and the direction of transcription of the VACV-COP genome, excluding the non-coding terminal inverted repeats. (A) Purine skew drawn using GraphDNA. Regions of the top strand that exhibit a purine bias will have a trend to the upward direction whereas regions that exhibit a pyrimidine bias will be drawn in the downward direction. Two example regions of changes in strand bias are shaded in green and marked (i) and (ii) (B) VACV-COP major ORFs drawn according to the strand of the genome on which each ORF is located. Beginning with a value of zero for the first major ORF of the genome, a numerical value of +1 or -1 is added to the value of the previous ORF depending on if the ORF is located on the top or bottom strand, respectively. (C) Gene orientation in two example regions demonstrating a change in strand bias. (i) Strand bias changes from a purine bias on the bottom strand, to a purine bias on the top strand that encompasses 1 gene on the top strand. (ii) Strand bias changes from a purine bias on the bottom strand, to a purine bias on the top strand that encompasses 4 genes located on the top strand.

When the purine skew (Figure 4A) slopes in the downward direction, this is due to a pyrimidine bias on the top strand and a commensurate purine bias on the bottom strand indicating that the major ORFs are located on the bottom strand. In regions where the purine skew changes direction from a downward slope to an upward slope or vice versa, these are regions on the genome where the transcription direction of the genes in the genome changes. For example, the purine skew appears to change direction from a downward slope to an upward slope at position 32,800 bp and then changes again from an upward slope to a downward slope at position 33,500 bp. Figure 4C (i) shows that within this region (32,800-33,500 bp), there is one gene (VACV-COP K7R) that is located on the top strand (upwards slope on purine skew) and is flanked by genes that are located on the bottom strand (downward slope on purine skew). A second example can be seen in Figure 4C (ii) where an upward slope in the purine skew occurs between positions 52,400 bp and 57,000 bp. In this case, the upward

sloping region encompasses four genes (VACV-COP E5R, E6R, E7R and E8R) and the two downward sloping regions flanking each side of this region encompass genes that are located on the bottom strand.

It was previously shown that minor ORFs in VACV-COP tend to have higher than average serine content as well as lower than average aspartate and glutamate content, due to the nature of the genetic code (Upton, 2000). Based on these observations and our current finding that the synonymous DNA strand is usually purine rich, we created a simple mathematical equation designed to provide a "quality" measure of each ORF. The results of the formula [Ser%-Asp%-Glu%+(50-AG%)], which essentially sums the trends in amino acid composition (3 amino acids) and purine content, are shown in Figure 5. If peptides are translated from ORFs on the non-synonymous strand, they tend to have a higher than average Ser%, but lower than average Asp% and Glu% (due to properties of the genetic code), and have a lower than average purine content. By subtracting the purine content of each ORF from the genome average for VACV-COP (50%), if the ORF is major, the numerical result of the equation is negative and if the results of the equation are positive, the ORF is predicted to be minor.

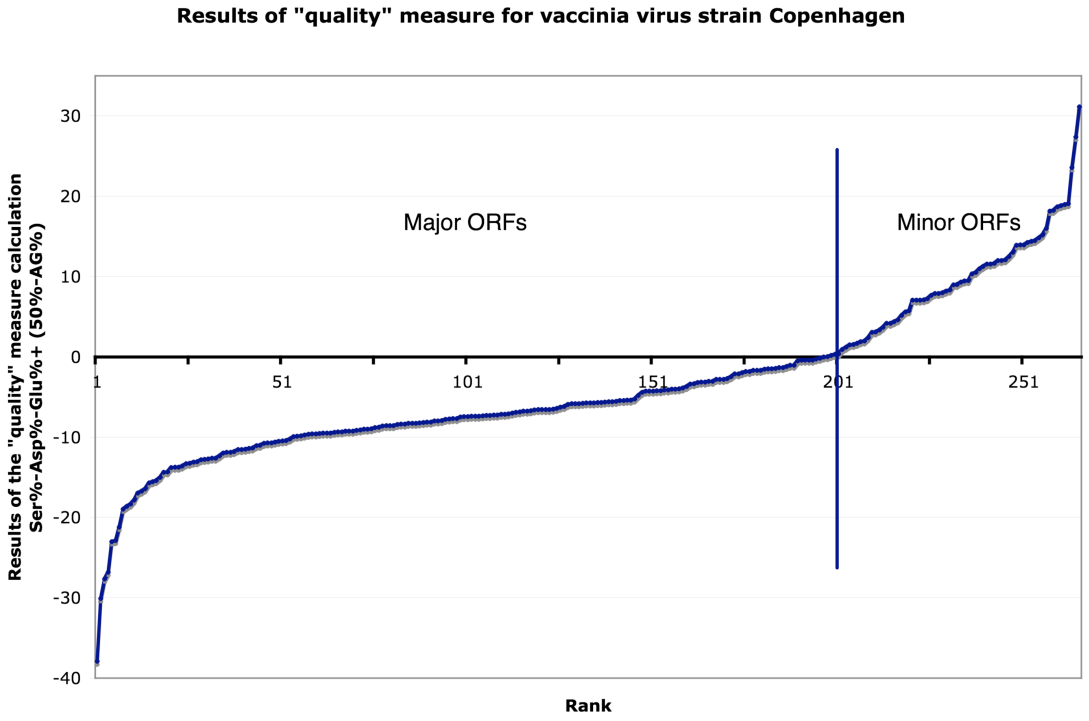


Figure 5. Results of the "quality" measure for VACV-COP.
Y-axis plots results of the "quality" calculation (Ser%-Asp%-Glu%+[50%-AG%]) and X-axis depicts rank of each ORF.

Plotting the results of this equation, we found that all but 6 of the 64 minor ORFs predicted in VACV-COP (VACV-COP A ORF G, VACV-COP A ORF T, VACV-COP B ORF G, VACV-COP C ORF F, VACV-COP E ORF D, and VACV-COP F ORF A) were classified correctly as being minor and all but 9 of the 202 major ORFs (VACV-COP A9L, VACV-COP A13L, VACV-COP A14L, VACV-COP A14.5L, VACV-COP A38L, VACV-COP A43R, VACV-COP C3L, VACV-COP I5L, VACV-COP I6L) were correctly classified as being major. Given that only 15 out of 266 genes were misclassified, the success rate of this measure for the VACV-COP genome was 94%. A possible reason why the 9 major ORFs were likely misclassified is because they are small membrane proteins that had a

lower aspartate and glutamate content than other major ORFs and the 6 minor ORFs were likely misclassified because they have a lower serine and higher purine percentage compared to other minor ORFs despite the fact that all but one minor ORF (VACV-COP A ORF T) overlap a major ORF on the opposite strand (Table 1). There were three genes that had initially been excluded from the annotation of VACV-COP due to their small size. Our measure successfully classified two of these genes (VACV-COP A2.5L and VACV-COP G5.5R) as major. One of these genes (VACV-COP A14.5L) was misclassified as minor likely due to the fact that it is a small membrane protein (Table 1).

Table 1. List of incorrectly classified VACV-COP ORFs.

Major ORFS incorrectly classified as minor						
ORF name	ORF size (bp)	Serine content (%)	Aspartate content (%)	Glutamate content (%)	Purine content (%)	Explanation
VACV-COP A13L	210	11.43	1.43	2.86	48.82	Small, membrane protein
VACV-COP A14L	270	11.11	3.33	0	45.79	Small, membrane protein
VACV-COP A14.5L	159	7.55	1.89	1.89	44.45	Small, membrane protein
VACV-COP A38L	831	7.94	3.97	2.53	47.25	Membrane protein
VACV-COP A43R	582	10.31	5.67	1.55	51.11	Membrane protein
VACV-COP C3L	789	13.31	4.18	3.8	52.27	High Ser%, low Asp% and Glu%
VACV-COP I5L	237	5.06	2.53	1.27	49.58	Small, membrane protein
VACV-COP I6L	1146	10.99	4.45	4.45	49.7	High Ser%, low Asp% and Glu%
Minor ORFs incorrectly classified as major						
ORF name	ORF size (bp)	Serine content (%)	Aspartate content (%)	Glutamate content (%)	Purine content (%)	Explanation
VACV-COP A ORF G	225	6.67	4	8	54.39	Low Ser%, high Asp% and AG%
VACV-COP A ORF T	243	1.23	3.7	2.47	51.63	Overlaps on same strand as major ORF
VACV-COP B ORF G	273	1.1	3.3	1.1	53.26	Low Ser%, high AG%
VACV-COP C ORF F	273	1.1	3.3	1.1	53.26	Low Ser%, high AG%
VACV-COP E ORF D	198	9.09	4.55	6.06	55.72	High Asp%, Glu%, AG%
VACV-COP F ORF A	201	4.48	4.48	0	50.49	Low Ser%

To further test this hypothesis, a similar analysis was repeated for the genome of *Amsacta moorei* entomopoxvirus (AMEV), an extremely AT-rich (82%) poxvirus (Bawden *et al.*, 2000). The AMEV genome was chosen for two reasons: (1) because it is not closely related to any known poxviruses and therefore its genome contains a large number of genes with unknown function and (2) its genome was liberally annotated and therefore it is questionable which ORFs are

likely to be functional genes. Thus, the "quality" measure was used to predict which AMEV ORFs are most likely to be minor. Figure 6 graphically depicts the results of the "quality" measure calculation for AMEV. Due to the extreme AT-richness of the AMEV genome, it was necessary to modify the "quality" measure to the following formula: [Ser%-Asp%-Glu%+(49%-AG%)]. 49% was chosen instead of 50% for the purine portion of this equation since the average purine content of the entire AMEV genome is 49%. As was the case with VACV-COP, if the ORF is minor, the results of the "quality" measure will be positive.

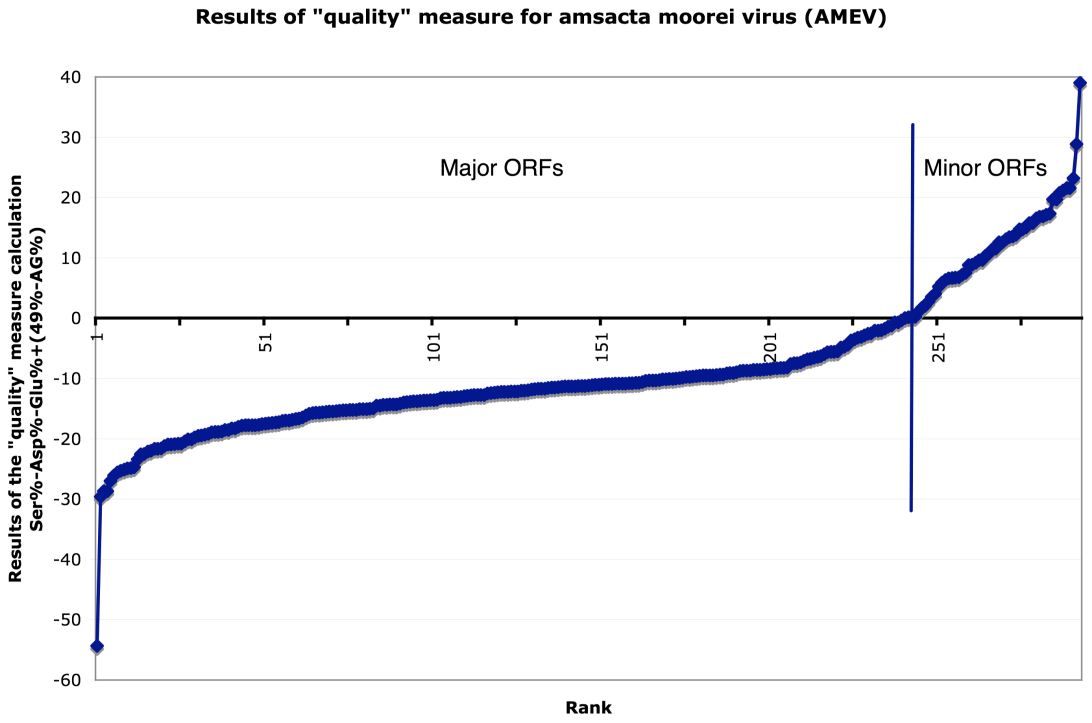


Figure 6. Results of the "quality" measure for *Amsacta moorei* virus (AMEV). Y-axis plots results of the "quality" calculation (Ser%-Asp%-Glu%+[49%-AG%]) and X-axis depicts rank of each ORF.

It was found that there were 51 ORFs that had a positive "quality" value and are therefore considered minor. Of these 51 ORFs, 41 ORFs further fit our

definition of a minor ORF as they overlapped another larger ORF on the opposite strand and 4 major ORFs (AMEV-161, AMEV-164, AMEV-171, and AMEV-183) were incorrectly classified as minor even though they each have orthologs in other poxviruses and are most likely functional genes (Table 2). The remaining 6 ORFs (AMEV-001, AMEV-089, AMEV-148, AMEV-198, AMEV-ITR02, and AMEV-ITR08) that were classified as minor using our "quality" measure were found not to overlap any ORFs on the opposite or same DNA strand and were further analyzed using the AMEV purine skew in order to try and determine the correct coding strand in each of these 6 regions (Table 3). For 5 (AMEV-001, AMEV-089, AMEV-148, AMEV-198, AMEV-ITR02) of these 6 ORFs, the coding strand is opposite to the strand on which these ORFs are located, indicating that these ORFs are minor. In 1 case (AMEV-ITR08), the ORF is located on the same strand as was predicted by the purine skew and therefore this ORF may actually be major. AMEV-ITR08 does not have any orthologs in other poxviruses but it does show a 73.6% amino acid identity with the AMEV-ITR07 ORF which was classified as being major using the "quality" calculation further supporting that AMEV-ITR08 is likely major. AMEV-ITR08 was predicted to contain a transmembrane domain (Bawden *et al.*, 2000), which could explain why it was misclassified.

Table 2. List of potentially incorrectly classified AMEV ORFs.

Major ORFS incorrectly classified as minor						
ORF name	ORF size (bp)	Serine content (%)	Aspartate content (%)	Glutamate content (%)	Purine content (%)	Explanation
AMEV-161	243	11.11	2.47	1.23	47.56	Membrane protein
AMEV-164	708	7.63	2.12	2.97	47.97	High Ser%, low Asp% and Glu%
AMEV-171	276	3.26	1.09	1.09	48.39	Low Asp% and Glu%
AMEV-183	675	6.67	3.11	1.33	51.18	Low AG% and low Glu%
Minor ORFs incorrectly classified as major						
ORF name	ORF size (bp)	Serine content (%)	Aspartate content (%)	Glutamate content (%)	Purine content (%)	Explanation
AMEV-152	225	0	12	1.33	60.97	Overlaps on same strand as major ORF
AMEV-189	180	1.67	8.33	1.67	43.17	Overlaps on opposite strand as major ORF
AMEV-191	228	0	2.63	10.53	61.9	Overlaps on same strand as major ORF

Table 3. List of 6 AMEV ORFs classified as minor that do not fit the definition of a minor ORF.

ORF name	DNA strand on which ORF is located	Direction of purine skew	Conclusion
AMEV-001	Top	Down	Minor
AMEV-089	Top	Down	Minor
AMEV-148	Bottom	Up	Minor
AMEV-198	Bottom	Up	Minor
AMEV-ITR02	Top	Down	Minor
AMEV-ITR08	Top	Up	May be major

There were three ORFs that had been classified as major (negative value for the "quality" measure) yet overlapped a larger gene on the opposite or same DNA strand (Table 2). Two of these ORFs (AMEV-152 and AMEV-191) overlap

a larger ORF on the same strand and therefore neither the purine skew nor the "quality" measure are capable of determining which ORF is major; and one ORF (AMEV-189) overlaps the much larger spheroidin gene on the opposite strand and was likely misclassified due to its lower than average serine content and higher than average aspartate content.

For the analyses shown in Figure 5 and Figure 6, the cut-off value used in both cases was zero. The value of zero was chosen in the training case (VACV-COP) because it represented a reasonable cut-off between genes that were known to be major and ORFs that were known to be minor with minimal misclassification of genes. With our test case (AMEV), since it was not known which ORFs were major or minor, a cut-off of zero was initially used with the presumption that the cut-off may need to be adjusted due to the extreme AT-richness of the AMEV genome. Analyzing the "quality" measure data obtained for AMEV with a cut-off of zero yielded satisfactory results in that the number of overlapping and therefore likely to be minor ORFs that were misclassified was relatively low and because of this we decided to maintain the zero cut-off. It is likely that a cut-off of zero worked well with AMEV despite its extremely AT-rich genome because the "quality" measure that was used reflected the average AG% of the genome. It is also likely that other poxvirus genomes that are analyzed using our method would use a cut-off of zero, provided the "quality" measure that was used was changed to reflect the average AG content of the genome.

Thus far we have shown that purine skews can be used to predict the coding strand of poxvirus genomes and that major ORFs in VACV-COP and in AMEV usually contain greater than 50% and 49% purines, respectively. In order to explain this purine richness in genes, the purine (R) to pyrimidine (Y) ratio (R/Y)

was calculated for each codon position of each coding and non-coding ORF in VACV-COP. A Student's T-test was used to compare the mean R/Y ratio values for the coding (genes) and non-coding ORFs at each codon position; means were considered statistically different when the p-value was less than 0.05. At the first nucleotide position in the codon, both VACV-COP major and minor ORFs exhibit a bias towards purines, but the major ORFs (genes) have significantly ($p<0.05$) higher levels of purines at this position (Table 4). At the second nucleotide position the major ORFs have a R/Y ratio of approximately 1 and the minor ORFs have a significantly lower R/Y ratio ($p<0.05$) indicating that minor ORFs are pyrimidine rich at the second codon position whereas major ORFs contain roughly equal amounts of purines and pyrimidines at this position. At position 3, no statistical difference was found, with both major and minor ORFs being rich in pyrimidines. Thus, for the first and second nucleotide positions of the codons, the major ORFs (genes) have significantly higher purine content than the minor ORFs.

Table 4. Mean purine to pyrimidine ratios for each codon position of vaccinia virus Copenhagen major and minor ORFs.
Positions marked with an asterisk (*) are statistically different.

	Purine/Pyrimidine (R/Y) ratio at each codon position		
	Position 1*	**Position 2***	**Position 3**
Major ORFs	1.77	0.99	0.93
Minor ORFs	1.21	0.75	0.96

It is important to remember that the use of purine and amino acid content of the coding strand and predicted protein, respectively, are just two measures that can be used to help predict whether an ORF is likely to be a functional gene and

that usually they are only useful in discriminating between coding and non-coding strands. However, ORFs that are fragments of *bone fide* genes may also be flagged as non-functional by this analysis, because protein sub-domains tend to have an amino acid composition that varies from the average. An example of this is the A25L ORF of VACV-COP that was flagged as non-functional by this method even though it is a fragment of the ATI protein. In a similar way, fragmentation of genes into smaller ORFs can also lead to unusual isoelectric points in the resulting predicted proteins; the 14 ORFs with a predicted pI of >9.6 are all minor ORFs or gene fragments. Thus, multiple approaches that may also include promoter analysis must be applied to attempt to correctly annotate small orphan ORFs in these genomes and there is no guarantee that the process will be 100% successful.

2.4 Conclusions

The overall goal of all poxvirus research currently being undertaken is to understand the replication and infection cycle that poxviruses undergo. In order to have this understanding, we need to know the function of every gene in the poxvirus genome and this cannot be done until we know which proteins the virus encodes. To that end, we have taken a bioinformatics approach to predict which ORFs are functional in AT-rich poxvirus genomes. We have successfully shown that in the case of AT-rich poxviruses, purine skews can be used to help predict the coding regions of the genome. This is particularly useful if predicted ORFs overlap each other and it is not apparently obvious which ORF is major (when neither ORF has an ortholog in another poxvirus genome). A second method that can be used in conjunction with purine skews is to calculate the

"quality" of each predicted ORF using information from amino acid composition and purine content. For a given ORF, if the results of this calculation are negative the ORF is predicted to be a functional gene, and if the results of the calculation are positive, the ORF is predicted to be minor.

By comparing purine to pyrimidine (R/Y) ratios at each codon position of major and minor vaccinia virus ORFs, it was found that the purine abundance seen for major ORFs stems primarily from the first codon position with both the second and third codon positions containing equal amounts of purines and pyrimidines.

The software used to create the purine skews (GraphDNA) and the VOCs database are both available for public use via the web at www.virology.ca (Thomas *et al.*, 2007, Ehlers *et al.*, 2002).

3.0 Host-derived pathogenicity islands in poxviruses

Published in *Virology Journal* in 2005 (Da Silva and Upton, 2005b).

3.1 Introduction

When making comparisons of genomes, alignment of orthologous regions is a common starting point. The alignment of large DNA sequences (>10 kb) is frequently not always possible because both local and global considerations must be taken into account. One of the first methods used to compare two genomes was the dotplot, which provides a graphical view of sequence similarity and can be used to evaluate alignments made by other algorithms (Sonnhammer and Durbin, 1995). Dotplots can be used for any length of DNA sequence and also for protein sequences, although they are especially useful when trying to get a global view of the relationship between large DNA sequences.

In the simplest form of a dotplot, one sequence is placed along the x-axis and the other sequence along the y-axis to create a matrix (Figure 7). Wherever one nucleotide of one sequence is identical to a nucleotide of the other sequence, the appropriate cell of the matrix is filled. Most dotplot programs, however, use a sliding window of user-defined size, and a measurement of similarity between the two windows instead of comparing each nucleotide along the genomes. When plotting the data, a greyscale can be used to record differing degrees of similarity. If the two sequences are identical, a fully black diagonal line is observed on the dotplot comprised of many dots, each drawn for a window that matches an identical window on the second sequence. The background of the dotplot is therefore comprised of greyscale dots representing random matches of varying similarity between the windows used to scan the two sequences. Figure 7 shows a simplified example of a typical dotplot. For each region of the

sequence plotted on the x-axis that is identical to the sequence on the y-axis, a diagonal line is plotted and, for the sake of this example, shaded an appropriate colour to show where each region is identical. Dotter (Sonnhammer and Durbin, 1995) and our java version of this program (JDotter) (Brodie *et al.*, 2004a) are especially useful implementations because the window size and degree of similarity used for the dotplot display can both be manipulated "on-the-fly", without requiring recalculation of the whole plot, which is CPU intensive. Changing these parameters can help visualize regions of relatively low similarity or to enhance or decrease the background (random) matches in the plot.

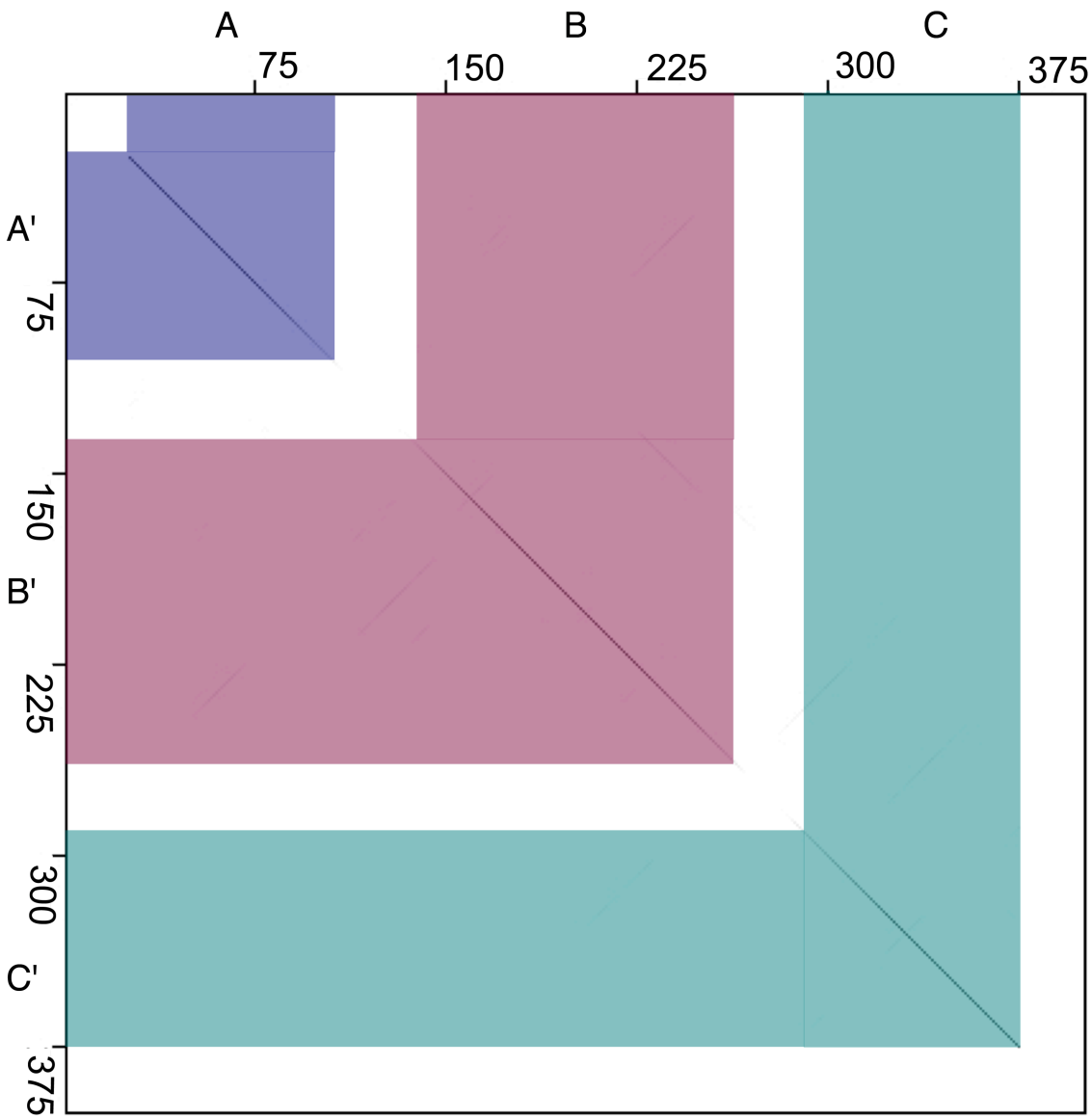


Figure 7. An example of a typical dotplot.
Regions on the x-axis that are identical or similar to regions on the y-axis are marked A, B and C and correspond to regions A', B', and C' on the y-axis respectively. Regions that correspond to each other are also colour-coded in purple for regions A and A', pink for regions B and B' and turquoise for regions C and C'.

This project arose because we noticed a background pattern resembling stripes when using a dotplot to locate repeat regions on the molluscum contagiosum virus genome (MOCV-1), which led us to question the causes of this stripe

pattern. We also observed this unusual background pattern when viewing many of the dotplots comparing different poxvirus genomes plotted against each other and themselves (Figure 8). These unusual non-random banding patterns suggested that the composition of discrete regions of the genomes differed significantly from the overall composition of the poxvirus genomes.

3.2 Methods

3.2.1 Creation of dotplots

Dotplots for the molluscum contagiosum virus genome were created and visualized using JDotter with a default window size of 26 nucleotides using a minimum cut-off score of 40 and a maximum cut-off score of 100 on the GreyMap tool in order to better visualize the background patterns (Brodie *et al.*, 2004a).

The dotplot shown in Figure 10 was created using the Dotter program with the complete molluscum contagiosum virus genome plotted on the x-axis and a 250 kb random sequence with increasing G+C content plotted on the y-axis (Sonnhammer and Durbin, 1995). The 250 kb sequence was created using DNACreator, a program that creates a random sequence with a given G+C content (Hu, 2002). Segments of 50 kb and varying G+C content were created and concatenated into one 250 kb sequence that contained an increasing G+C content.

3.2.2 Codon usage

3.2.2.1 Comparison between genes in regions 1 and 2 and 49 conserved MOCV-1 genes

The program CodonW was used to calculate the Relative Synonymous Codon Usage (RSCU) values of 49 molluscum contagiosum virus genes that are conserved in all poxviruses, and the 5 genes comprising each of region 1 and 2 (Figure 8 and Table 5) (Peden, 1999). The mean RSCU values for each codon excluding the codons for methionine and tryptophan (62 in total) from regions 1 and 2 were initially compared to each other using a two-tailed Student's T-test with the null hypothesis that the means are equal being rejected if the p-value was less than 0.05. From the results of these T-tests, the RSCU values for each region were pooled in order to create a larger data set to perform subsequent Student's T-tests. The mean RSCU values for four codons in regions 1 and 2 were found to be statistically different and were therefore treated as individual data sets in subsequent T-tests. These codons were CCA (proline), CAU and CAC (histidine), and CAG (glutamine).

With the RSCU values from regions 1 and 2 being pooled for all but 4 codons, the mean values for this pooled dataset were compared to the mean RSCU values obtained for the 49 conserved MOCV-1 genes again using a two-tailed Student's T-test. The null hypothesis that mean RSCU values from the two datasets are equal was rejected only when the p-value was less than 0.05.

3.2.2.2 Comparison between MOCV-1 and 50 human genes

Using the above outlined method, the mean RSCU values for the genes in regions 1 and 2 were compared to the mean RSCU values of 50 randomly selected human genes. The 50 human genes were randomly selected using the

random command in Microsoft Excel, from the dataset used by the codon usage database at the Kazusa DNA Research Institute (Nakamura *et al.*, 2000) that consists of 76,893 human coding sequences that were taken from the NCBI GenBank FlatFile release 145.0 (January 25, 2005). The comparison of the codon usage of 50 human genes with the codon usage of 49 MOCV-1 genes was performed using the same methods outlined above. In each of the analyses performed, the null hypothesis that mean RSCU values from the two datasets are equal was rejected when the p-value was less than 0.05

3.3 Results and discussion

A self-dotplot of MOCV-1, where the viral genome is plotted on both the x- and y-axes, is shown in Figure 8. Since the comparison uses the identical genome on each axis, there is an unbroken diagonal line running from the upper left corner to the lower right corner of the dotplot. A number of horizontal and vertical stripes can be seen scattered throughout the plot background; such patterns are observed in most poxvirus genome self-plots although the intensity varies considerably between stripes in a single plot and tend to be more intense in those genomes with the most extreme nucleotide composition, AT- or GC-rich genomes. Two of the most striking regions for MOCV-1 are marked on Figure 8. Regions 1 and 2 are located at 71,550 - 74,790 bp and 151,470 - 157,950 bp on the MOCV-1 genome, respectively. Because each region encompasses 5 genes, it was evident from the dotplot itself that sets of small DNA repeats were not responsible for generating these unusual stripes in the background of the plot. On thinking about how these stripes, which represent less than average random scoring background regions, could be produced, the simplest explanation was unusual base composition, and because these regions are so large, this suggested

these might be regions with interesting biology. Figure 9 shows that these two regions marked in Figure 8 do indeed have a G+C composition significantly lower than the genome average. Regions 1 and 2 have G+C averages of 52.58% and 50.38% respectively, whereas the G+C composition of MOCV-1 is 63.36% (also shown as a horizontal line in Figure 9A and B).

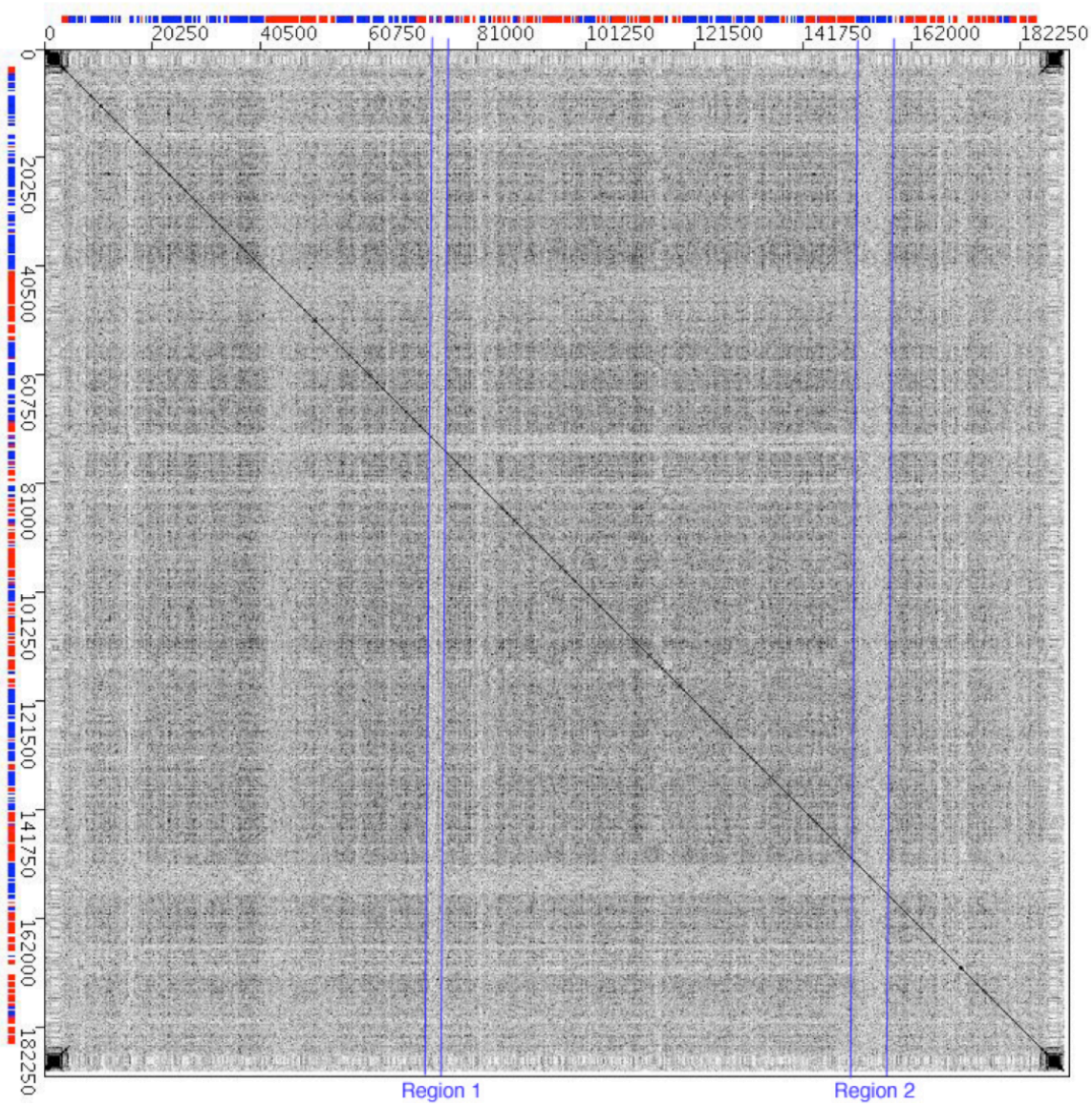
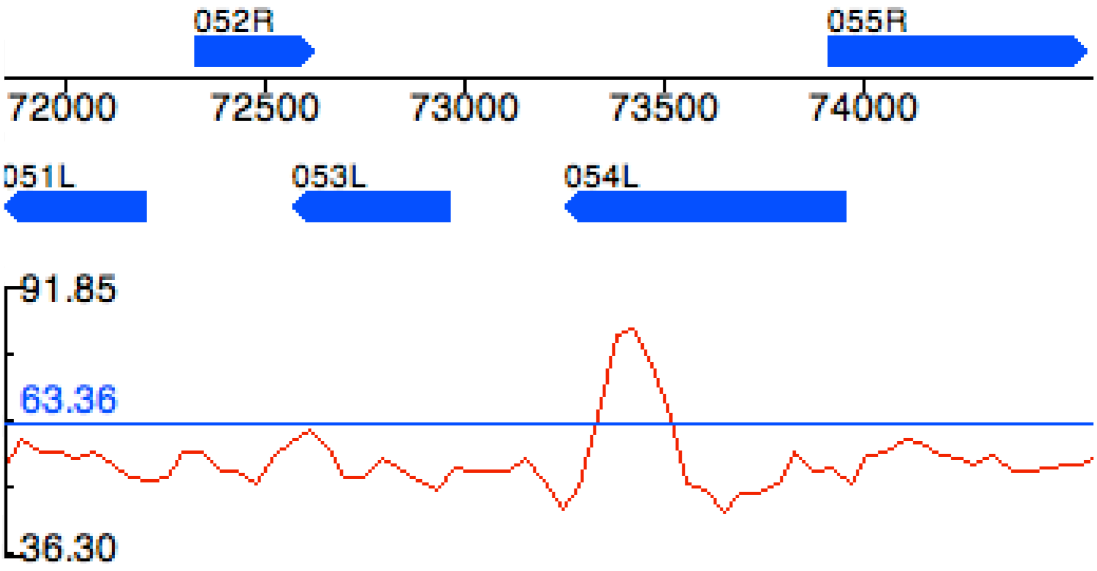
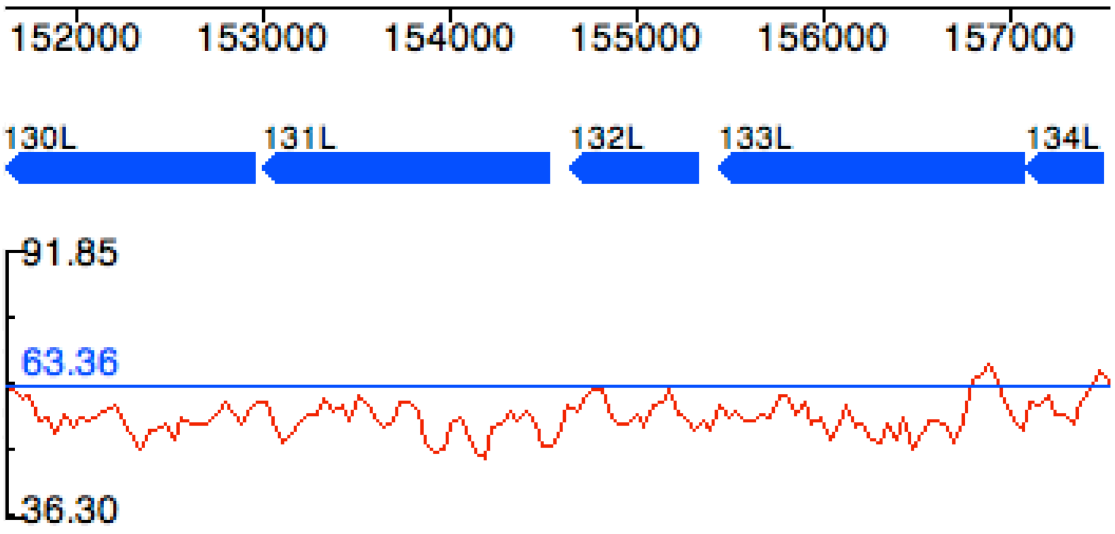


Figure 8. Dotplot depicting a comparison of the molluscum contagiosum virus genome to itself.

Region 1 begins at position 71,550 bp and ends at position 74,790 bp and contains 5 genes. Region 2 begins at position 151,470 bp and ends at 157,950 bp and contains 5 genes. Blue and red bars across each axis represent each gene as it is located on the MOCV-1 genome with red bars representing genes located on the top strand and blue bars representing genes located on the bottom strand.


052R
055R
72000
72500
73000
73500
74000
051L
053L
054L
91.85
63.36
36.30

A


152000
153000
154000
155000
156000
157000
130L
131L
132L
133L
134L
91.85
63.36
36.30

B

Figure 9. G+C composition plots created using viral genome organizer (VGO) (Upton *et al.*, 2000).

A window size of 135 bp was used for (A) region 1 of Figure 8 and (B) region 2 of Figure 8. The scale on the left hand side of the plot shows the maximum (91.85%) and minimum (36.30%) G+C content of the MOCV-1 genome. The line drawn through the plot indicates the average G+C content of the MOCV-1 genome (63.36%). Numbered blue bars represent the genes in each region and the scale across the top of each panel shows the location of each gene on the genome.

The genes present in regions 1 and 2 are listed in Table 5 together with the G+C content and putative function. Region 1 contains 5 genes (MOCV-1-051L, MOCV-1-052R, MOCV-1-053L, MOCV-1-054L, MOCV-1-055R), two of which have unknown function (MOCV-1-052R and MOCV-1-055R). Interestingly, three of the five genes in region 1 show significant similarity to each other. The gene product of MOCV-1-054L is a secreted glycoprotein that binds interleukin-18 (Xiang and Moss, 1999a) and shows sequence similarity to human and mouse IL-18 binding proteins with three conserved cysteine residues shared between the human and MOCV-1 IL-18 binding proteins (Xiang and Moss, 1999b); although no function has been predicted for the hypothetical products of related genes MOCV-1-051L and MOCV-1-053L, they are predicted to be secreted from infected cells (Xiang and Moss, 1999b). Since the other poxviruses that encode an ortholog of the MOCV-1 interleukin-18 binding protein do not contain orthologs of MOCV-1-051L and MOCV-1-053L, it appears likely that these genes have arisen from duplications of MOCV-1-054L after the divergence of MOCV from the other poxvirus genera.

Table 5. Description of genes in regions 1 and 2.

Gene names, protein length, GC% and putative function of all genes described located in regions 1 and 2 (Figure 8).

Region	Gene Name	Protein length (aa)	G+C%	Putative Function
1	MOCV-1-051L	120	55.92	Secreted glycoprotein
	MOCV-1-052R	100	54.54	Unknown
	MOCV-1-053L	133	55.47	Secreted glycoprotein
	MOCV-1-054L	235	58.05	Secreted IL-18 binding protein
	MOCV-1-055R	216	55.76	Unknown
2	MOCV-1-130L	451	57.01	A-Type inclusion (ATI) protein
	MOCV-1-131L	513	56.48	Intracellular mature virion surface protein (ATI-factor)
	MOCV-1-132L	229	58.7	Unknown
	MOCV-1-133L	546	57.4	Intracellular mature virion surface protein (ATI-factor)
	MOCV-1-134L	141	58.22	Intracellular mature virion membrane protein (associated with virus entry)

Region 2 is also comprised of 5 genes (MOCV-1-130L, MOCV-1-131L, MOCV-1-132L, MOCV-1-133L, and MOCV-1-134L). Three of the five genes (MOCV-1-130L, MOCV-1-131L and MOCV-1-133L) have low, but significant, similarity to each other and also to both the A-type inclusion (ATI) protein and orthologs of the vaccinia virus P4c gene (A26L; strain Copenhagen). In some strains of cowpox virus, the ATI protein functions to surround intracellular mature virus (IMV) particles in the cytoplasm of the host cell with the P4c protein playing a role in directing the IMV particles to the A-type inclusions (Funahashi *et al.*, 1988b, McKelvey *et al.*, 2002). In some orthopoxviruses such as vaccinia and

variola virus, the ATI gene is fragmented and is predicted to encode a truncated ATI protein. This apparently truncated gene is well conserved among the orthopoxviruses, but it is not known what, if any, function the truncated protein performs (Ulaeto *et al.*, 1996). Little is known about the role of the ATI proteins in the MOCV-1 replication cycle, however, given that these proteins are truncated compared to their cowpox orthologs, it is likely that they have a somewhat different, if any, function in MOCV-1. There are two other genes in region 2; MOCV-1-132L is not found in any other poxvirus and its function is unknown, and MOCV-1-134L is conserved in all poxviruses and shows significant sequence similarity (50% amino acid identity) to the A28L gene of vaccinia virus strain Copenhagen and is an intracellular mature virion membrane protein which is associated with virus entry into the host cell (Senkevich *et al.*, 2004a, Senkevich *et al.*, 2004b).

There are several possible explanations for the observed differences in G+C content of the genes in regions 1 and 2 compared to the rest of the genome. The MOCV-1-054L gene in region 1, for example, was most likely recently acquired from an AT-rich host or from an AT-rich virus with subsequent duplications and divergence resulting in the MOCV-1-051L and MOCV-1-053L genes. Since these genes make up a relatively AT-rich region on the MOCV-1 genome, they may have served as a target site for the subsequent acquisition, presumably through non-homologous recombination of genes MOCV-1-052R and MOCV-1-055R, neither of which have been assigned a function. Given that the ATI genes found in region 2 are relatively AT-rich compared to the MOCV-1 genome, these genes may have also been acquired relatively recently from an AT-rich poxvirus. Similar to the genes in region 1, an initial acquisition event leading to an AT-rich

region may have led to further acquisition events and the creation of region 2. Although the MOCV-1-134L gene appears to have been acquired from an AT-rich poxvirus, the situation is not simple because the gene has been found to be essential for vaccinia virus entry into the host cell (Senkevich *et al.*, 2004a, Senkevich *et al.*, 2004b). One explanation is that an essential MOCV-1 ortholog was replaced by a similarly functioning gene from an AT-rich poxvirus, thus explaining why it differs in G+C content despite being an essential gene. Alternatively, it is possible that the MOCV-1-134L gene was acquired at the same time that the ATI genes were, since the gene order is maintained when comparing the VACV and MOCV genomes.

It is interesting to note that there is a short sequence in the AT-rich region 1 that has a significantly higher G+C composition than the rest of the region (Figure 9A). This area can be seen on the dotplot as a thin, dark stripe located in the middle of the band for region 1 (Figure 8). This short spike of G+C rich DNA is located at the 3' end of the MOCV-1-054L gene in a region that does not align with either of the two related MOCV-1 genes (MOCV-1-051L and MOCV-1-053L) indicating that this region may have also been acquired or generated relatively recently; this region results in an in-frame insertion in the predicted MOCV-1-054L polypeptide.

To further examine the effect of nucleotide composition on dotplot results and to collect additional evidence that the unusual stripes seen in the background of the MOCV-1 dotplot could be due to a lower than average G+C content of these particular regions of the genome, a dotplot was created with the MOCV-1 genome plotted on the x-axis against a 250 kb DNA sequence consisting of five 50 kb randomized DNA segments with specific G+C content on the y-axis

(Figure 10). Regions 1 and 2 are marked on the figure and the individual 50 kb segments are clearly visible on the dotplot. As the G+C content increases along the 250 kb sequence on the y-axis, the stripes seen for regions 1 and 2 change colour from dark to light, and this reflects a reduction in random matches between the two sequences. At 60% G+C content, the bands for both regions 1 and 2 appear similar to those observed in the MOCV-1 self-dotplot (Figure 8). When the randomized sequence has a G+C content of 50%, the stripes for regions 1 and 2 appear to disappear, as they merge with the background, indicating that the G+C content of these regions is approximately 50%. As the G+C content further decreases to less than 40%, the bands seen for regions 1 and 2 are again visible but are the negative image of those seen in Figure 8; the bands are darker than the background because these regions have a G+C content that is more similar to the randomized sequence than the rest of the MOCV-1 genome. Similarly, each of the 50 kb segments of the random sequence used for Figure 10 produce a different level of background matches in the dotplot and appears as a broad horizontal stripe.

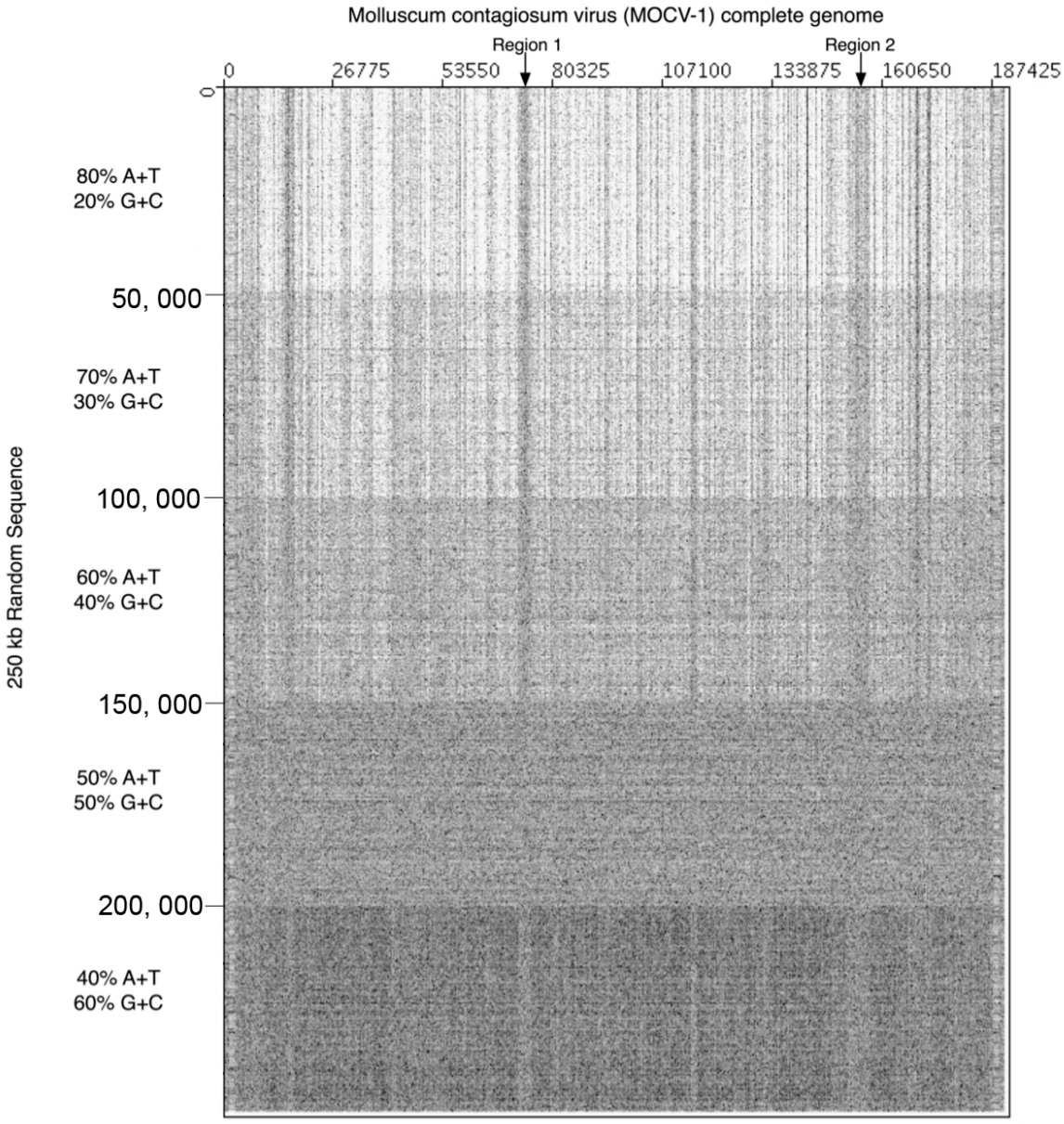


Figure 10. Dotplot comparing molluscum contagiosum virus genome to a random sequence of different nucleotide content.
The random sequence consists of 50 kb segments with differing A+T content, which is indicated on the dotplot. Regions corresponding to the regions seen in Figure 8 are indicated with arrows.

Locating regions of differing G+C content using dotplots with two genomes can be a difficult task given that the intensity of the banding patterns observed

are often diminished in poxvirus genomes that are not extremely AT- or GC-rich. However, we noticed that these banding patterns were easily distinguished when comparing the MOCV-1 genome to a random sequence with increasing A+T content (Figure 10). For example, areas on the MOCV-1 genome with higher than average G+C content that were previously undetected when observing a self-plot of the MOCV-1 genome can be visualized on the dotplot of the MOCV-1 genome versus a series of random sequences with decreasing A+T content (compare Figure 8 and Figure 10). Thus, this type of comparison can be used to enhance recognition of regions of unusual nucleotide composition. An advantage to using this albeit low-tech approach is that it uses a screening tool that is extremely sensitive to minor changes in patterns, the human eye.

In order to further illustrate the differences between the genes in regions 1 and 2 compared to the whole genome, the codon usage of 49 MOCV-1 genes that are conserved in all poxviruses was compared to the codon usage of the 10 genes in regions 1 and 2. The mean relative synonymous codon usage (RSCU) values of the two datasets were compared using a Student's T-test. The codons that were found to have statistically different codon usage between the two datasets are listed in Table 6. We found that the codon usage of 43 of 62 codons (69%) was statistically different between all genes in regions 1 and 2 and the 49 conserved MOCV-1 genes. Codon usage was deemed statistically different when the p-value was less than 0.05.

Table 6. Codon usage differences between regions 1 and 2 and 49 MOCV-1 genes conserved in all poxviruses.

The codon usage between regions 1 and 2 was first determined to be statistically equal and was then used to compare to the codon usage of 49 MOCV-1 genes that are conserved in all poxviruses. Student's T-test was used to compare codon usage with null hypothesis assuming codon usage was equal. *Codon usage was considered different if p-value < 0.05

Amino Acid	Fraction of codons with different usage	Codons that have statistically different usage*
Phe	2/2	UUU, UUC
Tyr	2/2	UAU, UAC
His	2/2	CAU, CAC
Gln	2/2	CAA, CAG
Asn	2/2	AAU, AAC
Lys	2/2	AAA, AAG
Asp	2/2	GAU, GAC
Glu	2/2	GAA, GAG
Cys	2/2	UGU, UGC
Leu	4/6	UUA, UUG, CUU, CUG
Ile	1/3	AUC
Val	2/4	GUU, GUG
Ser	2/6	UCU, AGC
Pro	3/4	CCU, CCC, CCA
Thr	2/4	ACU, ACA
Ala	2/4	GCU, GCA
Arg	5/6	CGU, CGC, CGA, CGG, AGA
Gly	3/4	GGU, GGC, GGA
Stop	1/3	UGA

The codon usage of the genes in regions 1 and 2 was subsequently compared to that of 50 human genes to see if a connection could be made between the codon usage of these genes in MOCV-1 to the host (human) genome, using the same method used for the comparison of regions 1 and 2 with the 49 conserved MOCV-1 genes. The codon usage of the genes in regions 1 and 2 was found to be 68% (48/62 codons) identical to the codon usage of the 50 human genes tested (Table 7). As a control, the codon usage of the same 50 human genes was

compared to the codon usage of the 49 conserved MOCV-1 genes and was found to be statistically different in 56 of 62 codons (90%) (Table 8).

Table 7. Codon usage differences between regions 1 and 2 and 50 human genes.

*Codon usage was considered different if p-value <0.05

Amino Acid	Fraction of codons with different usage	Codons that have statistically different usage*
Phe	0/2	-
Tyr	0/2	-
His	1/2	CAC
Gln	0/2	-
Asn	0/2	-
Lys	0/2	-
Asp	2/2	GAU, GAC
Glu	2/2	GAA, GAG
Cys	0/2	-
Leu	2/6	UUG, CUG
Ile	0/3	-
Val	0/4	-
Ser	1/6	UCG
Pro	2/4	CCU, CCG
Thr	2/4	ACC, ACG
Ala	2/4	GCU, GCG
Arg	4/6	CGU, CGC, AGA, AGG
Gly	0/4	-
Stop	2/3	UAG, UGA

Table 8. Codon usage differences between 49 MOCV-1 genes that are conserved in all poxviruses and 50 human genes.

*Codon usage was considered different if p-value <0.05

Amino Acid	Fraction of codons with different usage	Codons that have statistically different usage*
Phe	2/2	UUU, UUC
Tyr	2/2	UAU, UAC
His	2/2	CAU, CAC
Gln	2/2	CAA, CAG
Asn	2/2	AAU, AAC
Lys	2/2	AAA, AAG
Asp	2/2	GAU, GAC
Glu	2/2	GAA, GAG
Cys	2/2	UGU, GUC
Leu	6/6	UUA, UUG, CUU, CUC, CUA, CUG
Ile	3/3	AUU, AUC, AUA
Val	3/4	GUU, CUA, GUG
Ser	5/6	USU, UCA, UCG, AGU, AGC
Pro	4/4	CCU, CCC, CCA, CCG
Thr	3/4	ACU, ACA, ACG
Ala	3/4	GCU, CGA, GCG
Arg	6/6	CGU, CGC, CGA, CGG, AGA, AGG
Gly	3/4	GGU, GGC, GGA
Stop	2/3	UAG, UGA

Since the codon usage of the genes in regions 1 and 2 was found to be statistically different to the codon usage of 49 conserved MOCV-1 genes for 69% of the codons tested yet was found to be only 34% different from the codon usage of the 50 human genes tested, it is consistent with the hypothesis that these genes may be recent acquisitions from the host genome or a poxvirus genome with higher A+T composition. Recently, a paper has been published which identifies the MOCV-1 interleukin-18 binding protein (as well as several other poxvirus proteins) as a possible acquisition from the host genome (Hughes and Friedman, 2005), which further supports our hypothesis that the genes in regions 1 and 2 were likely acquired from external sources.

Pathogenicity islands (PAI) in bacterial genomes contain several distinct structural features, which include the presence of virulence genes within the PAI, a nucleotide composition of the PAI that is different from the remainder of the genome, and occupancy of large regions of the bacterial chromosome (Schmidt and Hensel, 2004). The genes in regions 1 and 2 also occupy a relatively large region of the MOCV-1 genome, and some of the genes in these regions are known virulence factors with the function of the remaining genes yet to be determined. Thus, we suggest that these regions may be analogous to bacterial pathogenicity islands. This notion is further supported by the observations in this section, which showed the differences in nucleotide composition and codon usage of the genes in regions 1 and 2.

3.4 Conclusions

The data presented in this chapter suggest that the unusual striped banding pattern seen on the dotplots of poxvirus genomes is due to differences in nucleotide content of the genes in these regions compared to the remainder of the genome. The difference between the genes in these regions and 49 MOCV-1 genes that are conserved in all poxviruses was further shown by comparing the codon usage of 62 codons in each of these datasets. The codon usage was found to differ significantly in 69% of the codons tested. The codon usage of the genes in regions 1 and 2 was found to be statistically similar to 50 human genes in 68% of the codons tested.

We conclude that the genes in regions 1 and 2 have a relatively low G+C content and that they may have been acquired from either an AT-rich host or virus. Thus, these regions appear to be viral counterparts of bacterial pathogenicity islands in the MOCV-1 genome.

This work demonstrates the usefulness of dotplots in the characterization of poxvirus genomes. In addition to highlighting break points for similarity alignments, by "overexposing" the background of random matches they can be used to examine the nucleotide composition of genomes and possibly the evolution of these genomes. Using dotplots in this way would be of particular interest to researchers studying poxvirus evolution because the current methods of creating full genome multiple alignments are prone to errors and are very memory intense.

4.0 Predicted function of the vaccinia virus G5R protein

Published in *Bioinformatics* in 2006 (Da Silva *et al.*, 2006).

4.1 Introduction

Despite having been sequenced over 15 years ago (Goebel *et al.*, 1990a), the vaccinia virus (VACV) strain Copenhagen genome still contains numerous genes that have yet to be assigned a function. Of particular interest are such proteins that are also expressed by all poxviruses since these are likely to play key roles in the replication cycle, one of which is the VACV G5R protein (434 aa). Recently, one of our major research objectives has been to use bioinformatics tools to predict the function of these unknown proteins. Our usual method to identifying the function of these proteins typically starts with a protein database search using PSI-BLAST. Unfortunately, what is often seen in the results of the PSI-BLAST are hits to the other orthologous poxvirus proteins in the database, with hits from other organisms either coming up with insignificant E-values or not being seen at all in the results table. Domain searches of these proteins using methods such as InterProScan are also not helpful since the results of these scans often identify domains entered into the database that were derived from orthologous poxvirus proteins. The lack of meaningful results using these two standard bioinformatics methods spurred us to seek out newer algorithms that could help identify the function of these unknown proteins. One technique which we employed was a Hidden Markov Model (HMM) comparison search tool (HHsearch) (Söding *et al.*, 2005, Söding, 2005) which creates an HMM profile of the query sequence and compares this profile to the HMM profiles of the proteins in several protein structure databases including PDB, SCOP and CATH. Using HHsearch to screen for the functions of these unknown poxvirus proteins,

we found that one of these, G5R, produced a significant hit (96% probability), suggesting that the G5R protein is a homolog of the class of proteins known as the flap endonucleases (FEN-1).

Although little is known about the function G5R plays in the poxvirus life-cycle, it is known to be essential (da Fonseca *et al.*, 2004) and is conserved in all poxvirus species sequenced to date, implying that its role in the poxvirus replication cycle is crucial. Initial characterization of the G5R protein found that G5R is expressed at early stages of poxvirus infection and becomes associated with the viral cores (pellet fraction of solubilized intracellular mature virions), a role that is predominantly associated with proteins expressed at late times of infection although a few early proteins are associated with the viral cores (Banham and Smith, 1992, Doglio *et al.*, 2002b). However, two recent studies that aimed to identify each protein associated with the intracellular mature virion using mass spectrometry were unable to detect G5R within the intracellular mature virion (Yoder *et al.*, 2006, Chung *et al.*, 2006).

Essential in a variety of organisms including human, yeast, bacterial and archaeal species, the FEN-1 protein plays a key role in the removal of RNA primers from the Okazaki fragments during lagging strand DNA synthesis and also plays a role in long-patch base excision repair (Shen *et al.*, 2005). FEN-1 is considered a structure-specific metallonuclease that requires a structured substrate rather than a specific sequence in order to cleave DNA (Shen *et al.*, 2005)(Figure 11) and relies on the binding of two divalent cations for catalysis. FEN-1 acts not only as a 5'-flap endonuclease, but also as a 5' to 3' exonuclease cleaving DNA that contains nicks and gaps (Shen *et al.*, 2005). The crystal structures of human FEN-1 (hFEN-1, PDB ID: 1UL1) (Sakurai *et al.*, 2005),

Archaeoglobus fulgidus (PDB ID: 1RXW) (Chapados *et al.*, 2004), *Methanococcus jannaschii* (PDB ID: 1A77) (Hwang *et al.*, 1998), *Pyrococcus furiosus* (PDB ID: 1B43) (Hosfield *et al.*, 1998) and T5 phage (PDB ID: 1XO1) (Garforth *et al.*, 1999) have been solved and have aided in the characterization of the amino acids important in FEN-1 activity.

Five regions of the FEN-1 protein are important for activity (Figure 12) (Shen *et al.*, 2005). First, the active site that consists of 10 residues (9 of which are charged) binds two divalent cations; one aids in catalysis and the other in protein stabilization (Figure 12, green). Second, the helical clamp region (Figure 12, purple) consists of a helix-loop-helix motif that contains mainly positively charged and hydrophobic residues and binds to the 5'-flap of the DNA substrate (Figure 11). The third region is a hydrophobic wedge (Figure 12, pink) that binds the 3'-flap of the DNA substrate. The fourth and fifth regions contain positively charged residues that stabilize the double-stranded portions of the DNA substrate (Figure 12,red and gold; Figure 11).

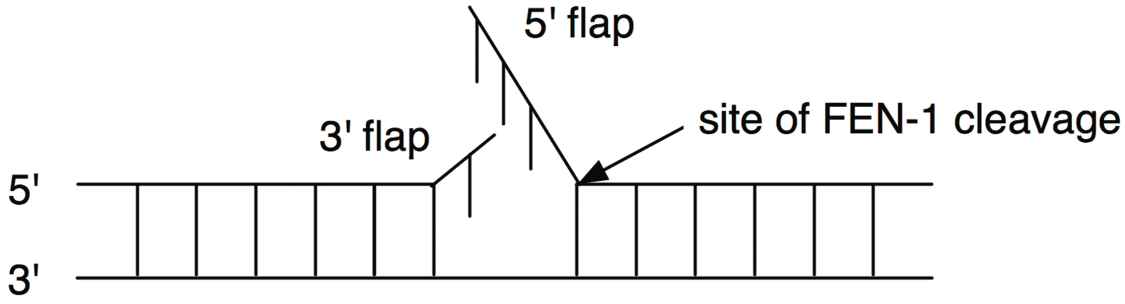


Figure 11. Optimal double-flap DNA substrate used by human FEN-1.

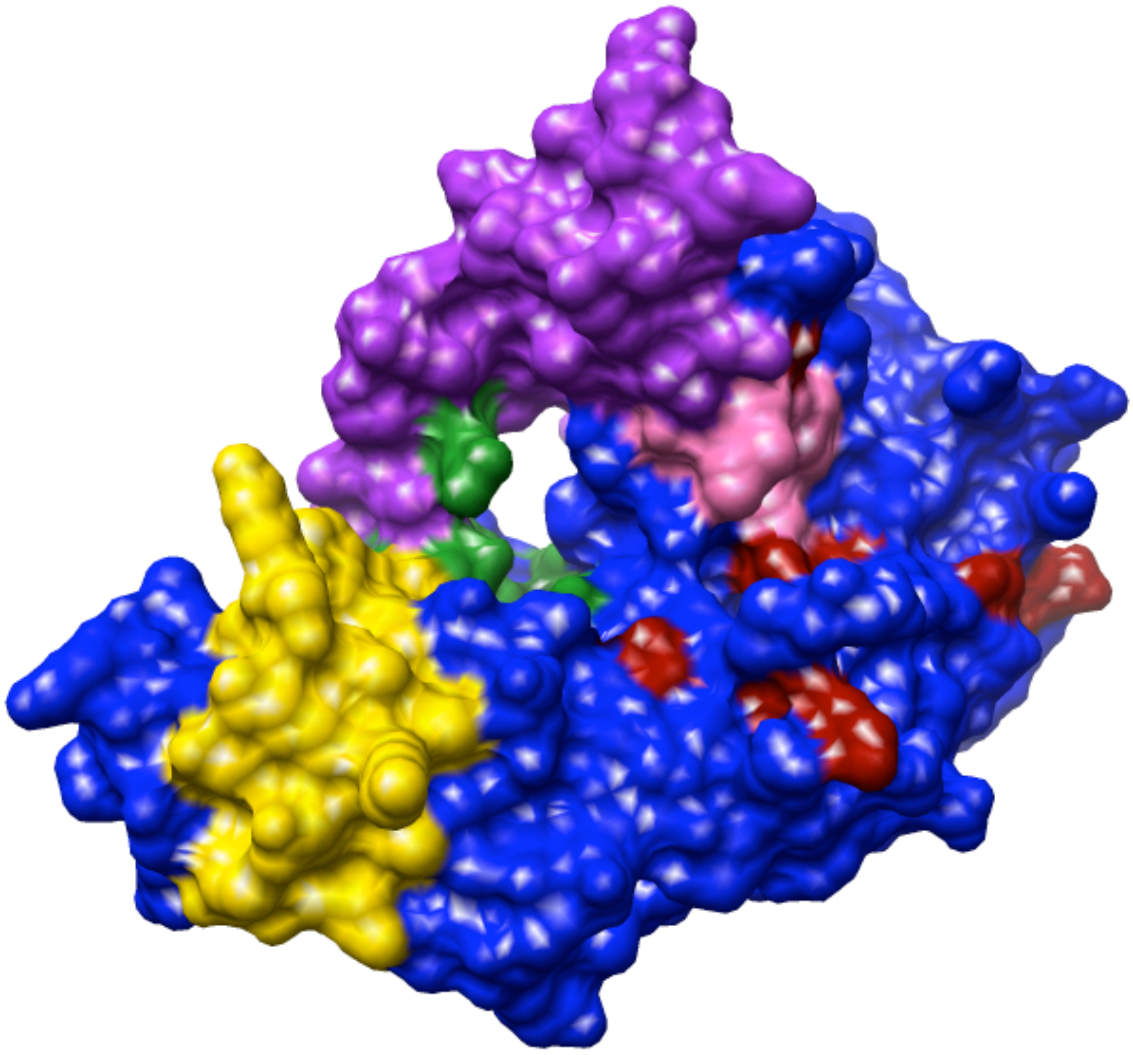


Figure 12. The five important regions of the FEN-1 protein.
Surface diagram of the *A. fulgidus* crystal structure (PDB ID: 1RXW) with the active site coloured in green, the helical clamp region coloured in purple, the hydrophobic wedge coloured in pink, the upstream DNA substrate binding region coloured in red and the downstream DNA substrate binding region coloured in gold.

The results presented in this chapter focus on a series of bioinformatics analyses that support the hypothesis that the VACV G5R protein is structurally similar to the FEN-1 family of nucleases.

4.2 Methods

The initial prediction of G5R function was carried out using a multiple alignment of 16 G5R protein sequences. The alignment was created using the T-Coffee alignment software (Notredame *et al.*, 2000) found in Base-By-Base (Brodie *et al.*, 2004b), with minor adjustments made to the alignment manually. The alignment was used as input for the HHsearch tool (Söding, 2005, Söding *et al.*, 2005).

The VACV G5R protein sequence was used as input in the Robetta protein structure prediction server (Chivian *et al.*, 2003, Rohl *et al.*, 2004, Chivian *et al.*, 2005, Kim *et al.*, 2004). Robetta is a fully automated structure prediction server that uses comparative modeling methods to model protein structure if a related structure exists in the PDB database or *ab initio* prediction methods if no structure exists. Robetta successfully created 5 potential structural models of the VACV G5R protein using the crystal structure of the *A. fulgidus* FEN-1 protein (PDB ID: 1RXV) as the template. Protein structures were superimposed using the MatchMaker feature of the Chimera visualization software (Pettersen *et al.*, 2004). Electrostatic surface images were created with a PyMOL (http://www.pymol.org) plug-in (Lerner, MG, Carlson, HA. APBS plug-in for PyMOL. 2006, unpublished) that utilizes the APBS electrostatic surface evaluation tool (Baker *et al.*, 2001). Electrostatic potential was visualized using PyMOL with positive potential in blue and negative potential in red.

The comparative alignment between the human FEN-1 protein sequence and the VACV G5R and *A. fulgidus* protein sequences was created by performing an initial T-Coffee alignment using default parameters (Notredame *et al.*, 2000) in Base-By-Base (Brodie *et al.*, 2004b) and manually editing the alignment to align

active site residues that were conserved on the 3D model. The secondary structure of the proteins was determined using the ksdssp command of the Chimera program (Pettersen *et al.*, 2004). This command assigns secondary structure elements based on the information found in the PDB file. Truncated versions of the two protein structures (from position 1-331 in G5R and position 1-332 in hFEN-1 and 1-326 in *A. fulgidus*) were used for both the pairwise alignment and the structure comparisons since the C-terminal domains (approximately 100 amino acids) of these proteins are not conserved.

4.3 Results and discussion

4.3.1 HHsearch results

The VACV strain Copenhagen G5R gene is present in all poxviruses sequenced to date and it encodes a 434 amino acid protein although the size of this protein in other poxviruses ranges from 416 amino acids to 505 amino acids with the average size being 435 amino acids. Traditional BLASTp and PSI-BLAST (Altschul *et al.*, 1997, Schäffer *et al.*, 2001) searches provided no clues as to the function of this protein; all significant hits were to other poxvirus G5R protein sequences. Using a diverse set of poxvirus G5R protein sequences that included each unique G5R sequence in the VOCs database (Ehlers *et al.*, 2002), a multiple alignment was created and used as input for the HHsearch program. HHsearch creates an HMM from the input alignment and uses it to search HMMs created from other protein databases including the PDB and SCOP databases. Profile-profile search tools such as HHsearch are thought to be more sensitive since they compare overall profiles of protein sequences rather than comparing a single protein sequence to an alignment of potentially similar proteins (Söding, 2005, Söding *et al.*, 2005). HHsearch showed, with a 96% probability, that G5R was

similar to both the human (E-value: $1.1x10^{-5}$) and *A. fulgidus* (archaeal; E-value: $3.2x10^{-3}$) FEN-1 proteins. This is notable since the human and *A. fulgidus* protein sequences are themselves only 40% identical and the G5R and *A. fulgidus* and human protein sequences are 13% and 16% identical respectively.

A multiple alignment of the VACV G5R, hFEN-1 and *A. fulgidus* proteins (Figure 13) show conservation of essential components of the hFEN-1 active site. The poxvirus G5R orthologs are themselves very diverse; the VACV G5R protein shares only 17% aa identity with the entomopoxvirus proteins, and 33-47% aa identity when compared to lepori, sui, capri, yata or avipoxvirus orthologs. It is therefore significant that the hFEN-1 active site residues are also completely conserved in all the poxvirus proteins suggesting strong functional conservation.

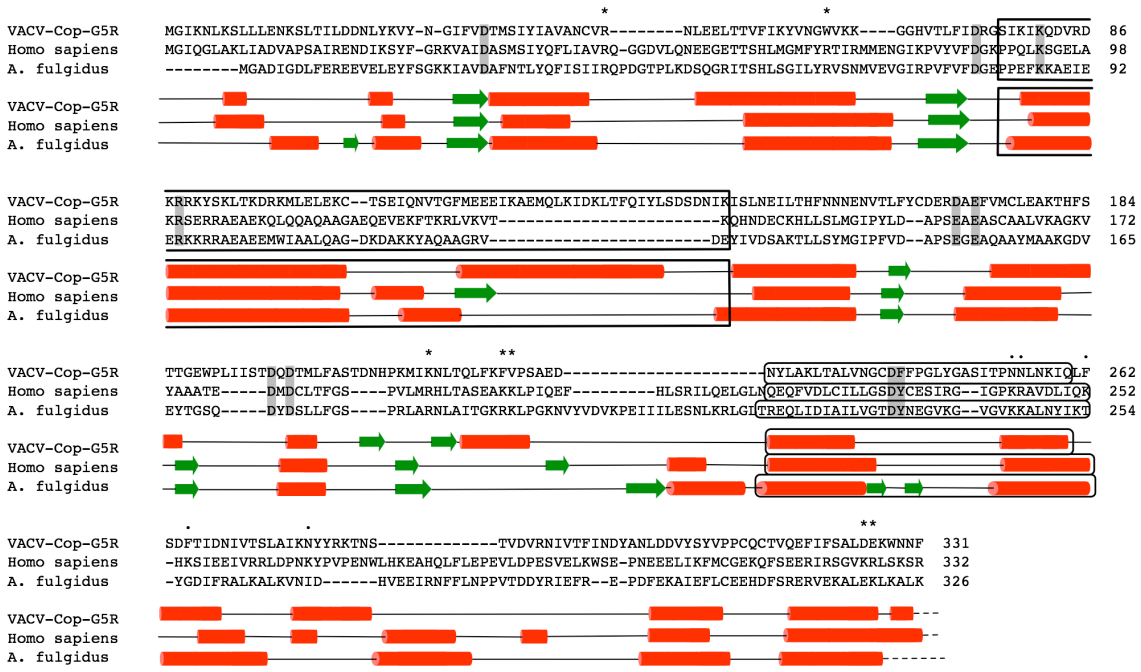


Figure 13. Multiple alignment between vaccinia virus G5R, human FEN-1 and *A. fulgidus* protein sequences.

Alignment created using all defaults of the T-Coffee alignment algorithm (Notredame *et al.*, 2000) with minor adjustments being made by hand. Red cylinders represent α-helices and green arrows represent β-strands that were extracted by ksdssp except in the helical clamp region of G5R and hFEN-1, which was predicted using Jpred. Grey shaded boxes represent active site residues; a black box represents the helical clamp region and a rounded box represents the H3TH motif. Asterisks and dots represent key residues in the upstream DNA binding region and downstream DNA binding region respectively.

4.3.2 Structural modeling of the VACV G5R protein

The significant sequence conservation and availability of FEN-1 crystal structures prompted us to test the following hypothesis: Can the G5R protein adopt a structure similar to FEN-1. We tested this hypothesis by creating a structural model of the VACV G5R protein (Figure 14) using the Robetta protein structure prediction server. The 5 models created with Robetta were very similar; model 4 was chosen as the representative model since it had the most

compact protein structure. Since the *A. fulgidus* and human proteins are distantly related we chose to model the VACV protein on the *A. fulgidus* FEN-1 structure and carry out subsequent analyses with the human structure (hFEN-1) to permit more rigorous comparisons.

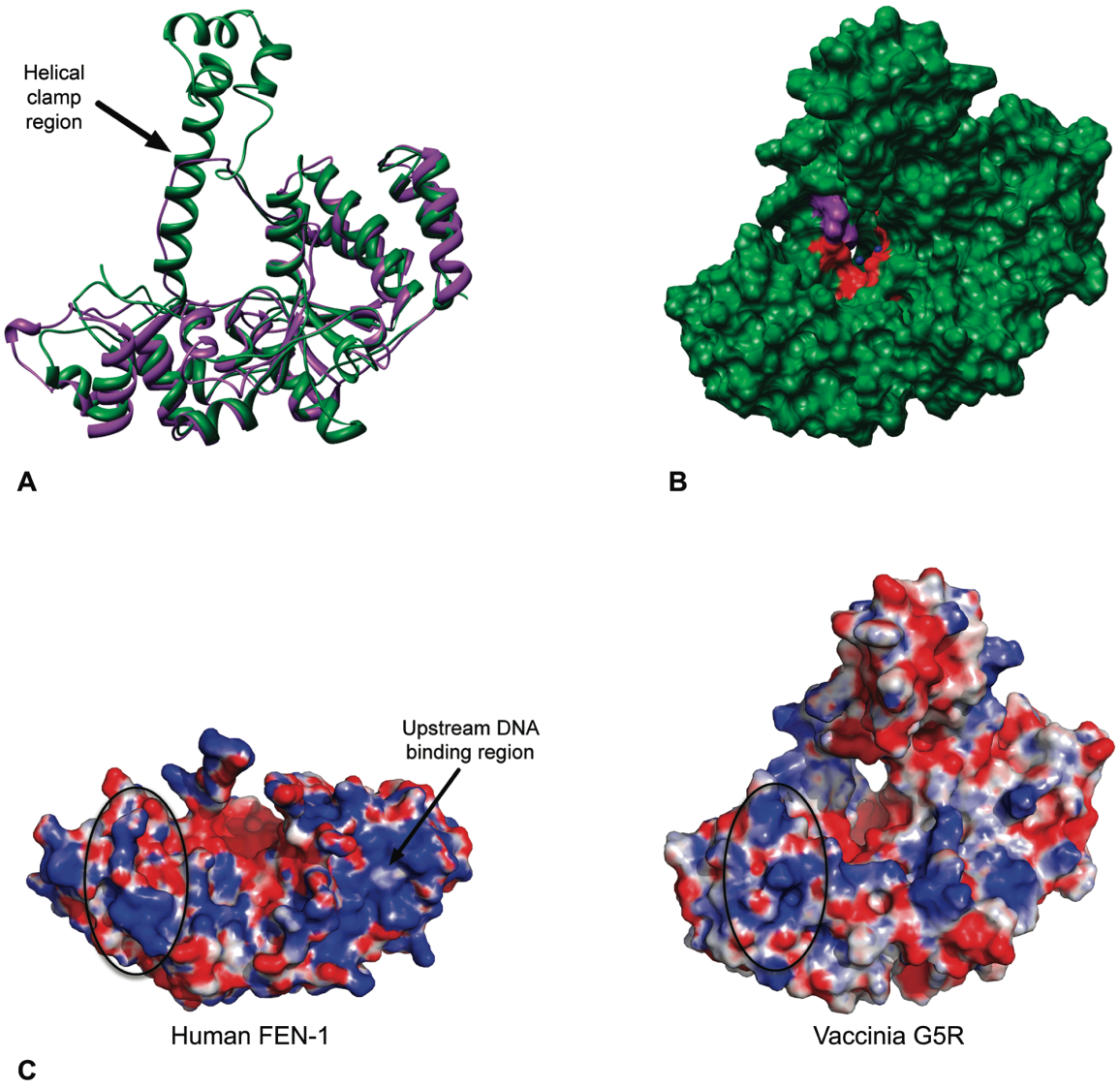


Figure 14. Tertiary structure comparisons between human FEN-1 and vaccinia G5R. (A) Superimposed structures of human FEN-1 (PDB ID: 1UL1, chain X) (purple) and vaccinia G5R (green). (B) Surface diagram of vaccinia virus G5R with negatively charged active site residues in red, positively charged and hydrophobic amino acid residues in purple and two magnesium ions in blue. (C) Surface diagrams of hFEN-1 and vaccinia virus G5R coloured by electrostatic properties as calculated by APBS using default values. Positively charged regions are coloured in blue, negatively charged regions in red. Ovals over both diagrams represent the helix-3turn-helix motif (downstream DNA binding region) that is conserved in both.

4.3.2.1 Secondary and tertiary structure comparison

In addition to the high conservation of active site residues, the secondary structure of the G5R and hFEN-1 proteins is also well conserved. Of the 18 α-helices found in hFEN-1, all but 4 align with similar secondary structure in G5R (Figure 13). Likewise, 4 of the 7 β-strands found in hFEN-1, align very well with G5R. The relatively short length of the β-strands in these proteins may explain this lower modeling success rate. Due to the inability to observe residues in the helical clamp region in the crystal structure of chain X in hFEN-1 (Sakurai *et al.*, 2005), the secondary structure in this region for both G5R and hFEN-1 was predicted using Jpred (Cuff *et al.*, 1998). The secondary structure of this missing region in hFEN-1 as well as the corresponding region in G5R consists of two α-helices and a short β-strand at the end of the clamp region for hFEN-1 and just 2 α-helices for this region in G5R (Figure 13, black box). Archaeal FEN-1 proteins also have helical secondary structure in this region, further supporting the α-helices predicted for the model.

The tertiary structure of hFEN-1 and G5R is also very well conserved. With the exception of the unmapped helical clamp region of hFEN-1, the positioning of the majority of alpha-helices and beta-strands in G5R is conserved and the RMSD of the superposition was found to be 1.4Å over 251 alpha-carbon pairs (Figure 14A). For comparison, the RMSD of the superposition of G5R and the *A. fulgidus* structures is 0.81Å over 281 alpha-carbon pairs and for *A. fulgidus* and hFEN-1 structures is 1.2 Å over 261 alpha-carbon pairs. Thus, the superposition of the G5R protein with human FEN-1 is almost as good as the superposition of the human and archaeal FEN-1 proteins suggesting that the structural model of the G5R protein reflects its true structure in nature. Conservation of the surface

electrostatic properties between VACV G5R and hFEN-1 is shown in Figure 14C; blue areas of hFEN-1 are positively charged and represent regions responsible for contacting substrate DNA (Figure 14C).

4.3.2.2 Important residues in the hFEN-1 active site

Ten amino acids are believed to comprise the active site of hFEN-1, seven are negatively charged and coordinate the binding of two divalent cations, two are positively charged and are thought to contribute to the nucleophilic attack that breaks the phosphate backbone of the substrate DNA and one is hydrophobic and may be involved in hydrogen bonding with other active site residues (Shen *et al.*, 2005). Six of the seven negatively charged residues are fully conserved in G5R (Figure 13 and Figure 14B); glutamate 158 and tyrosine 234 in hFEN-1 are replaced by aspartate and phenylalanine, respectively, in the corresponding positions of G5R. Both substitutions maintain the biochemical properties of the amino acids found in FEN-1 and therefore are likely to be functional in G5R. The two positively charged residues that are key for FEN-1 activity are lysine 93 and arginine 100 (Shen *et al.*, 2005). These two residues are absolutely conserved in all poxvirus G5R proteins and are also positionally conserved on the structural model of G5R compared to hFEN-1.

4.3.2.3 Regions important in substrate binding

There are 4 regions that are important in binding the substrate DNA in hFEN-1. The first is the region important in binding the 5'-flap of substrate DNA, which is located at the helical clamp region and contains the Helix-Loop-Helix (HLH) motif that binds ssDNA (Shen *et al.*, 2005). The G5R structural model has 4 α-helices in this region that when combined into two sets of two helices separated by a six amino acid loop, closely resembles an HLH motif (Figure

14A). When the secondary structure of G5R was predicted for this region using Jpred, the HLH motif was found to be conserved and consists of 2 α-helices separated by a 9 amino acid loop (Figure 13). The helical clamp region of hFEN-1 consists of bulky and positively charged amino acid residues on the inner side of each of the helices of the HLH motif that are responsible for contacting the 5'-flap DNA (Shen *et al.*, 2005). Of the 70 amino acids comprising the helical clamp region in G5R, 20 of them (28.6%) are positively charged or hydrophobic compared to hFEN-1, which has 12 of 47 amino acids (25.5%) being positively charged or hydrophobic in the helical clamp region.

The binding of the 3'-flap of the substrate DNA occurs at the hydrophobic wedge in hFEN-1 (Shen *et al.*, 2005). This region begins at position 42 of hFEN-1 and consists of the sequence: FLIAV (Figure 13). The corresponding region in G5R, although not identical to that of hFEN-1 begins at position 41, on both the structural model (Figure 15) and the alignment (Figure 13), and consists of the sequence: VANCV. This sequence in G5R is primarily hydrophobic suggesting that it may also be capable of binding to a 3'-flap DNA region.

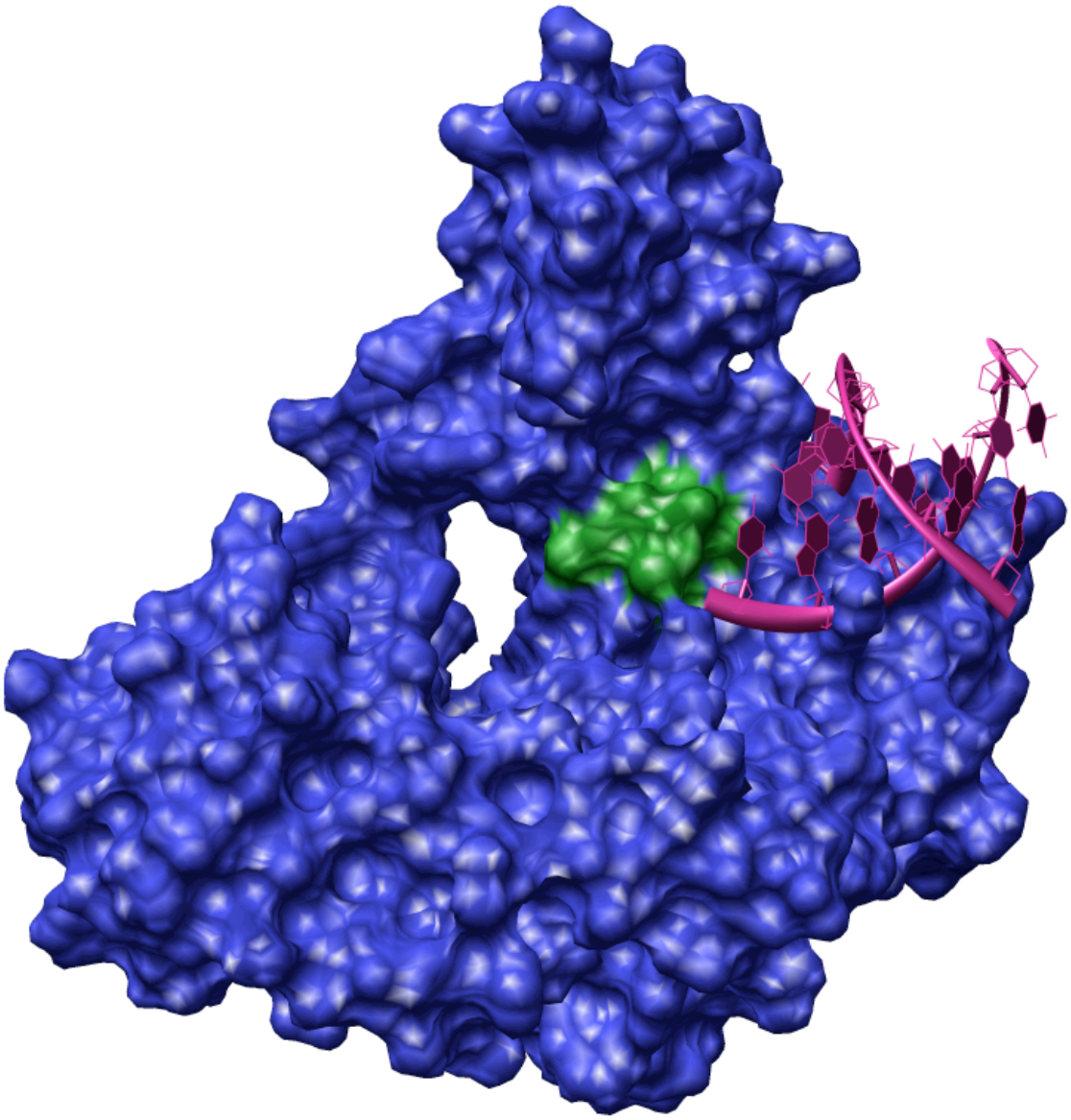


Figure 15. The hydrophobic wedge region of the G5R structural model.
Surface diagram of the VACV G5R protein structural model (blue) with the hydrophobic wedge region coloured in green. A short segment of substrate DNA showing the single 3′-flap nucleotide and a region of duplex DNA is depicted in pink.

The remaining two regions are responsible for binding the upstream and downstream double-stranded substrate DNA. Each of these regions contains key positively charged residues that are responsible for making contact with the substrate DNA. The upstream DNA binding region includes the following

residues of hFEN-1: R47, R70, R192, K200, K201, K326, and R327. Of these seven residues, one (R47 in hFEN-1) is conserved in the G5R protein (R46 in G5R). The alignment shown in Figure 13 shows the poor conservation of these residues, however it is important to note that this alignment was manually edited to ensure no gaps were located within the alpha-helices and beta-strands and can be further improved upon. For example, Figure 16 B shows an unedited, T-Coffee alignment of the region spanning amino acids 179 and 207 of the human FEN-1 protein compared to the same region aligned according to Figure 13 (Figure 16A), which includes three of the seven important upstream DNA substrate binding residues. Only one of these three amino acids is conserved in the original alignment (Figure 13 and Figure 16A) however, once this region is realigned, there are now two conserved amino acids (Figure 16B). Electrostatic surface diagrams were subsequently used to determine the extent of positively charged residues in this region and they showed that the corresponding upstream DNA binding region of the G5R protein (Figure 14C) consists of the positively charged residues that could contribute to DNA binding. Thus, despite the apparent lack of sequence matching in the alignment seen in Figure 13, these regions may be in fact well conserved functionally.

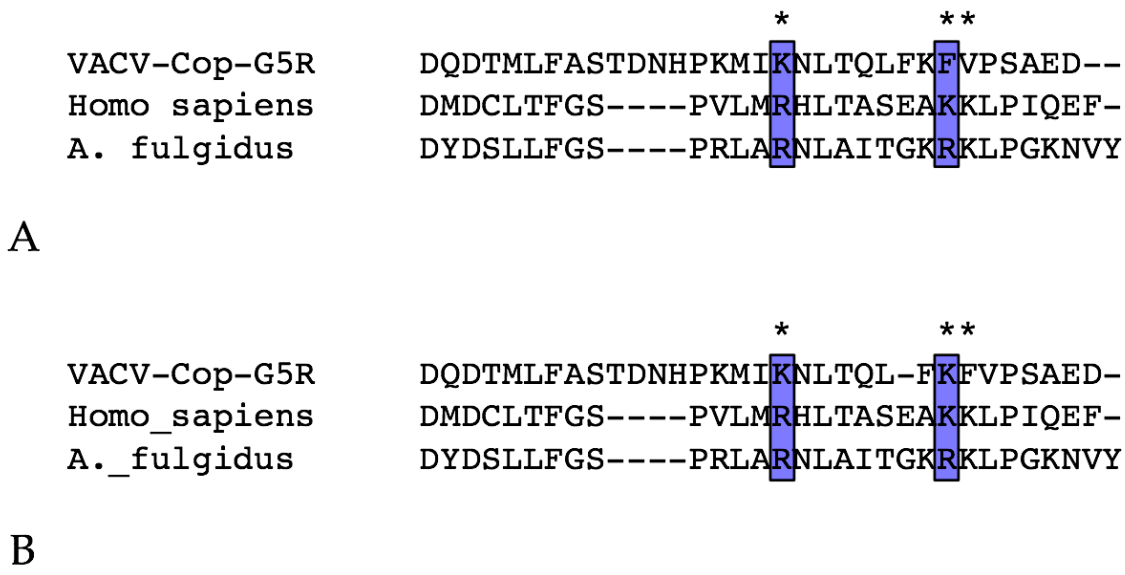


Figure 16. Comparison between two alignments of the same upstream DNA binding region.

(A) Original alignment of the region between amino acids 179 and 207 of the human FEN-1 protein as seen in Figure 13. (B) Unedited T-Coffee alignment of the region shown in panel A. Asterisks above both alignments show the three positively charged resides that are important in upstream DNA substrate binding. Blue boxes show the two amino acids that are important for upstream DNA binding and are conserved between the G5R and hFEN-1 proteins.

Analysis of the downstream DNA substrate binding regions of the 2 structures revealed some regions of conservation and some differences between the proteins. This region on hFEN-1 consists of 5 important positively charged amino acids as well as a Helix-3Turn-Helix (H3TH) motif. The five residues that are involved in hFEN-1 downstream DNA binding are: K244, R245, K252, K254, and K267. These residues appear to correspond to N254, N255, I259, L261, and V271 respectively, on the G5R model suggesting a loss of positive charge. However, the H3TH region of hFEN-1 is highly conserved in the G5R structure (Figure 13, rounded boxes; Figure 14C). On hFEN-1, the first helix of the H3TH motif is located at Q220 and ends at G231, which corresponds to the G5R helix

that begins at N228 and ends at V237. The 3T region of hFEN-1 is slightly shorter than the G5R 3T region beginning at S232 and ending at G242 and corresponds to positions N238 through T252 in G5R, respectively. The second helix of G5R is slightly shorter than that seen for hFEN-1 with the hFEN-1 helix beginning at P243 and ending at K252 and corresponding to P253-Q260 respectively for G5R.

4.3.2.4 Other features that may play a role in G5R activity

There are 2 other features that have been observed in hFEN-1 that may play a role in G5R activity. The first is the post-translational phosphorylation of S187 in hFEN-1, which has been shown to play a role in the regulation of exo/endonuclease activity in hFEN-1 (Shen *et al.*, 2005). The corresponding residue in G5R is S204 and is conserved in both the alignment and the structural model.

Not only are the N-terminal sequences ending at position 10 of the G5R and FEN-1 proteins very highly conserved among their respective groups, but also that there is significant conservation between the two sets of proteins (Figure 13). Although the exact function of this short sequence is not known, the crystal structure of hFEN-1 revealed that it is in close proximity to the active site and H3TH residues and thus may stabilize interactions with the substrate (Sakurai *et al.*, 2005).

Finally, there are 61 identical amino acids that are found in all poxviruses (except the entomopoxviruses) and may contribute to the function of the G5R protein. Of these 61 amino acids, 15 are also identical in the hFEN-1 protein and of those 15 that are identical in hFEN-1, seven are part of the active site of the protein. Thus, eight amino acids are identical between hFEN-1 and all poxviruses

but have yet to be assigned a functional role in the activity of the protein. This leaves 46 amino acids that are not conserved in the hFEN-1 protein but are identical in all poxviruses suggesting that these amino acids play some kind of functional role in the G5R protein and are excellent candidates for future mutational studies.

4.4 Conclusions

Given the multiple lines of evidence presented here, we conclude that the poxvirus G5R proteins are likely to share significant structural similarity to the FEN-1 family. Herpes (Oroskar and Read, 1989, Taddeo *et al.*, 2002) and iridoviruses also encode related proteins, and these may represent very ancient viral genes that have evolved from a FEN-1-like ancestor. However, it is not clear whether endonuclease activity has been preserved and given the wide range of activities described for FEN-1 proteins including its participation in the base excision repair pathway, genome replication and probable participation in apoptosis (Shen *et al.*, 2005) it is difficult to predict the natural substrates for the viral proteins.

Early experiments have shown that poxvirus DNA replication may involve lagging strand DNA synthesis (Pogo and O'Shea, 1978), and a recent study has found that the vaccinia virus D5R protein is capable of synthesizing RNA primers suggesting that poxvirus DNA replication likely does involve lagging strand DNA synthesis (De Silva *et al.*, 2007). The likelihood that lagging strand synthesis is occurring in poxviruses suggests that the G5R protein may play a role similar to that of hFEN-1 in RNA primer removal during DNA replication. It is important to note, however, that since the human and vaccinia proteins are so divergent from one another, the G5R protein may have evolved novel

functions in poxviruses in a manner similar to the uracil DNA glycosylase (De Silva and Moss, 2003). It is likely that the C-terminal domain, which appears to be unique to the poxvirus proteins, plays a role in its specific function.

5.0 VACV-G8R: A comparison to human proliferating cell nuclear antigen

5.1 Introduction

Perhaps one of the most important features of FEN-1 is that it cleaves DNA more efficiently when bound to proliferating cell nuclear antigen (PCNA) since the PCNA protein is able to freely slide along the DNA molecule bringing the FEN-1 protein into contact with 5'-flap regions, which the FEN-1 protein can then cleave (Li *et al.*, 1995). In fact, previous studies have shown that when FEN-1 interacts with PCNA, its activity increases between 10- and 50-fold (Li *et al.*, 1995, Tom *et al.*, 2000). The PCNA protein is classified as a sliding clamp and is found in all organisms including prokaryotes, where it is called a beta-sliding clamp (Bruck and O'Donnell, 2001) and in the case of eukaryotes functions as a homotrimer, wrapping itself around DNA and leaving its outer edge for interactions with other proteins (clients) (Kelman, 1997). The client proteins with which PCNA interacts can be grouped into three categories: DNA replication, DNA repair and cell cycle regulation (Tsurimoto, 1998, Warbrick, 2000). DNA repair client proteins include FEN-1, Xeroderma pigmentosa G (XPG) protein and uracil DNA glycosylase (Warbrick, 2000). The XPG protein belongs to the same nuclease superfamily as FEN-1 since it is also a structure-specific endonuclease and plays a role in nucleotide excision repair by cleaving ultra violet light damaged DNA at the 3'-end of the DNA repair bubble structure (Nishino *et al.*, 2006). This function should not be confused with the function of FEN-1 in long-patch base excision repair, a different DNA repair pathway, where FEN-1 cleaves the resulting 5'-flap from a repaired piece of DNA in response to damage to DNA caused by oxidation, methylation or the spontaneous removal

of the DNA base (Wood, 1996). DNA replication proteins include DNA polymerase delta; FEN-1; and the clamp loading complex replication factor C (RFC), which functions to assemble the three subunits of PCNA around the DNA to be replicated (Warbrick, 2000).

Possibly the more intriguing function of PCNA is its role in cell cycle regulation, which is somewhat related to its function in DNA repair since it involves a cell-cycle regulating protein (p21) which inhibits two cyclin-dependent kinases in response to DNA damage (Warbrick, 2000, Tom *et al.*, 2000). A peptide derived from the C-terminus of the p21 protein has been co-crystallized with human PCNA and shows its extensive interactions with the PCNA protein (Warbrick, 2000). Although the complete PCNA~p21 complex has not been co-crystallized, mainly due to the multiple functional domains of the p21 protein, PCNA and p21 have been shown to co-immunoprecipitate suggesting that these two proteins do interact with each other (Warbrick, 2000, Tsurimoto, 1998). By interacting with PCNA, p21 blocks the DNA polymerase complex from associating with PCNA and effectively shuts off DNA replication (Warbrick, 2000). This, in turn, also indirectly prevents long-patch base excision repair, which requires a functioning DNA polymerase complex to function correctly, however, short-patch DNA repair can still occur in response to DNA damage because it does not require a functioning DNA polymerase complex (Warbrick, 2000). It has also been speculated that during apoptosis, when caspase-3 cleaves p21 and other proteins with a PCNA binding motif, the resulting peptide can also bind to PCNA preventing it from binding to the DNA polymerase complex and thus pushing the cell towards apoptosis (Tsurimoto, 1998).

Structurally, the PCNA trimer has pseudo-sixfold symmetry with two equivalent structural domains per subunit (Figure 20). There are two regions of PCNA that are important for its functionality. The first is the DNA binding region that consists of 4 alpha helices per subunit for a total of 12 alpha helices per trimer. The trimer forms a circle similar to a donut in which the center of the "donut" represents the region where PCNA comes into contact with DNA and is where the alpha helices are located. The second region is the client-binding region, which consists of a long loop called the interdomain connecting loop. This loop serves two purposes, first in connecting the two identical structural domains to each other per subunit and the second is client binding. Each client protein that interacts with PCNA has a PCNA interacting protein (PIP) motif, which consists of the consensus sequence: Qxx[M/L/I]xxF[Y/F] where x represents any amino acid. The ability of PCNA to slide along DNA and transiently bind to a multitude of client proteins, has, until recently, been limited to clients that functioned in DNA replication and repair. Given this strong interaction between PCNA and DNA, it is not surprising that it has now been reported that PCNA plays a role in transcription by binding to transcription factors that regulate gene expression (Schultz-Norton *et al.*, 2007).

Knowing that FEN-1 requires PCNA to function efficiently, we hypothesized that vaccinia virus encoded a PCNA-like protein. To try and find this poxvirus equivalent PCNA protein, we used an InterProScan of the vaccinia virus genome, using JiPS (Syed and Upton, 2006), to try and identify any potential sliding-clamp proteins. One protein listed as a sliding clamp was the vaccinia virus G8R protein and upon further inspection, the class of sliding clamp that G8R was listed as being was one that contained the PCNA protein. Further

HMM comparison searches using HHsearch further showed that the G8R protein was similar to PCNA. This chapter presents the results of the bioinformatics comparisons made between the vaccinia virus G8R structural model and human PCNA, which is currently a work in progress.

5.2 Methods

The initial vaccinia genome-wide InterProScan (Zdobnov and Apweiler, 2001) was performed using the JIPS program developed in our laboratory (Syed and Upton, 2006). Results of the scan were manually parsed looking for the keywords "sliding clamp". The one protein (VACV G8R) to fit this criteria was examined in further detail at the InterProScan website (http://www.ebi.ac.uk/InterProScan/).

A T-Coffee alignment of 16 poxvirus G8R sequences was used as input in the HHsearch HMM comparison search tool and the VACV G8R sequence was used as input in the Robetta protein structure prediction server (Chivian *et al.*, 2003, Chivian *et al.*, 2005, Kim *et al.*, 2004, Rohl *et al.*, 2004). Robetta successfully created 5 potential structural models of the VACV G8R protein using the crystal structure of yeast PCNA (PDB ID: 1SXJ) (Bowman *et al.*, 2004) as a template. Comparisons were subsequently made to the crystal structure of human PCNA (PDB ID: 1UL1) (Sakurai *et al.*, 2004). Protein structures were visualized and superimposed using the Chimera visualization software (Pettersen *et al.*, 2004). Electrostatic surface images were created with the APBS plug-in for PyMOL (http://www.pymol.org) with positive potential coloured in blue and negative potential in red.

The comparative alignment between VACV G8R, human and yeast PCNA protein sequences was created using T-Coffee with manual manipulations to

ensure that the secondary structure that was being mapped contained no gaps. The secondary structure displayed with the alignment was derived directly from the crystal structures of the human and yeast PCNA proteins and from the model of the VACV G8R protein.

5.3 Results and discussion

The VACV G8R gene encodes a 260 amino acid protein that is found in all chordopoxviruses (everything except entomopoxviruses), expressed at intermediate stages of infection, and is thought to play a part in poxvirus late transcription (Zhang *et al.*, 1992b, Dellis *et al.*, 2004, Wright and Coroneos, 1993). In searching the vaccinia virus genome, using InterProScan, for a protein that may be similar to human PCNA, we found only 1 protein (G8R) that fit the criteria of a sliding clamp. The InterProScan results for the G8R protein shows a classification into two different categories. The first is as a viral trans-activator protein (InterProScan ID: IPR005022), which is not surprising given the role that the G8R protein plays in late transcription. The second category in which the G8R protein was classified using InterProScan was the DNA clamp category, which links to the DNA clamp superfamily in the SCOP (structural classification of proteins) database (SCOP ID: 55979). This superfamily is subdivided into two families, the DNA polymerase III beta subunit family which contains multiple domains from the beta subunit of the DNA polymerase III protein in *E. coli,* and the DNA polymerase processivity factor family which contains sliding clamps from several organisms including herpes virus, human, yeast and 3 archaeal species. The InterProScan did not assign the G8R protein to a particular family.

In order to garner some extra information about the classification of the G8R protein, we created a multiple alignment of the protein sequences of 16 diverse

G8R orthologs using T-Coffee and used this alignment as input for the HHsearch tool (Söding, 2005, Söding *et al.*, 2005). The HHsearch tool resulted in a top hit being to the yeast PCNA protein with a probability of 95.5% and an E-value of 10. Identical results were seen when just the VACV G8R protein sequence was used as input for HHsearch. It is important to note that although the E-value for the top hit between the G8R protein and yeast PCNA is poor (high), it cannot be interpreted in the exact manner as an E-value would be interpreted when performing a BLASTp or PSI-BLAST. The authors of the HHsearch tool state that since their E-value calculation does not take into account conserved secondary structural elements between the query and the hit, that the probability measure is a more accurate statistic (Söding, 2005, Söding *et al.*, 2005). The authors also recommend that results that have a high probability yet a poor E-value warrant further investigation and as such, since the top hit for the G8R protein was to a yeast PCNA protein with a complete structure in the PDB database, we decided to model the structure of the G8R protein using the Robetta protein structure prediction server as was performed for the G5R protein (Chapter 4, this dissertation).

5.3.1 Secondary and tertiary structure conservation

The T-Coffee alignment of the G8R, human and yeast PCNA proteins (Figure 17) shows 35% sequence identity between human and yeast PCNA and only 9% and 12% sequence identity between the G8R protein and the human and yeast proteins, respectively. Despite the relatively low sequence identity between the G8R protein and these eukaryotic proteins, comparison of the secondary structure of the G8R protein with the secondary structures of the human and yeast PCNA proteins reveals strong conservation between all three proteins

(Figure 17). It is known that the human and yeast PCNA proteins have a pseudo sixfold symmetry which means that each subunit has two key domains, which in the case of PCNA, are divided by the interdomain connecting loop (Sakurai *et al.*, 2004). For all three proteins (G8R, human and yeast PCNA), the first domain has the following secondary structure associated with it: SHSSSSSHSSS where S is a beta-sheet and H is an alpha-helix. For the G8R protein and the human PCNA protein, the second domain, although it is, by definition of the pseudo-sixfold axis of symmetry, supposed to be identical in secondary structure to the first domain, is missing a beta-sheet and has the following secondary structure associated with it: SHSSSSHSSS. The second domain in the yeast PCNA protein has an identical secondary structure pattern as the first domain and is as follows: SHSSSSSHSSS. The interdomain connecting loop, shown in Figure 17 surrounded by a rounded box, consists of 18 amino acids for the G8R protein and human PCNA, and 17 amino acids for yeast PCNA.

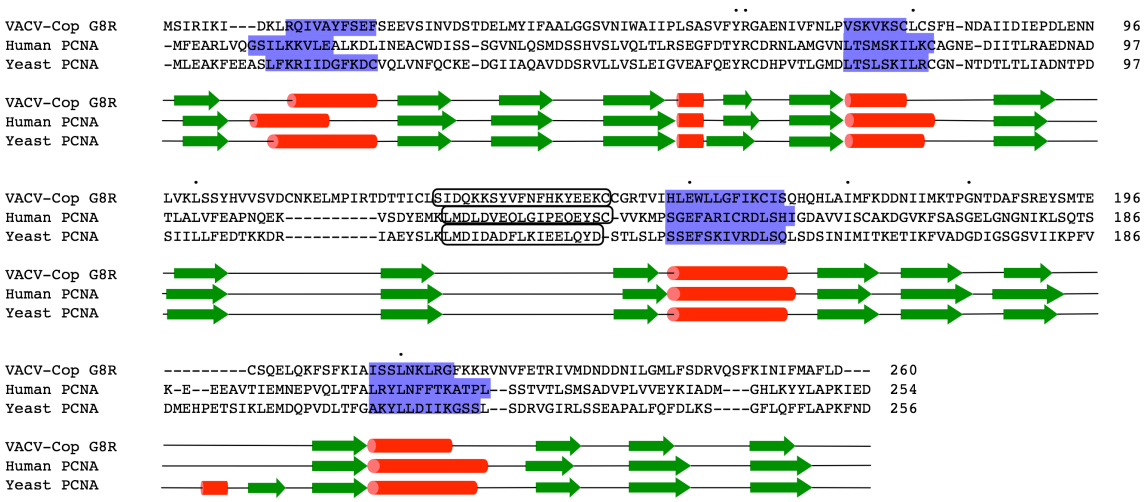


Figure 17. Comparative alignment between VACV G8R, human and yeast PCNA. Alignment created using all default values of the T-Coffee alignment algorithm with minor adjustments made by hand (Notredame *et al.*, 2000). Green arrows represent beta-sheets, and red cylinders represent alpha-helices. Black rounded box shows the interdomain connecting loop, blue shaded boxes represent the helices that contribute to DNA binding and dots represent identical residues between the 3 protein sequences.

Using the Robetta protein structure prediction server, the complete G8R protein sequence was successfully modeled using the yeast PCNA crystal structure (PDB ID: 1PLQ) as a template. As with the secondary structure, the tertiary structure of the structural model of the G8R protein is highly conserved when compared to the crystal structures of both human and yeast PCNA (Figure 18). The RMSD for the superposition of the human and yeast crystal structures is 1.24Å over 249 alpha-carbon pairs, and in comparison, the RMSD for the superposition of G8R and human PCNA is 1.45Å over 228 alpha-carbon pairs and is 0.74Å over 226 alpha-carbon pairs when superimposing the G8R structural model over the yeast PCNA crystal structure. The RMSD values show that despite the low sequence identity between the G8R and human and yeast proteins, the structural model of the G8R protein is as similar to the human

PCNA crystal structure as the yeast PCNA structure is to the human PCNA structure. Not surprising is the lower RMSD value between the yeast PCNA crystal structure and the G8R model since the G8R structure was modeled using the yeast crystal structure as a template. Since PCNA acts as an intermediary between DNA and many diverse client proteins and because of this does not have a defined active site, analyzing its structure and that of the G8R protein is very important in determining interactions between PCNA, its clients and DNA (Section 5.3.2).

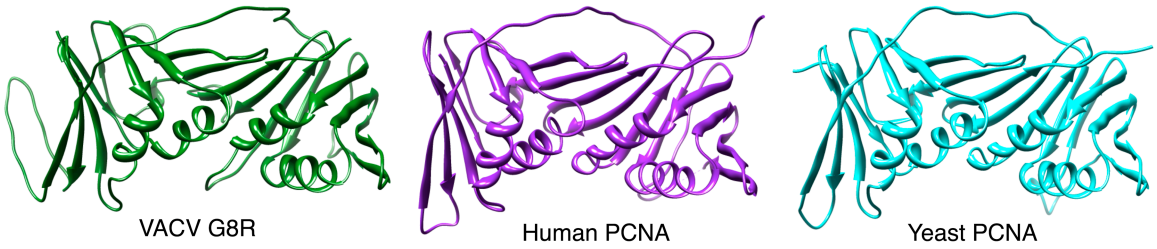


Figure 18. Structures of the VACV G8R, human and yeast PCNA proteins.

There are three major differences that are observed upon further comparison of the human PCNA and VACV G8R structures. The first is the overall shape of the PCNA subunit, which with human PCNA, is curved into a shape resembling a semi-circle, but is somewhat stretched horizontally in the G8R structural model (Figure 19). The main reason for this is how the loop regions on the G8R model are placed in 3-dimensional space. The positioning of the helices and the beta-sheets is nearly identical between both but the loop regions exhibit enough variability to change the overall shape of the G8R molecule into something that is more like a flattened semi-circle. The second major difference is the loop region located between beta-sheets 7 and 8 (Figure 19). In the G8R protein, this

loop is 16 amino acids long and extends out towards the back of the protein whereas in human PCNA this loop is considerably shorter at only 6 amino acids and it extends downwards (Figure 19). The potential role that this extended loop plays in the G8R protein is unclear and it is not possible to predict whether it plays a role in promoting or preventing a trimerization of the three G8R subunits to create a complete protein complex. However, given that this loop region is conserved in all chordopoxviruses with minor amino acid differences seen primarily with the avipoxviruses and the molluscum contagiosum virus, compared to the VACV G8R protein, it is likely to play an important role in the function of the G8R protein. The other major difference is the loop region between alpha-helix 4 and beta-sheet 15, which is 9 amino acids long in the G8R protein and is only a short 2 amino acid loop in human PCNA (Figure 19). Since this loop extends into the DNA binding region of the G8R protein, it is possible that it plays a role in stabilizing the interaction between DNA and the G8R protein.

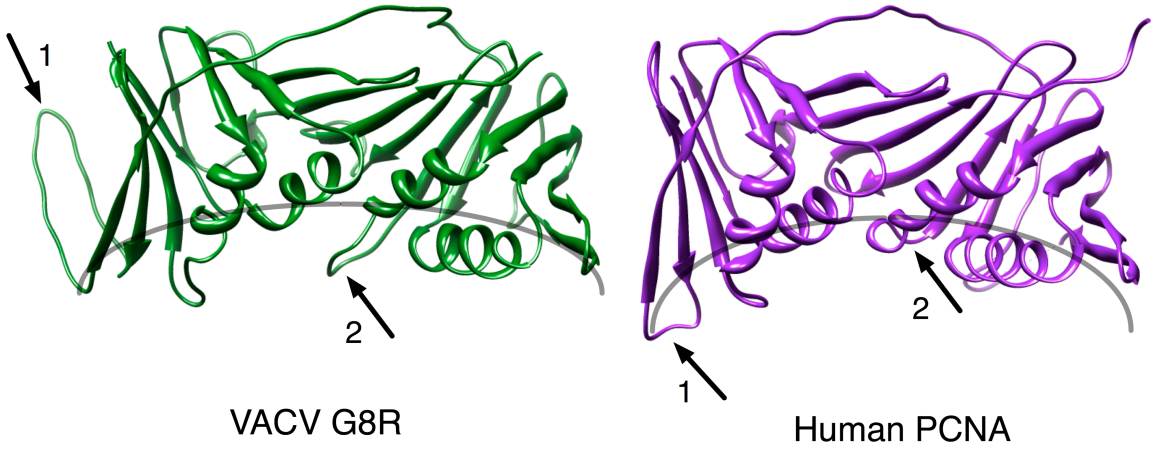


Figure 19. Three major differences between the VACV G8R and human PCNA structures.
Arrows labelled 1 mark the loop region between beta-sheets 7 and 8 PCNA; arrows labelled 2 mark the loop region between alpha-helix 4 and beta-sheet 15; opaque black semi-circular lines mark the overall shape of the VACV G8R and human PCNA structures.

As was alluded to in the previous paragraph, the human PCNA protein functions as a homotrimer, which forms the shape of a donut around the DNA that fits in the centre. Using three copies of the VACV G8R structural model, we attempted to model the PCNA homotrimer (Figure 20). The hypothetical G8R trimer looks strikingly similar to the human PCNA trimer and it appears as though the extended loop region seen between beta-sheets 7 and 8 on the VACV G8R protein is not in a position to be inhibiting the formation of the trimer (Figure 20). Interestingly, the idea that the G8R protein could be forming a trimer with itself is not unwarranted, since a recent yeast two hybrid assay showed that the VACV G8R protein has a strong interaction with itself (Dellis *et al.*, 2004).

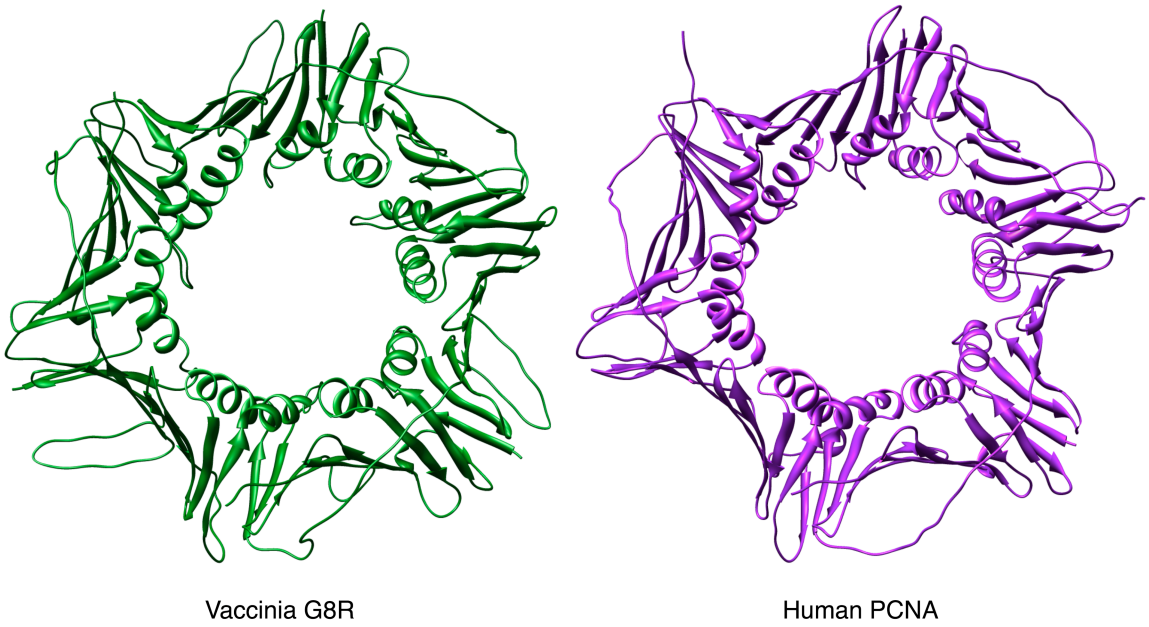


Figure 20. Complete trimer of the VACV G8R and human PCNA protein structures.

5.3.2 Functional domain comparison

Since the role of the PCNA protein is to act as an intermediary between certain client proteins and the DNA genome, it is not considered an enzyme and thus does not have an active site. It does, however, have two functional domains that are important in both DNA and client binding. The first functional domain is the client binding domain, which consists of the interdomain connecting loop on the PCNA protein. The amino acids comprising this long loop interact with client proteins that contain the following PIP (PCNA interacting protein) motif: Qxx[M/L/I]xxF[Y/F] where x is any amino acid. In the case of the human FEN-1 client protein, this motif is located at the C-terminus of the protein, although in some client proteins, the PIP motif may be located at the N-terminus or even near the middle of the protein exposed on its surface (Doré *et al.*, 2006). In the case of the VACV G5R protein, which may be a FEN-1-like protein, no motif was identified that was identical to the motif described above, but a similar motif

QDTMLF was found in the active site of the G5R protein. Obviously given that this potential motif is located in the active site of the G5R protein and is not well exposed, it likely does not associate with the interdomain connecting loop of the G8R protein. The lack of a PIP motif in the G5R protein does not necessarily mean that the G8R and G5R proteins do not interact, as it is possible that through the evolution of these proteins in the poxvirus, their interacting sites may have changed and only further biochemical experiments will be able to determine the extent of this interaction.

The other important functional domain in the PCNA protein is the DNA binding domain. This domain consists of the 4 alpha helices of the protein that cluster towards the center of the trimer, forming a hole into which the DNA fits (Figure 17 blue shaded boxes, and Figure 20). Since these helices are capable of binding to DNA, they must contain positively charged amino acids that can electrostatically interact with the negatively charged DNA molecule and help stabilize it in the centre of the protein. A simple count of the number of basic, positively charged amino acids (lysine, histidine and arginine) contained in the 4 helices of a set of diverse PCNA proteins (Table 9) shows that the eukaryotic and archaeal PCNA proteins have 9 positively charged amino acids in their helical regions whereas the VACV G8R protein has only 7. Although it may appear that the VACV G8R protein has a somewhat lower number of positively charged amino acids in this region, given that the zebrafish and silkworm PCNA proteins also have 7 positively charged amino acids in this region, the G8R count is consistent with other species. To get a better idea of the locations of these and other residues that contribute to overall electrostatic charge, the electrostatic potential of the human, yeast, archaeal and G8R proteins was calculated and

mapped onto surface diagrams of each protein (Figure 21). The electrostatic surface diagrams reveal that despite the VACV G8R having the lowest number of positively charged amino acids in its helical region, the electrostatics in the helical region is similar to what is seen for the other 3 PCNA proteins. The positively charged residues tend to cluster primarily near the centre of the helical region for the G8R protein and this clustering towards the centre of the protein is also seen for the archaeal PCNA protein (Figure 21). The human and yeast proteins both have some positively charged residues near the centre of the helical region, but overall tend to have a more spread-out charge pattern compared to the archaeal and G8R proteins. Given that the spread of positive charge observed in the G8R protein is consistent with that seen with the other three proteins, it is quite possible that the G8R protein is capable of binding to DNA although more biochemical evidence is required to show this. Given the role that the G8R protein plays in poxvirus late transcription, it is perhaps not surprising that it may be able to bind DNA, but whether it plays a role solely in transcription or whether it plays a role similar to that of PCNA remains to be determined.

Table 9. Number of positively charged amino acids in the helical regions of various PCNA proteins.

Organism	Number of positively charged amino acids
Bacteriophage T4	4
VACV G8R	7
Zebrafish (*Danio rerio*)	7
Silkworm (*Bombyx Mori*)	7
Frog (*Xenopus laevis)*	8
Plasmodium falciparum	8
Human (*Homo sapiens)*	9
Yeast (*Saccharomyces cerevisiae*)	9
Archaeoglobus fulgidus (Archaeal)	9
Pyrococcus furiosus (Archaeal)	9
Rat (*Rattus rattus*)	9
Chicken (*Gallus gallus*)	9
Mouse (*Mus musculus*)	9

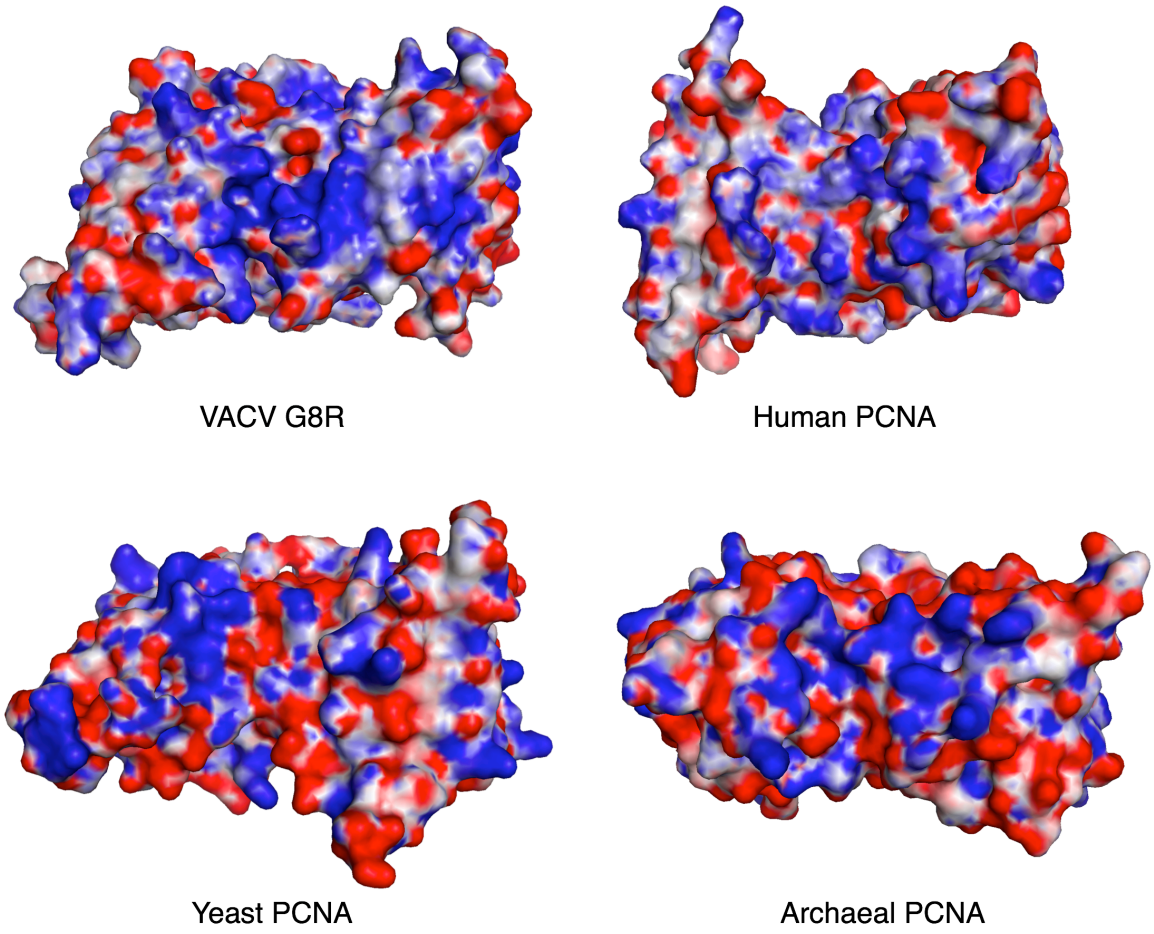


Figure 21. Electrostatic surface diagrams of the VACV G8R protein and 3 other PCNA proteins.

Surface diagrams are coloured by electrostatic properties as calculated by APBS using all default values with positively charged regions coloured in blue and negatively charged regions in red.

5.4 Conclusions

The work presented in this chapter has shown that the VACV G8R protein, despite having low sequence identity to human and yeast PCNA, has a highly conserved secondary and tertiary structure and should therefore be classified in the DNA sliding clamp superfamily where PCNA is classified. Although it remains to be tested biochemically whether DNA and the G8R protein are capable of interacting with each other, given the distribution of the positively

charged residues in the helical region of the G8R protein, it is quite likely that it can bind DNA. It is important to note that as with the G5R protein, the origin of the G8R protein is not known and because of this, the function of the G8R protein may have evolved to be different than that of the PCNA protein, only retaining the DNA binding characteristic of the PCNA protein and nothing more. The DNA binding ability of the G8R protein has never been experimentally determined and given the novel role that human PCNA plays in transcription (Schultz-Norton *et al.*, 2007) and the knowledge that the G8R protein functions in poxvirus late transcription, it is of utmost importance to biochemically determine whether the G8R protein binds to DNA.

Homologs of the PCNA protein are expressed in at least two dsDNA virus families (*Baculoviridae* and *Herpesviridae*) (Matsumoto *et al.*, 1987, Zuccola *et al.*, 2000, Kool *et al.*, 1994) so it is not unlikely that a similar protein exists in the poxvirus. An interesting development in the ongoing search for a poxvirus sliding clamp protein occurred in July 2007 when the crystal structure of the VACV uracil DNA glycosylase (UDG) protein was determined (Schorman *et al.*, 2007). The poxvirus UDG protein has been previously shown to play a dual role in DNA replication and repair and the researchers who crystallized the protein revealed that it might act as a sliding clamp similar to PCNA, wrapping around the DNA as a homotetramer. Structurally, the UDG protein is not at all similar to the human PCNA protein, but as a homotetramer, it does create a channel in the center of the complex that would be large enough to fit duplex DNA. If the UDG protein were to be a sliding clamp similar in function to PCNA, however, there is one important detail that must be investigated further. The parts of the UDG homotetramer that come into contact with DNA would have to be somewhat

positively charged in order to interact with the negatively charged DNA molecule. Looking at the crystal structure of the UDG protein however, shows that the regions that would come into contact with the DNA molecule are primarily uncharged with at least two residues that are negatively charged. How the UDG tetramer in this conformation could successfully bind to DNA remains to be seen, but if it does bind to DNA in that conformation and if it is further shown to be the sliding clamp important in poxvirus replication, it could mean that the G8R protein with a structure that is very similar to that of PCNA could be the sliding clamp that is important for poxvirus transcription –two structurally different viral proteins performing similar functions during different stages of the poxvirus life cycle.

6.0 Discussion

Bioinformatics, by definition, is the application of statistical, mathematical and computational methods to the analysis of biological and molecular data (Luscombe *et al.*, 2001). Today, with the sequencing of thousands of genomes, bioinformatics involves two overall areas: data management and organization and, data analysis. Data management and organization can be difficult depending on the type of data to be stored (DNA or protein sequences versus structures), the number of sequences or structures being stored and in the case of databases that store sequence information, the length of the sequences that are stored since one thousand short sequences of 100 nucleotides does not take up as much space in a database as one thousand long sequences of a million nucleotides or longer. Often when creating databases, researchers try to add value to the data by providing search capabilities and links to tools that may be useful in subsequently analysing the data, which can raise some speed concerns when using such databases. One would think, in the case of smaller sized genomes coming from viruses, that data management and organization would be a relatively simplistic task, although this is not the case. For example, even though poxvirus genomes are relatively small (~200 kbp) compared to the human genome, housing sequence and gene information in a database that is easy and quick to search through has been a challenge. Fortunately, with newer, more powerful computers, a database with these capabilities has been developed, known as the virus orthologous clusters (VOCs) database (Ehlers *et al.*, 2002). VOCs is a database that began as a repository for poxvirus sequences and has now developed to include the sequences of 12 other virus families, all housed in a single location with easy access to all sequences of interest. It also

provides users with the ability to perform BLAST searches without having to copy and paste the sequence and navigate to the BLAST website, and to create multiple alignments from a subset of gene or protein sequences without having to download a separate alignment program amongst other features that researchers of these viruses find useful. Databases like VOCs have made the analysis of viral genomes, and in particular poxvirus genomes, a quick, efficient and understandable task.

The second area in which bioinformatics is involved is data analysis, and it can include using tools to determine the location of promoters and genes on the genome to modelling the structure of a protein using computationally expensive *ab initio* approaches. In the case of poxvirus genomes, bioinformatics has contributed in four different ways. The first, phylogeny, involves the determination of the origins of the different poxvirus species and likely began with the sequencing of the vaccinia and variola virus genomes in the 1990's (Shchelkunov *et al.*, 1993, Shchelkunov *et al.*, 1994, Shchelkunov *et al.*, 1996, Goebel *et al.*, 1990b). More recently, phylogenetics has contributed to a better understanding of the origins of certain strains of monkeypox virus (Chen *et al.*, 2005, Likos *et al.*, 2005) and to the understanding of the evolution of several different strains of variola virus (Esposito *et al.*, 2006). The second, comparative genomics, is related to the first in that it uses a different approach to determine the origins or evolution of poxvirus genomes. Comparative genomics in poxviruses involves the pairwise comparison of each gene in two poxvirus genomes with the hopes that differences between these genomes may help determine why one genome is more virulent than the other for example. The lack of certain genes in one genome may also help determine which genes are

virulence genes and which likely are not. The third way in which bioinformatics contributes to poxvirus research is through similarity and motif searches which involve the use of nucleotide, protein or profile databases to determine the function of a given protein. It was through the use of early protein database similarity searches that the functions of many proteins involved in poxvirus immune modulation were determined, and some of these even occurred prior to the sequencing of their respective genomes (Smith *et al.*, 1990b, Eppstein *et al.*, 1985). The final way in which bioinformatics contributes is through expression analysis, which involves the determination of which ORFs in the poxvirus genome are likely to be expressed and also involves the prediction of promoters as an added measure in the prediction of expression. Each of these ways contributes to the definition and determination of how poxviruses replicate and infect the host cell and with this understanding of the poxvirus life cycle comes the identification of which proteins could be better drug targets and the development of an improved vaccine that uses protein subunits rather than live attenuated virus.

The common theme of this dissertation is the use of bioinformatics to analyse poxvirus genomes. Each chapter uses bioinformatics methods that were outlined in the previous paragraph to help gain a better understanding of the poxvirus replication and infection cycle. Chapter 2 of this dissertation focuses on the determination of which genes are likely to be expressed in the virus. A novel method to predict expressed genes in poxvirus genomes by looking at the amino acid usage and purine content of each predicted ORF was presented. By using this method on newly sequenced or even existing poxvirus genomes, the poxvirus researcher will first be able to predict which ORFs are likely to be

expressed and he/she will then be able to narrow down which ORFs should be examined in more detail using biochemical techniques to determine expression. Expansion of this method to include information on the presence of a strong or weak poxvirus promoter would make this method even more useful.

Chapter 3 of this dissertation focuses on an interesting aspect of poxvirus phylogeny. By examining the background of self-dotplots, we found regions on the molluscum contagiosum genome that differed in nucleotide content compared to the remainder of the genome. The genes in these regions were also found to have a different codon usage compared to the remainder of the genes in the genome suggesting that these genes may have been acquired from a different virus or from the host. These results provide a new method of determining regions on the genome that may have been acquired elsewhere and they positively contribute to the general hypothesis that some poxvirus genes, especially genes involved in host immune modulation, have been acquired from the host.

Chapters 4 and 5 of this dissertation focus on the use of motif search tools in order to determine the function of two poxvirus proteins. In Chapter 4, results were presented that predicted the structure and function of the VACV G5R protein as a putative flap endonuclease protein using a HMM comparison search tool (HHsearch). Determining the function of all of the "unknown" proteins that are found in all poxviruses will not only lead to a better understanding of the poxvirus replication cycle but may also lead to the identification of new drug targets, and this work represents a start in this direction. Using bioinformatics to predict the function of these "unknown" genes allows the researcher to have a place to begin biochemical experiments rather than beginning with a guess about

protein function if bioinformatics is not used, and in the case of the G5R protein, further biochemical experiments would likely include testing the nuclease activity of the G5R protein, identifying its optimal substrate and determining where it localizes in an infected cell.

Related to the results presented in Chapter 4, Chapter 5 focuses on the use of a motif search tool (InterProScan) to identify potential sliding clamp proteins similar to PCNA, encoded by the VACV genome. Using InterProScan, the G8R protein was identified as being a sliding clamp similar to PCNA and these results were further confirmed by using HHsearch and by modeling the structure of the G8R protein. Prior to this study, previous work had shown that the G8R protein played a role in poxvirus transcription, although no direct interactions between the G8R protein and DNA were shown. Although it remains to be determined whether the G8R protein functions similarly to PCNA or whether it simply has a similar structure to PCNA because it binds to DNA during transcription, what is important to note is that had the VACV genome not been rescanned using the InterProScan motif search tool, this interesting connection between PCNA and G8R would not have been identified. This raises an important point that despite preliminary classifications of many poxvirus genes as playing a part in replication or transcription, periodic re-scans of each poxvirus genome including proteins with a putative function may help identify new functional domains on poxvirus proteins that have thus far been overlooked. Given the results presented in Chapter 5, poxvirus researchers should now be able to focus their biochemical experiments to determine if the G8R protein is able to bind to DNA and to determine what, if any, other proteins act as clients for the G8R protein.

The work presented in the previous four chapters represents a small step towards the understanding of various aspects of the poxvirus life cycle using bioinformatics. There are several other questions concerning the virus life cycle that remain to be elucidated including: What is the exact process by which poxviruses replicate their genomes? Can the locations of each poxvirus promoter be determined with accuracy? What are the functions of the "unknown" proteins in each poxvirus genome? Can the structures of these and other essential poxvirus proteins be solved? Each of these questions will hopefully be answered in the near future likely using a combination of bioinformatics and wet-lab techniques with bioinformatics playing a large preliminary role. Two of these questions require the development of more sophisticated algorithms that take into consideration aspects of the sequence like AT-content, in the case of promoter identification, or tertiary structure of distantly related proteins, in the case of identifying functions in "unknown" proteins, before they can be answered. The development of these new algorithms will likely come about in response to the increasing numbers of sequences and structures that are published and submitted to their respective databases. Perhaps one of the largest challenges that bioinformaticans will face as they begin to develop and improve upon existing algorithms, will be how to present a complex algorithm to the researcher in an understandable fashion, where the results of the algorithm are presented clearly and are easily interpreted. If bioinformaticians can overcome this challenge, then bioinformatics will be used progressively more by researchers as a first approach in the analysis of their genes and proteins of interest.

Bibliography

1. Viral Orthologous Clusters (VOCS) [http://www.virology.ca/pbr/vocs]

2. Ahn, B.Y., Moss, B. (1992). RNA polymerase-associated transcription specificity factor encoded by vaccinia virus. *Proceedings of the National Academy of Science of the United States of America.* **89**(8): 3536-3540.

3. Alcamí, A., Khanna, A., Paul, N.L., Smith, G.L. (1999). Vaccinia virus strains Lister, USSR and Evans express soluble and cell-surface tumour necrosis factor receptors. *The Journal of General Virology.* **80 (Pt 4)**: 949-959.

4. Alcamí, A., Symons, J.A., Smith, G.L. (2000). The vaccinia virus soluble alpha/beta interferon (IFN) receptor binds to the cell surface and protects cells from the antiviral effects of IFN. *Journal of Virology.* **74**(23): 11230-11239.

5. Aldaz-Carroll, L., Whitbeck, J.C., Ponce de Leon, M., Lou, H., Pannell, L.K., Lebowitz, J., Fogg, C., White, C.L., Moss, B., Cohen, G.H., Eisenberg, R.J. (2005). Physical and immunological characterization of a recombinant secreted form of the membrane protein encoded by the vaccinia virus L1R gene. *Virology.* **341**(1): 59-71.

6. Altschul, S.F., Koonin, E.V. (1998). Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. *Trends Biochem Sci.* **23**(11): 444-447.

7. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* **25**: 3389-3402.

8. Amegadzie, B.Y., Ahn, B.Y., Moss, B. (1992). Characterization of a 7-kilodalton subunit of vaccinia virus DNA- dependent RNA polymerase with structural similarities to the smallest subunit of eukaryotic RNA polymerase II. *Journal of Virology.* **66**(5): 3003-3010.

9. Artenstein, A.W., Johnson, C., Marbury, T.C., Morrison, D., Blum, P.S., Kemp, T., Nichols, R., Balser, J.P., Currie, M., Monath, T.P. (2005). A novel, cell culture-derived smallpox vaccine in vaccinia-naive adults. *Vaccine.* **23**(25): 3301-3309.

10. Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome.

Proceedings of the National Academy of Science of the United States of America. **98**(18): 10037-10041.

11. Banham, A.H., Smith, G.L. (1992). Vaccinia Virus Gene-B1R Encodes a 34-kDa Serine Threonine Protein Kinase That Localizes in Cytoplasmic Factories and Is Packaged into Virions. *Virology.* **191**: 803-812.

12. Bawden, A.L., Glassberg, K.J., Diggans, J., Shaw, R., Farmerie, W., Moyer, R.W. (2000). Complete genomic sequence of the Amsacta moorei entomopoxvirus: analysis and comparison with other poxviruses. *Virology.* **274**(1): 120-139.

13. Bayliss, C.D., Smith, G.L. (1997). Vaccinia virion protein VP8, the 25 kDa product of the L4R gene, binds single-stranded DNA and RNA with similar affinity. *Nucleic Acids Research.* **25**(20): 3984-3990.

14. Beattie, E., Paoletti, E., Tartaglia, J. (1995). Distinct patterns of IFN sensitivity observed in cells infected with vaccinia K3L- and E3L- mutant viruses. *Virology.* **210**(2): 254-263.

15. Beaud, G. (1995). Vaccinia virus DNA replication: A short review. *Biochimie.* **77**: 774-779.

16. Beaud, G., Beaud, R., Leader, D.P. (1995). Vaccinia virus gene H5R encodes a protein that is phosphorylated by the multisubstrate vaccinia virus B1R protein kinase. *Journal of Virology.* **69**(3): 1819-1826.

17. Beaud, G., Sharif, A., Topa-Masse, A., Leader, D.P. (1994). Ribosomal protein S2/Sa kinase purified from HeLa cells infected with vaccinia virus corresponds to the B1R protein kinase and phosphorylates in vitro the viral ssDNA-binding protein. *The Journal of General Virology.* **75 (Pt 2)**: 283-293.

18. Behbehani, A.M. (1983). The smallpox story: life and death of an old disease. *Microbiological Reviews.* **47**(4): 455-509.

19. Bell, S.J., Forsdyke, D.R. (1999). Deviations from Chargaff's second parity rule correlate with direction of transcription. *Journal of Theoretical Biology.* **197**(1): 63-76.

20. Belongia, E.A., Naleway, A.L. (2005). Smallpox vaccine: the good, the bad, and the ugly. *Clinical Medicine & Research.* **1**(2): 87-92.

21. Bertholet, C., Van, M.E., ten, H.-., Wittek, R. (1987). Vaccinia virus produces late mRNAs by discontinuous synthesis. *Cell.* **50**(2): 153-162.

22. Betakova, T., Wolffe, E.J., Moss, B. (1999). Regulation of vaccinia virus morphogenesis: phosphorylation of the A14L and A17L membrane proteins and C-terminal truncation of the A17L protein are dependent on the F10L kinase. *Journal of Virology.* **73**(5): 3534-3543.

23. Betakova, T., Wolffe, E.J., Moss, B. (2000). The vaccinia virus A14.5L gene encodes a hydrophobic 53-amino-acid virion membrane protein that enhances virulence in mice and is conserved among vertebrate poxviruses. *Journal of Virology.* **74**(9): 4085-492..

24. Black, E.P., Condit, R.C. (1996). Phenotypic characterization of mutants in vaccinia virus gene G2R, a putative transcription elongation factor. *Journal of Virology.* **70**(1): 47-54.

25. Black, M.E., Hruby, D.E. (1992). A single amino acid substitution abolishes feedback inhibition of vaccinia virus thymidine kinase. *The Journal of Biological Chemistry.* **267**(14): 9743-9748.

26. Bowman, G.D., O'Donnell, M., Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature.* **429**(6993): 724-730.

27. Brodie, R., Roper, R.L., Upton, C. (2004). JDotter: a Java interface to multiple dotplots generated by dotter. *Bioinformatics.* **20**(2): 279-281.

28. Brodie, R., Smith, A.J., Roper, R.L., Tcherepanov, V., Upton, C. (2004). Base-By-Base: single nucleotide-level analysis of whole viral genome alignments. *BMC Bioinformatics.* **5**(1): 96.

29. Brown, N.G., D, N.M., Beaud, G., Hardie, G., Leader, D.P. (2000). Identification of sites phosphorylated by the vaccinia virus B1R kinase in viral protein H5R. *BMC Biochemistry.* **1**: 2.

30. Broyles, S.S. (1991). A role for ATP hydrolysis in vaccinia virus early gene transcription. Dissociation of the early transcription factor-promoter complex. *The Journal of Biological Chemistry.* **266**(23): 15545-15548.

31. Broyles, S.S. (1993). Vaccinia virus encodes a functional dUTPase. *Virology.* **195**(2): 863-865.

32. Broyles, S.S. (2003). Vaccinia virus transcription. *The Journal of General Virology.* **84**(Pt 9): 2293-2303.

33. Broyles, S.S., Fesler, B.S. (1990). Vaccinia virus gene encoding a component of the viral early transcription factor. *J.Virol.* **64**(4): 1523-1529.

34. Broyles, S.S., Liu, X., Zhu, M., Kremer, M. (1999). Transcription factor YY1 is a vaccinia virus late promoter activator. *The Journal of Biological Chemistry.* **274**(50): 35662-35667.

35. Bruck, I., O'Donnell, M. (2001). The ring-type polymerase sliding clamp family. *Genome Biology.* **2**(1): REVIEWS3001.1-REVIEWS3001.3.

36. Calderara, S., Xiang, Y., Moss, B. (2001). Orthopoxvirus IL-18 binding proteins: affinities and antagonist activities. *Virology.* **279**(1): 22-6..

37. Cao, J.X., Koop, B.F., Upton, C. (1997). A human homolog of the vaccinia virus HindIII K4L gene is a member of the phospholipase D superfamily. *Virus Research.* **48**(1): 11-18.

38. Carpenter, M.S., DeLange, A.M. (1992). Identification of a temperature-sensitive mutant of vaccinia virus defective in late but not intermediate gene expression. *Virology.* **188**(1): 233-244.

39. Carter, G.C., Law, M., Hollinshead, M., Smith, G.L. (2005). Entry of the vaccinia virus intracellular mature virion and its interactions with glycosaminoglycans. *The Journal of General Virology.* **86**(Pt 5): 1279-1290.

40. Cassetti, M.A., Moss, B. (1996). Interaction of the 82-kDa subunit of the vaccinia virus early transcription factor heterodimer with the promoter core sequence directs downstream DNA binding of the 70-kDa subunit. *Proceedings of the National Academy of Science of the United States of America.* **93**(15): 7540-7545.

41. Centers for Disease Control and Prevention (CDC). (2007). Vulvar vaccinia infection after sexual contact with a military smallpox vaccinee--Alaska, 2006. *MMWR. Morbidity and Mortality Weekly Report.* **56**(17): 417-419.

42. Centers for Disease Control and Prevention (CDC). (2007). Household transmission of vaccinia virus from contact with a military smallpox vaccinee--Illinois and Indiana, 2007. *MMWR. Morbidity and Mortality Weekly Report.* **56**(19): 478-481.

43. Challberg, M.D., Englund, P.T. (1979). Purification and properties of the deoxyribonucleic acid polymerase induced by vaccinia virus. *The Journal of Biological Chemistry.* **254**(16): 7812-7819.

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

45. Chen, N., Li, G., Liszewski, M.K., Atkinson, J.P., Jahrling, P.B., Feng, Z., Schriewer, J., Buck, C., Wang, C., Lefkowitz, E.J., Esposito, J.J., Harms, T., Damon, I.K., Roper, R.L., Upton, C., Buller, R.M. (2005). Virulence differences between monkeypox virus isolates from West Africa and the Congo basin. *Virology.* **340**(1): 46-63.

46. Chivian, D., Kim, D.E., Malmström, L., Bradley, P., Robertson, T., Murphy, P., Strauss, C.E., Bonneau, R., Rohl, C.A., Baker, D. (2003). Automated prediction of CASP-5 structures using the Robetta server. *Proteins.* **53 Suppl 6**: 524-533.

47. Chivian, D., Kim, D.E., Malmström, L., Schonbrun, J., Rohl, C.A., Baker, D. (2005). Prediction of CASP6 structures using automated Robetta protocols. *Proteins.* **61 Suppl 7**: 157-166.

48. Christen, L.M., Sanders, M., Wiler, C., Niles, E.G. (1998). Vaccinia virus nucleoside triphosphate phosphohydrolase I is an essential viral early gene transcription termination factor. *Virology.* **245**(2): 360-371.

49. Chung, C.S., Chen, C.H., Ho, M.Y., Huang, C.Y., Liao, C.L., Chang, W. (2006). Vaccinia virus proteome: identification of proteins in vaccinia virus intracellular mature virion particles. *Journal of Virology.* **80**(5): 2127-2140.

50. Chung, C.S., Hsiao, J.C., Chang, Y.S., Chang, W. (1998). A27L protein mediates vaccinia virus interaction with cell surface heparan sulfate. *Journal of Virology.* **72**(2): 1577-1585.

51. Chung, C.S., Huang, C.Y., Chang, W. (2005). Vaccinia virus penetration requires cholesterol and results in specific viral envelope proteins associated with lipid rafts. *Journal of Virology.* **79**(3): 1623-1634.

52. Codontree [http://bioweb.pasteur.fr/seqanal/interfaces/codontree.html]

53. Colamonici, O.R., Domanski, P., Sweitzer, S.M., Larner, A., Buller, R.M. (1995). Vaccinia virus B18R gene encodes a type I interferon-binding protein that blocks interferon alpha transmembrane signaling. *The Journal of Biological Chemistry.* **270**(27): 15974-15978.

54. Condit, R.C., Moussatche, N., Traktman, P. (2006). In a nutshell: structure and assembly of the vaccinia virion. *Advances in Virus Research.* **66**: 31-124.

55. Cowley, R., Greenaway, P.J. (1990). Nucleotide sequence comparison of homologous genomic regions from variola, monkeypox, and vaccinia viruses. *Journal of Medical Virology.* **31**(4): 267-271.

56. Cresawn, S.G., Condit, R.C. (2007). A targeted approach to identification of vaccinia virus postreplicative transcription elongation factors: genetic evidence for a role of the H5R gene in vaccinia transcription. *Virology.* **363**(2): 333-341.

57. Cudmore, S., Blasco, R., Vincentelli, R., Esteban, M., Sodeik, B., Griffiths, G., Krijnse Locker, J. (1996). A vaccinia virus core protein, p39, is membrane associated. *Journal of Virology.* **70**(10): 6909-6921.

58. Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M., Barton, G.J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics.* **14**(10): 892-893.

59. Cyrklaff, M., Risco, C., Fernández, J.J., Jiménez, M.V., Estéban, M., Baumeister, W., Carrascosa, J.L. (2005). Cryo-electron tomography of vaccinia virus. *Proceedings of the National Academy of Science of the United States of America.* **102**(8): 2772-2777.

60. da Fonseca, F.G., Silva, R.L., Marques, J.T., Ferreira, P.C., Kroon, E.G. (1999). The genome of cowpox virus contains a gene related to those encoding the epidermal growth factor, transforming growth factor alpha and vaccinia growth factor. *Virus Genes.* **18**(2): 151-160.

61. da Fonseca, F.G., Weisberg, A.S., Caeiro, M.F., Moss, B. (2004). Vaccinia virus mutants with alanine substitutions in the conserved G5R gene fail to initiate morphogenesis at the nonpermissive temperature. *Journal of Virology.* **78**(19): 10238-10248.

62. da Fonseca, F.G., Wolffe, E.J., Weisberg, A., Moss, B. (2000). Effects of deletion or stringent repression of the H3L envelope gene on vaccinia virus replication. *Journal of Virology.* **74**(16): 7518-7528.

63. da Fonseca, F.G., Wolffe, E.J., Weisberg, A., Moss, B. (2000). Characterization of the vaccinia virus H3L envelope protein: topology and posttranslational membrane insertion via the C-terminal hydrophobic tail. *Journal of Virology.* **74**(16): 7508-7517.

64. 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.

65. Da Silva, M., Upton, C. (2005). Using purine skews to predict genes in AT-rich poxviruses. *BMC Genomics.* **6**(1): 22.

66. Da Silva, M., Upton, C. (2005). Host-derived pathogenicity islands in poxviruses. *Virology Journal.* **2**(1): 30.

67. Dales, S., Mosbach, E.H. (1968). Vaccinia as a model for membrane biogenesis. *Virology.* **35**(4): 564-583.

68. Damaso, C.R., Esposito, J.J., Condit, R.C., Moussatché, N. (2000). An emergent poxvirus from humans and cattle in Rio de Janeiro State: Cantagalo virus may derive from Brazilian smallpox vaccine. *Virology.* **277**(2): 439-449.

69. Darai, G., Reisner, H., Scholz, J., Schnitzler, P., Lorbacher de Ruiz, H. (1986). Analysis of the genome of molluscum contagiosum virus by restriction endonuclease analysis and molecular cloning. *Journal of Medical Virology.* **18**(1): 29-39.

70. Davies, M.V., Chang, H.W., Jacobs, B.L., Kaufman, R.J. (1993). The E3L and K3L Vaccinia Virus Gene Products Stimulate Translation Through Inhibition of the Double-Stranded RNA- Dependent Protein Kinase by Different Mechanisms. *Journal of Virology.* **67**(3): 1688-1692.

71. Davis, J.A., Takagi, Y., Kornberg, R.D., Asturias, F.A. (2002). Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Molecular Cell.* **10**(2): 409-415.

72. Davison, A.J., Moss, B. (1989). Structure of vaccinia virus early promoters. *Journal of Molecular Biology.* **210**(4): 749-769.

73. Davison, A.J., Moss, B. (1989). Structure of vaccinia virus late promoters. *Journal of Molecular Biology.* **210**(4): 771-784.

74. De Silva, F.S., Lewis, W., Berglund, P., Koonin, E.V., Moss, B. (2007). Poxvirus DNA primase. *Proceedings of the National Academy of Science of the United States of America.* .

75. De Silva, F.S., Moss, B. (2003). Vaccinia virus uracil DNA glycosylase has an essential role in DNA synthesis that is independent of its glycosylase activity: catalytic site mutations reduce virulence but not virus replication in cultured cells. *Journal of Virology.* **77**(1): 159-166.

76. Dellis, S., Strickland, K.C., McCrary, W.J., Patel, A., Stocum, E., Wright, C.F. (2004). Protein interactions among the vaccinia virus late transcription factors. *Virology.* **329**(2): 328-336.

77. DeMasi, J., Traktman, P. (2000). Clustered charge-to-alanine mutagenesis of the vaccinia virus H5 gene: isolation of a dominant, temperature-sensitive mutant with a profound defect in morphogenesis. *Journal of Virology.* **74**(5): 2393-2405.

78. Deng, L., Shuman, S. (1997). Elongation properties of vaccinia virus RNA polymerase: pausing, slippage, 3' end addition, and termination site choice. *Biochemistry.* **36**(50): 15892-15899.

79. Deng, L., Shuman, S. (1998). Vaccinia NPH-I, a DExH-box ATPase, is the energy coupling factor for mRNA transcription termination. *Genes & Development.* **12**(4): 538-546.

80. Dobbelstein, M., Shenk, T. (1996). Protection against apoptosis by the vaccinia virus SPI-2 (B13R) gene product. *Journal of Virology.* **70**(9): 6479-6485.

81. Doglio, L., De Marco, A., Schleich, S., Roos, N., Krijnse Locker, J. (2002). The Vaccinia virus E8R gene product: a viral membrane protein that is made

early in infection and packaged into the virions' core. *Journal of Virology.* **76**(19): 9773-9786.

82. Doglio, L., De Marco, A., Schleich, S., Roos, N., Krijnse Locker, J. (2002). The Vaccinia virus E8R gene product: a viral membrane protein that is made early in infection and packaged into the virions' core. *Journal of Virology.* **76**(19): 9773-9786.

83. Doré, A.S., Kilkenny, M.L., Jones, S.A., Oliver, A.W., Roe, S.M., Bell, S.D., Pearl, L.H. (2006). Structure of an archaeal PCNA1-PCNA2-FEN1 complex: elucidating PCNA subunit and client enzyme specificity. *Nucleic Acids Research.* **34**(16): 4515-4526.

84. Dubochet, J., Adrian, M., Richter, K., Garces, J., Wittek, R. (1994). Structure of intracellular mature vaccinia virus observed by cryoelectron microscopy. *Journal of Virology.* **68**(3): 1935-1941.

85. Earl, P.L., Jones, E.V., Moss, B. (1986). Homology between DNA polymerases of poxviruses, herpesviruses, and adenoviruses: nucleotide sequence of the vaccinia virus DNA polymerase gene. *Proceedings of the National Academy of Science of the United States of America.* **83**(11): 3659-3663.

86. Eckert, D., Williams, O., Meseda, C.A., Merchlinsky, M. (2005). Vaccinia virus nicking-joining enzyme is encoded by K4L (VACWR035). *Journal of Virology.* **79**(24): 15084-15090.

87. Ehlers, A., Osborne, J., Slack, S., Roper, R.L., Upton, C. (2002). Poxvirus Orthologous Clusters (POCs). *Bioinformatics.* **18**(11): 1544-1545.

88. El Omari, K., Solaroli, N., Karlsson, A., Balzarini, J., Stammers, D.K. (2006). Structure of vaccinia virus thymidine kinase in complex with dTTP: insights for drug design. *BMC Structural Biology.* **6**: 22.

89. Eppstein, D.A., Marsh, Y.V., Schreiber, A.B., Newman, S.R., Todaro, G.J., Nestor, J.J. (1985). Epidermal growth factor receptor occupancy inhibits vaccinia virus infection. *Nature.* **318**(6047): 663-665.

90. Esposito, J.J., Sammons, S.A., Frace, A.M., Osborne, J.D., Olsen-Rasmussen, M., Zhang, M., Govil, D., Damon, I.K., Kline, R., Laker, M., Li, Y., Smith, G.L., Meyer, H., Leduc, J.W., Wohlhueter, R.M. (2006). Genome sequence diversity and clues to the evolution of variola (smallpox) virus. *Science.* **313**(5788): 807-812.

91. Evans, E., Klemperer, N., Ghosh, R., Traktman, P. (1995). The vaccinia virus D5 protein, which is required for DNA replication, is a nucleic acid-independent nucleoside triphosphatase. *Journal of Virology.* **69**(9): 5353-5361.

92. Evans, E., Traktman, P. (1987). Molecular genetic analysis of a vaccinia virus gene with an essential role in DNA replication. *Journal of Virology.* **61**(10): 3152-3162.

93. Everett, H., McFadden, G. (2002). Poxviruses and apoptosis: a time to die. *Current Opinion in Microbiology.* **5**(4): 395-402.

94. Fenner, F., Henderson, D.A., Arita, I., Jezek, Z., and Ladnyi, I. (1988). *Smallpox and its Eradication.* World Health Organization: Geneva.

95. Funahashi, S., Sato, T., Shida, H. (1988). Cloning and characterization of the gene encoding the major protein of the A-type inclusion body of cowpox virus. *The Journal of General Virology.* **69 (Pt 1)**: 35-47.

96. Funahashi, S., Sato, T., Shida, H. (1988). Cloning and characterization of the gene encoding the major protein of the A-type inclusion body of cowpox virus. *J.Gen.Virol.* **69**: 35-47.

97. Garcia, A.D., Moss, B. (2001). Repression of vaccinia virus Holliday junction resolvase inhibits processing of viral DNA into unit-length genomes. *Journal of Virology.* **75**(14): 6460-6471.

98. Garforth, S.J., Ceska, T.A., Suck, D., Sayers, J.R. (1999). Mutagenesis of conserved lysine residues in bacteriophage T5 5'-3' exonuclease suggests separate mechanisms of endo-and exonucleolytic cleavage. *Proceedings of the National Academy of Science of the United States of America.* **96**(1): 38-43.

99. Gershon, P.D., Moss, B. (1990). Early transcription factor subunits are encoded by vaccinia virus late genes. *Proceedings of the National Academy of Science of the United States of America.* **87**(11): 4401-4405.

100. Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P., Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. *Virology.* **179**: 247-266.

101. Goebel, S.J., Johnson, G.P., Perkus, M.E., Davis, S.W., Winslow, J.P., Paoletti, E. (1990). The complete DNA sequence of vaccinia virus. *Virology.* **179**(1): 247-66, 517-63.

102. Griffiths, G., Roos, N., Schleich, S., Locker, J.K. (2001). Structure and assembly of intracellular mature vaccinia virus: thin-section analyses. *Journal of Virology.* **75**(22): 11056-11070.

103. Griffiths, G., Wepf, R., Wendt, T., Locker, J.K., Cyrklaff, M., Roos, N. (2001). Structure and assembly of intracellular mature vaccinia virus: isolated-particle analysis. *Journal of Virology.* **75**(22): 11034-11055.

104. Gross, C.P., Sepkowitz, K.A. (1998). The myth of the medical breakthrough: smallpox, vaccination, and Jenner reconsidered. *International Journal of Infectious Diseases.* **3**(1): 54-60.

105. Gubser, C., Hué, S., Kellam, P., Smith, G.L. (2004). Poxvirus genomes: a phylogenetic analysis. *The Journal of General Virology.* **85**(Pt 1): 105-117.

106. Gunasinghe, S.K., Hubbs, A.E., Wright, C.F. (1998). A vaccinia virus late transcription factor with biochemical and molecular identity to a human cellular protein. *The Journal of Biological Chemistry.* **273**(42): 27524-27530.

107. Haga, I.R., Bowie, A.G. (2005). Evasion of innate immunity by vaccinia virus. *Parasitology.* **130 Suppl**: S11-S25.

108. Heljasvaara, R., Rodríguez, D., Risco, C., Carrascosa, J.L., Esteban, M., Rodríguez, J.R. (2001). The major core protein P4a (A10L gene) of vaccinia virus is essential for correct assembly of viral DNA into the nucleoprotein complex to form immature viral particles. *Journal of Virology.* **75**(13): 5778-5795.

109. Henderson, D.A., Inglesby, T.V., Bartlett, J.G., Ascher, M.S., Eitzen, E., Jahrling, P.B., Hauer, J., Layton, M., McDade, J., Osterholm, M.T., O'Toole, T., Parker, G., Perl, T., Russell, P.K., Tonat, K. (1999). Smallpox as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *JAMA : The Journal of the American Medical Association.* **281**(22): 2127-2137.

110. Heuser, J. (2005). Deep-etch EM reveals that the early poxvirus envelope is a single membrane bilayer stabilized by a geodetic "honeycomb" surface coat. *The Journal of Cell Biology.* **169**(2): 269-283.

111. Hollinshead, M., Vanderplasschen, A., Smith, G.L., Vaux, D.J. (1999). Vaccinia virus intracellular mature virions contain only one lipid membrane. *Journal of Virology.* **73**(2): 1503-1517.

112. Hooda-Dhingra, U., Thompson, C.L., Condit, R.C. (1989). Detailed phenotypic characterization of five temperature-sensitive mutants in the 22- and 147-kilodalton subunits of vaccinia virus DNA-dependent RNA polymerase. *Journal of Virology.* **63**(2): 714-729.

113. Hosfield, D.J., Mol, C.D., Shen, B., Tainer, J.A. (1998). Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell.* **95**(1): 135-146.

114. Howell, M.L., Sanders-Loehr, J., Loehr, T.M., Roseman, N.A., Mathews, C.K., Slabaugh, M.B. (1992). Cloning of the vaccinia virus ribonucleotide reductase small subunit gene. Characterization of the gene product expressed in Escherichia coli. *The Journal of Biological Chemistry.* **267**(3): 1705-1711.

115. Hsiao, J.C., Chung, C.S., Chang, W. (1998). Cell surface proteoglycans are necessary for A27L protein-mediated cell fusion: identification of the N-terminal region of A27L protein as the glycosaminoglycan-binding domain. *Journal of Virology.* **72**(10): 8374-8379.

116. Hsiao, J.C., Chung, C.S., Chang, W. (1999). Vaccinia virus envelope D8L protein binds to cell surface chondroitin sulfate and mediates the adsorption of intracellular mature virions to cells. *Journal of Virology.* **73**(10): 8750-8761.

117. Hu, X., Wolffe, E.J., Weisberg, A.S., Carroll, L.J., Moss, B. (1998). Repression of the A8L gene, encoding the early transcription factor 82-kilodalton subunit, inhibits morphogenesis of vaccinia virions. *Journal of Virology.* **72**(1): 104-112.

118. Hu, Z. (2002). DNACreator version 1.0. *Center for Computational Research at State University of New York at Buffalo.* .

119. Hubbs, A.E., Wright, C.F. (1996). The A2L intermediate gene product is required for in vitro transcription from a vaccinia virus late promoter. *Journal of Virology.* **70**(1): 327-331.

120. Hughes, A.L., Friedman, R. (2005). Poxvirus genome evolution by gene gain and loss. *Molecular Phylogenetics and Evolution.* **35**(1): 186-195.

121. Hughes, S.J., Johnston, L.H., de Carlos, A., Smith, G.L. (1991). Vaccinia virus encodes an active thymidylate kinase that complements a cdc8 mutant of Saccharomyces cerevisiae. *The Journal of Biological Chemistry.* **266**(30): 20103-20109.

122. Hwang, K.Y., Baek, K., Kim, H.Y., Cho, Y. (1998). The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. *Nature Structural Biology.* **5**(8): 707-713.

123. Ichihashi, Y., Oie, M., Tsuruhara, T. (1984). Location of DNA-binding proteins and disulfide-linked proteins in vaccinia virus structural elements. *Journal of Virology.* **50**(3): 929-938.

124. Ishii, K., Moss, B. (2002). Mapping interaction sites of the A20R protein component of the vaccinia virus DNA replication complex. *Virology.* **303**(2): 232-239.

125. Johnson, E.S. (2004). Protein modification by SUMO. *Annual Review of Biochemistry.* **73**: 355-382.

126. Kane, E.M., Shuman, S. (1992). Temperature-sensitive mutations in the vaccinia virus H4 gene encoding a component of the virion RNA polymerase. *Journal of Virology.* **66**(10): 5752-5762.

127. Karkas, J.D., Rudner, R., Chargaff, E. (1968). Seapration of B. subtilis DNA into complementary strands. II. Template functions and composition as determined by transcription with RNA polymerase. *Proceedings of the National Academy of Science of the United States of America.* **60**(3): 915-920.

128. Kato, S.E., Condit, R.C., Moussatché, N. (2007). The vaccinia virus E8R gene product is required for formation of transcriptionally active virions. *Virology.* : doi:10.1016/j.virol.2007.05.002.

129. Kato, S.E., Strahl, A.L., Moussatche, N., Condit, R.C. (2004). Temperature-sensitive mutants in the vaccinia virus 4b virion structural protein assemble malformed, transcriptionally inactive intracellular mature virions. *Virology.* **330**(1): 127-146.

130. Katsafanas, G.C., Moss, B. (2004). Vaccinia virus intermediate stage transcription is complemented by Ras-GTPase-activating protein SH3 domain-binding protein (G3BP) and cytoplasmic activation/proliferation-

associated protein (p137) individually or as a heterodimer. *The Journal of Biological Chemistry.* **279**(50): 52210-52217.

131. Keck, J.G., Baldick, C.J., Moss, B. (1990). Role of DNA replication in vaccinia virus gene expression: a naked template is required for transcription of three late trans-activator genes. *Cell.* **61**(5): 801-809.

132. Keck, J.G., Feigenbaum, F., Moss, B. (1993). Mutational analysis of a predicted zinc-binding motif in the 26-kilodalton protein encoded by the vaccinia virus A2L gene: correlation of zinc binding with late transcriptional transactivation activity. *Journal of Virology.* **67**(10): 5749-5753.

133. Keck, J.G., Kovacs, G.R., Moss, B. (1993). Overexpression, Purification, and Late Transcription Factor Activity of the 17-Kilodalton Protein Encoded by the Vaccinia Virus A1L-Gene. *Journal of Virology.* **67**(10): 5740-5748.

134. Kelman, Z. (1997). PCNA: structure, functions and interactions. *Oncogene.* **14**(6): 629-640.

135. Kettle, S., Alcamí, A., Khanna, A., Ehret, R., Jassoy, C., Smith, G.L. (1997). Vaccinia virus serpin B13R (SPI-2) inhibits interleukin-1beta-converting enzyme and protects virus-infected cells from TNF- and Fas-mediated apoptosis, but does not prevent IL-1beta-induced fever. *The Journal of General Virology.* **78 (Pt 3)**: 677-685.

136. Kibler, K.V., Shors, T., Perkins, K.B., Zeman, C.C., Banaszak, M.P., Biesterfeldt, J., Langland, J.O., Jacobs, B.L. (1997). Double-stranded RNA is a trigger for apoptosis in vaccinia virus-infected cells. *Journal of Virology.* **71**(3): 1992-2003.

137. Kim, D.E., Chivian, D., Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research.* **32**(Web Server issue): W526-W531.

138. Knutson, B.A., Liu, X., Oh, J., Broyles, S.S. (2006). Vaccinia virus intermediate and late promoter elements are targeted by the TATA-binding protein. *Journal of Virology.* **80**(14): 6784-6793.

139. Kool, M., Ahrens, C.H., Goldbach, R.W., Rohrmann, G.F., Vlak, J.M. (1994). Identification of genes involved in DNA replication of the Autographa californica baculovirus. *Proceedings of the National Academy of Science of the United States of America.* **91**(23): 11212-11216.

140. Kovacs, G.R., Moss, B. (1996). The vaccinia virus H5R gene encodes late gene transcription factor 4: purification, cloning, and overexpression. *Journal of Virology.* **70**(10): 6796-6802.

141. Lackner, C.A., Condit, R.C. (2000). Vaccinia virus gene A18R DNA helicase is a transcript release factor. *The Journal of Biological Chemistry.* **275**(2): 1485-1494.

142. Latner, D.R., Thompson, J.M., Gershon, P.D., Storrs, C., Condit, R.C. (2002). The positive transcription elongation factor activity of the vaccinia virus J3 protein is independent from its (nucleoside-2'-O-) methyltransferase and poly(A) polymerase stimulatory functions. *Virology.* **301**(1): 64-80.

143. Latner, D.R., Xiang, Y., Lewis, J.I., Condit, J., Condit, R.C. (2000). The vaccinia virus bifunctional gene J3 (nucleoside-2'-O-)-methyltransferase and poly(A) polymerase stimulatory factor is implicated as a positive transcription elongation factor by two genetic approaches. *Virology.* **269**(2): 345-355.

144. Lee, S.B., Esteban, M. (1994). The interferon-induced double-stranded RNA-activated protein kinase induces apoptosis. *Virology.* **199**(2): 491-496.

145. Li, J., Broyles, S.S. (1993). Recruitment of Vaccinia Virus RNA Polymerase to an Early Gene Promoter by the Viral Early Transcription Factor. *The Journal of Biological Chemistry.* **268**(4): 2773-2780.

146. Li, X., Li, J., Harrington, J., Lieber, M.R., Burgers, P.M. (1995). Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. *The Journal of Biological Chemistry.* **270**(38): 22109-22112.

147. Likos, A.M., Sammons, S.A., Olson, V.A., Frace, A.M., Li, Y., Olsen-Rasmussen, M., Davidson, W., Galloway, R., Khristova, M.L., Reynolds, M.G., Zhao, H., Carroll, D.S., Curns, A., Formenty, P., Esposito, J.J., Regnery, R.L., Damon, I.K. (2005). A tale of two clades: monkeypox viruses. *The Journal of General Virology.* **86**(Pt 10): 2661-2672.

148. Lin, C.L., Chung, C.S., Heine, H.G., Chang, W. (2000). Vaccinia virus envelope H3L protein binds to cell surface heparan sulfate and is important for intracellular mature virion morphogenesis and virus infection in vitro and in vivo. *Journal of Virology.* **74**(7): 3353-3365.

149. Lin, S., Chen, W., Broyles, S.S. (1992). The vaccinia virus B1R gene product is a serine/threonine protein kinase. *Journal of Virology.* **66**(5): 2717-2723.

150. Lobry, J.R. (1996). A simple vectorial representation of DNA sequences for the detection of replication origins in bacteria. *Biochimie.* **78**(5): 323-326.

151. Locker, J.K., Kuehn, A., Schleich, S., Rutter, G., Hohenberg, H., Wepf, R., Griffiths, G. (2000). Entry of the two infectious forms of vaccinia virus at the plasma membane is signaling-dependent for the IMV but not the EEV. *Molecular Biology of the Cell.* **11**(7): 2497-2511.

152. Lofquist, J.M., Weimert, N.A., Hayney, M.S. (2003). Smallpox: a review of clinical disease and vaccination. *American Journal of Health-System Pharmacy : AJHP : Official Journal of the American Society of Health-System Pharmacists.* **60**(8): 749-56; quiz 757-8.

153. Luscombe, N.M., Greenbaum, D., Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine.* **40**(4): 346-358.

154. Mahalingam, S., Karupiah, G. (2000). Modulation of chemokines by poxvirus infections. *Current Opinion in Immunology.* **12**(4): 409-412.

155. Mallardo, M., Leithe, E., Schleich, S., Roos, N., Doglio, L., Krijnse Locker, J. (2002). Relationship between vaccinia virus intracellular cores, early mRNAs, and DNA replication sites. *Journal of Virology.* **76**(10): 5167-5183.

156. Matsumoto, K., Moriuchi, T., Koji, T., Nakane, P.K. (1987). Molecular cloning of cDNA coding for rat proliferating cell nuclear antigen (PCNA)/cyclin. *The EMBO Journal.* **6**(3): 637-642.

157. McFadden, G., Murphy, P.M. (2000). Host-related immunomodulators encoded by poxviruses and herpesviruses. *Current Opinion in Microbiology.* **3**(4): 371-378.

158. McKelvey, T.A., Andrews, S.C., Miller, S.E., Ray, C.A., Pickup, D.J. (2002). Identification of the orthopoxvirus p4c gene, which encodes a structural protein that directs intracellular mature virus particles into A-type inclusions. *Journal of Virology.* **76**(22): 11216-11225.

159. Mercer, J., Traktman, P. (2003). Investigation of structural and functional motifs within the vaccinia virus A14 phosphoprotein, an essential component of the virion membrane. *Journal of Virology.* **77**(16): 8857-8871.

160. Meyer, H., Rziha, H.J. (1993). Characterization of the gene encoding the A-type inclusion protein of camelpox virus and sequence comparison with other orthopoxviruses. *The Journal of General Virology.* **74 (Pt 8)**: 1679-1684.

161. Mohamed, M.R., Niles, E.G. (2000). Interaction between nucleoside triphosphate phosphohydrolase I and the H4L subunit of the viral RNA polymerase is required for vaccinia virus early gene transcript release. *The Journal of Biological Chemistry.* **275**(33): 25798-25804.

162. Mohamed, M.R., Niles, E.G. (2001). The viral RNA polymerase H4L subunit is required for Vaccinia virus early gene transcription termination. *The Journal of Biological Chemistry.* **276**(23): 20758-20765.

163. Moore, Z.S., Seward, J.F., Lane, J.M. (2006). Smallpox. *Lancet.* **367**(9508): 425-435.

164. Moss, B. (2001). Poxviridae: The Viruses and Their Replication. In *Fundamental Virology.* Edited by Fields, B.N., Knipe, D.M., Howley, P.M.. Lippincott-Raven Publishers: Philadelphia. pp. 2849-2883.

165. Moss, B. (2006). Poxvirus entry and membrane fusion. *Virology.* **344**(1): 48-54.

166. Mossman, K., Upton, C., McFadden, G. (1995). The myxoma virus-soluble interferon-gamma receptor homolog, M-T7, inhibits interferon-gamma in a species-specific manner. *The Journal of Biological Chemistry.* **270**(7): 3031-3038.

167. Mullick, J., Bernet, J., Panse, Y., Hallihosur, S., Singh, A.K., Sahu, A. (2005). Identification of complement regulatory domains in vaccinia virus complement control protein. *Journal of Virology.* **79**(19): 12382-12393.

168. Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., Darst, S.A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science.* **296**(5571): 1285-1290.

169. Nagasse-Sugahara, T.K., Kisielius, J.J., Ueda-Ito, M., Curti, S.P., Figueiredo, C.A., Cruz, A.S., Silva, M.M., Ramos, C.H., Silva, M.C., Sakurai, T., Salles-

Gomes, L.F. (2004). Human vaccinia-like virus outbreaks in São Paulo and Goiás States, Brazil: virus detection, isolation and identification. *Rev Inst Med Trop Sao Paulo.* **46**(6): 315-322.

170. Nakamura, Y., Gojobori, T., Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research.* **28**(1): 292.

171. Nichols, R.J., Wiebe, M.S., Traktman, P. (2006). The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. *Mol Biol Cell.* **17**(5): 2451-2464.

172. Nishino, T., Ishino, Y., Morikawa, K. (2006). Structure-specific DNA nucleases: structural basis for 3D-scissors. *Current Opinion in Structural Biology.* **16**(1): 60-67.

173. Notredame, C., Higgins, D.G., Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology.* **302**(1): 205-217.

174. Oda, K.I., Joklik, W.K. (1967). Hybridization and sedimentation studies on "early" and "late" vaccinia messenger RNA. *Journal of Molecular Biology.* **27**(3): 395-419.

175. Oroskar, A.A., Read, G.S. (1989). Control of mRNA stability by the virion host shutoff function of herpes simplex virus. *Journal of Virology.* **63**(5): 1897-1906.

176. Palacios, S., Perez, L.H., Welsch, S., Schleich, S., Chmielarska, K., Melchior, F., Locker, J.K. (2005). Quantitative SUMO-1 modification of a vaccinia virus protein is required for its specific localization and prevents its self-association. *Molecular Biology of the Cell.* **16**(6): 2822-2835.

177. Parrish, S., Resch, W., Moss, B. (2007). Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression. *Proceedings of the National Academy of Science of the United States of America.* **104**(7): 2139-2144.

178. Passarelli, A.L., Kovacs, G.R., Moss, B. (1996). Transcription of a vaccinia virus late promoter template: requirement for the product of the A2L intermediate-stage gene. *Journal of Virology.* **70**(7): 4444-4450.

179. Patel, D.D., Pickup, D.J. (1989). The second-largest subunit of the poxvirus RNA polymerase is similar to the corresponding subunits of procaryotic and eucaryotic RNA polymerases. *Journal of Virology.* **63**(3): 1076-1086.

180. Peden J. (1999). **Analysis of Codon Usage.** *PhD Dissertation.* University of Nottingham.

181. Pedersen, K., Snijder, E.J., Schleich, S., Roos, N., Griffiths, G., Locker, J.K. (2000). Characterization of vaccinia virus intracellular cores: implications for viral uncoating and core structure. *Journal of Virology.* **74**(8): 3525-3536.

182. Person-Fernandez, A., Beaud, G. (1986). Purification and characterization of a protein synthesis inhibitor associated with vaccinia virus. *The Journal of Biological Chemistry.* **261**(18): 8283-8289.

183. Pesole, G., Attimonelli, M., Liuni, S. (1988). A backtranslation method based on codon usage strategy. *Nucleic Acids Research.* **16**(5): 1715-1728.

184. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry.* **25**(13): 1605-1612.

185. Pogo, B.G. (1980). Changes in parental vaccinia virus DNA after viral penetration into cells. *Virology.* **101**(2): 520-524.

186. Pogo, B.G., Dales, S. (1969). Two deoxyribonuclease activities within purified vaccinia virus. *Proceedings of the National Academy of Science of the United States of America.* **63**(3): 820-827.

187. Pogo, B.G., O'Shea, M., Freimuth, P. (1981). Initiation and termination of vaccinia virus DNA replication. *Virology.* **108**(1): 241-248.

188. Pogo, B.G., O'Shea, M.T. (1978). The mode of replication of vaccinia virus DNA. *Virology.* **84**(1): 1-8.

189. Radetsky, M. (1999). Smallpox: a history of its rise and fall. *The Pediatric Infectious Disease Journal.* **18**(2): 85-93.

190. Ravanello, M.P., Hruby, D.E. (1994). Conditional lethal expression of the vaccinia virus L1R myristylated protein reveals a role in virion assembly. *Journal of Virology.* **68**(10): 6401-6410.

191. Reckmann, I., Higley, S., Way, M. (1997). The vaccinia virus F17R protein interacts with actin. *FEBS Letters.* **409**(2): 141-146.

192. Rodriguez, D., Bárcena, M., Möbius, W., Schleich, S., Esteban, M., Geerts, W.J., Koster, A.J., Griffiths, G., Locker, J.K. (2006). A vaccinia virus lacking A10L: viral core proteins accumulate on structures derived from the endoplasmic reticulum. *Cellular Microbiology.* **8**(3): 427-437.

193. Rodriguez, D., Esteban, M., Rodriguez, J.R. (1995). Vaccinia virus A17L gene product is essential for an early step in virion morphogenesis. *J.Virol.* **69**(8): 4640-4648.

194. Rodríguez, J.R., Risco, C., Carrascosa, J.L., Esteban, M., Rodríguez, D. (1997). Characterization of early stages in vaccinia virus membrane biogenesis: implications of the 21-kilodalton protein and a newly identified 15-kilodalton envelope protein. *Journal of Virology.* **71**(3): 1821-1833.

195. Rodríguez, J.R., Risco, C., Carrascosa, J.L., Esteban, M., Rodríguez, D. (1998). Vaccinia virus 15-kilodalton (A14L) protein is essential for assembly and attachment of viral crescents to virosomes. *Journal of Virology.* **72**(2): 1287-1296.

196. Rohl, C.A., Strauss, C.E., Chivian, D., Baker, D. (2004). Modeling structurally variable regions in homologous proteins with rosetta. *Proteins.* **55**(3): 656-677.

197. Roper, R.L., Wolffe, E.J., Weisberg, A., Moss, B. (1998). The envelope protein encoded by the A33R gene is required for formation of actin-containing microvilli and efficient cell-to-cell spread of vaccinia virus. *Journal of Virology.* **72**(5): 4192-4204.

198. Rosales, R., Harris, N., Ahn, B.Y., Moss, B. (1994). Purification and identification of a vaccinia virus-encoded intermediate stage promoter-specific transcription factor that has homology to eukaryotic transcription factor SII (TFIIS) and an additional role as a viral RNA polymerase subunit. *The Journal of Biological Chemistry.* **269**(19): 14260-14267.

199. Rudner, R., Karkas, J.D., Chargaff, E. (1968). Separation of B. subtilis DNA into complementary strands, I. Biological properties. *Proceedings of the National Academy of Science of the United States of America.* **60**(2): 630-635.

200. Sahu, A., Isaacs, S.N., Soulika, A.M., Lambris, J.D. (1998). Interaction of vaccinia virus complement control protein with human complement proteins: factor I-mediated degradation of C3b to iC3b1 inactivates the alternative complement pathway. *Journal of Immunology.* **160**(11): 5596-5604.

201. Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H., Hakoshima, T. (2004). Structural basis for recruitment of human flap endonuclease 1 to PCNA. *The EMBO Journal.* **24**(4): 683-693.

202. 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.

203. Sanderson, C.M., Frischknecht, F., Way, M., Hollinshead, M., Smith, G.L. (1998). Roles of vaccinia virus EEV-specific proteins in intracellular actin tail formation and low pH-induced cell-cell fusion. *The Journal of General Virology.* **79 (Pt 6)**: 1415-1425.

204. Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Barrell, B.G., Brown, N.L., Fiddes, J.C., Hutchison, C.A.3., Slocombe, P.M., Smith, M. (1978). The nucleotide sequence of bacteriophage phiX174. *Journal of Molecular Biology.* **125**(2): 225-246.

205. Sanz, P., Moss, B. (1999). Identification of a transcription factor, encoded by two vaccinia virus early genes, that regulates the intermediate stage of viral gene expression. *Proceedings of the National Academy of Science of the United States of America.* **96**(6): 2692-2697.

206. Schäffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V., Altschul, S.F. (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Research.* **29**(14): 2994-3005.

207. Schmidt, H., Hensel, M. (2004). Pathogenicity islands in bacterial pathogenesis. *Clinical Microbiology Reviews.* **17**(1): 14-56.

208. Schnierle, B.S., Gershon, P.D., Moss, B. (1992). Cap-specific mRNA (nucleoside-O2'-)-methyltransferase and poly(A) polymerase stimulatory activities of vaccinia virus are mediated by a single protein. *Proceedings of the National Academy of Science of the United States of America.* **89**(7): 2897-2901.

209. Schorman, N., Grigorian, A., Samal, A., Krishnan, R., Delucas, L., Chattopadhyay, D. (2007). Crystal structure of vaccinia virus uracil-DNA glycosylase reveals dimeric assembly. *BMC Structural Biology.* **7**(1): 45.

210. Schramm, B., Locker, J.K. (2005). Cytoplasmic organization of POXvirus DNA replication. *Traffic.* **6**(10): 839-846.

211. Schultz-Norton, J.R., Gabisi, V.A., Ziegler, Y.S., McLeod, I.X., Yates, J.R., Nardulli, A.M. (2007). Interaction of estrogen receptor {alpha} with proliferating cell nuclear antigen. *Nucleic Acids Research.* .

212. Schwer, B., Visca, P., Vos, J.C., Stunnenberg, H.G. (1987). Discontinuous transcription or RNA processing of vaccinia virus late messengers results in a 5' poly(A) leader. *Cell.* **50**(2): 163-169.

213. Senkevich, T.G., Ojeda, S., Townsley, A., Nelson, G.E., Moss, B. (2005). Poxvirus multiprotein entry-fusion complex. *Proceedings of the National Academy of Science of the United States of America.* **102**(51): 18572-18577.

214. Senkevich, T.G., Ward, B.M., Moss, B. (2004). Vaccinia virus A28L gene encodes an essential protein component of the virion membrane with intramolecular disulfide bonds formed by the viral cytoplasmic redox pathway. *Journal of Virology.* **78**(5): 2348-2356.

215. Senkevich, T.G., Ward, B.M., Moss, B. (2004). Vaccinia virus entry into cells is dependent on a virion surface protein encoded by the A28L gene. *Journal of Virology.* **78**(5): 2357-2366.

216. Senkevich, T.G., White, C.L., Koonin, E.V., Moss, B. (2002). Complete pathway for protein disulfide bond formation encoded by poxviruses. *Proceedings of the National Academy of Science of the United States of America.* **99**(10): 6667-6672.

217. Senkevich, T.G., White, C.L., Weisberg, A., Granek, J.A., Wolffe, E.J., Koonin, E.V., Moss, B. (2002). Expression of the vaccinia virus A2.5L redox protein is required for virion morphogenesis. *Virology.* **300**(2): 296-303.

218. Shchelkunov, S.N., Blinov, V.M., Resenchuk, S.M., Totmenin, A.V., Olenina, L.V., Chirikova, G.B., Sandakhchiev, L.S. (1994). Analysis of the nucleotide sequence of 53 kbp from the right terminus of the genome of variola major virus strain India-1967. *Virus Research.* **34**(3): 207-236.

219. Shchelkunov, S.N., Resenchuk, S.M., Totmenin, A.V., Blinov, V.M., Marennikova, S.S., Sandakhchiev, L.S. (1993). Comparison of the Genetic Maps of Variola and Vaccinia Viruses. *FEBS Letters.* **327**(3): 321-324.

220. Shchelkunov, S.N., Safronov, P.F., Totmenin, A.V., Petrov, N.A., Ryazankina, O.I., Gutorov, V.V., Kotwal, G.J. (1998). The genomic sequence analysis of the left and right species-specific terminal region of a cowpox virus strain reveals unique sequences and a cluster of intact ORFs for immunomodulatory and host range proteins. *Virology.* **243**(2): 432-460.

221. Shchelkunov, S.N., Totmenin, A.V., Sandakhchiev, L.S. (1996). Analysis of the nucleotide sequence of 23.8 kbp from the left terminus of the genome of variola major virus strain India-1967. *Virus Research.* **40**(2): 169-183.

222. Shen, B., Singh, P., Liu, R., Qiu, J., Zheng, L., Finger, L.D., Alas, S. (2005). Multiple but dissectible functions of FEN-1 nucleases in nucleic acid processing, genome stability and diseases. *Bioessays.* **27**(7): 717-729.

223. Shors, S.T., Beattie, E., Paoletti, E., Tartaglia, J., Jacobs, B.L. (1998). Role of the vaccinia virus E3L and K3L gene products in rescue of VSV and EMCV from the effects of IFN-alpha. *Journal of Interferon & Cytokine Research : The Official Journal of the International Society for Interferon and Cytokine Research.* **18**(9): 721-729.

224. Shuman, S. (1991). Site-specific DNA cleavage by vaccinia virus DNA topoisomerase I. Role of nucleotide sequence and DNA secondary structure. *The Journal of Biological Chemistry.* **266**(3): 1796-1803.

225. Sieczkarski, S.B., Whittaker, G.R. (2005). Viral entry. *Current Topics in Microbiology and Immunology.* **285**: 1-23.

226. Simpson, D.A., Condit, R.C. (1995). Vaccinia virus gene A18R encodes an essential DNA helicase. *Journal of Virology.* **69**(10): 6131-6139.

227. Slabaugh, M.B., Davis, R.E., Roseman, N.A., Mathews, C.K. (1993). Vaccinia Virus Ribonucleotide Reductase Expression and Isolation of the Recombinant Large Subunit. *The Journal of Biological Chemistry.* **268**(24): 17803-17810.

228. Smith, C., Davis, T., Anderson, D., Solam, L., Beckmann, M., Jerzy, R., Dower, S., Cosman, D., Goodwin, R. (1990). A receptor for tumor necrosis factor defines an unusual family of cellular and viral proteins. *Science.* **248**(4958): 1019-1023.

229. Smith, C.A., Davis, T., Anderson, D., Solam, L., Beckmann, M.P., Jerzy, R., Dower, S.K., Cosman, D., Goodwin, R.G. (1990). A receptor for tumor necrosis factor defines an unusual family of cellular and viral proteins. *Science.* **248**: 1019-1023.

230. Smith, G.L., Chan, Y.S. (1991). Two vaccinia virus proteins structurally related to the interleukin-1 receptor and the immunoglobulin superfamily. *The Journal of General Virology.* **72 (Pt 3)**: 511-518.

231. Smith, G.L., Murphy, B.J., Law, M. (2003). Vaccinia virus motility. *Annual Review of Microbiology.* **57**: 323-342.

232. Smith, G.L., Vanderplasschen, A., Law, M. (2002). The formation and function of extracellular enveloped vaccinia virus. *The Journal of General Virology.* **83**(Pt 12): 2915-2931.

233. Smith, V.P., Bryant, N.A., Alcamí, A. (2000). Ectromelia, vaccinia and cowpox viruses encode secreted interleukin-18-binding proteins. *The Journal of General Virology.* **81**(Pt 5): 1223-1230.

234. Sodeik, B., Krijnse-Locker, J. (2002). Assembly of vaccinia virus revisited: de novo membrane synthesis or acquisition from the host? *Trends in Microbiology.* **10**(1): 15-24.

235. Sonnhammer, E.L., Durbin, R. (1995). A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene.* **167**(1-2): GC1-G10.

236. Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics.* **21**(7): 951-960.

237. Söding, J., Biegert, A., Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research.* **33**(Web Server issue): W244-W248.

238. Stanitsa, E.S., Arps, L., Traktman, P. (2006). Vaccinia virus uracil DNA glycosylase interacts with the A20 protein to form a heterodimeric processivity factor for the viral DNA polymerase. *The Journal of Biological Chemistry.* **281**(6): 3439-3451.

239. Stuart, D.T., Upton, C., Higman, M.A., Niles, E.G., McFadden, G. (1993). A poxvirus-encoded uracil DNA glycosylase is essential for virus viability. *Journal of Virology.* **67**(5): 2503-2512.

240. Stuckey, J.A., Dixon, J.E. (1999). Crystal structure of a phospholipase D family member. *Nature Structural Biology.* **6**(3): 278-284.

241. Syed, A., Upton, C. (2006). Java GUI for InterProScan (JIPS): a tool to help process multiple InterProScans and perform ortholog analysis. *BMC Bioinformatics.* **7**: 462.

242. Szybalski, W., Kubinski, H., Sheldrick, P. (1966). Pyrimidine clusters on the transcribing strand of DNA and their possible role in the initiation of RNA synthesis. *Cold Spring Harbor Symposia on Quantitative Biology.* **31**: 123-127.

243. Taddeo, B., Esclatine, A., Roizman, B. (2002). The patterns of accumulation of cellular RNAs in cells infected with a wild-type and a mutant herpes simplex virus 1 lacking the virion host shutoff gene. *Proceedings of the National Academy of Science of the United States of America.* **99**(26): 17031-17036.

244. Thomas, J.M., Horspool, D., Brown, G., Tcherepanov, V., Upton, C. (2007). GraphDNA: a Java program for graphical display of DNA composition analyses. *BMC Bioinformatics.* **8**: 21.

245. Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.* **22**(22): 4673-4680.

246. Tolonen, N., Doglio, L., Schleich, S., Krijnse Locker, J. (2001). Vaccinia virus DNA replication occurs in endoplasmic reticulum-enclosed cytoplasmic mini-nuclei. *Molecular Biology of the Cell.* **12**(7): 2031-2046.

247. Tom, S., Henricksen, L.A., Bambara, R.A. (2000). Mechanism whereby proliferating cell nuclear antigen stimulates flap endonuclease 1. *The Journal of Biological Chemistry.* **275**(14): 10498-10505.

248. Traktman, P., Liu, K., DeMasi, J., Rollins, R., Jesty, S., Unger, B. (2000). Elucidating the essential role of the A14 phosphoprotein in vaccinia virus morphogenesis: construction and characterization of a tetracycline-inducible recombinant. *Journal of Virology.* **74**(8): 3682-3695.

249. Tsurimoto, T. (1998). PCNA, a multifunctional ring on DNA. *Biochimica et Biophysica Acta.* **1443**(1-2): 23-39.

250. Tulman, E.R., Delhon, G., Afonso, C.L., Lu, Z., Zsak, L., Sandybaev, N.T., Kerembekova, U.Z., Zaitsev, V.L., Kutish, G.F., Rock, D.L. (2006). Genome of horsepox virus. *Journal of Virology.* **80**(18): 9244-9258.

251. Twardzik, D.R., Brown, J.B., Ranchalis, J.E., Todaro, G.J., Moss, B. (1985). Vaccinia virus-infected cells release a novel polypeptide functionally related to transforming and epidermal growth factors. *Proceedings of the National Academy of Science of the United States of America.* **82**(16): 5300-5304.

252. Ulaeto, D., Grosenbach, D., Hruby, D.E. (1996). The vaccinia virus 4c and A-type inclusion proteins are specific markers for the intracellular mature virus particle. *Journal of Virology.* **70**(6): 3372-3377.

253. Upton, C. (2000). Screening predicted coding regions in poxvirus genomes. *Virus Genes.* **20**(2): 159-164.

254. Upton, C., Hogg, D., Perrin, D., Boone, M., Harris, N.L. (2000). Viral genome organizer: a system for analyzing complete viral genomes. *Virus Research.* **70**(1-2): 55-64.

255. Upton, C., Mossman, K., McFadden, G. (1992). Encoding of a homolog of the IFN-gamma receptor by myxoma virus. *Science.* **258**: 1369-1373.

256. Upton, C., Slack, S., Hunter, A.L., Ehlers, A., Roper, R.L. (2003). Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *Journal of Virology.* **77**(13): 7590-7600.

257. Upton, C., Stuart, D.T., McFadden, G. (1993). Identification of a poxvirus gene encoding a uracil DNA glycosylase. *Proceedings of the National Academy of Science of the United States of America.* **90**: 4518-4522.

258. Vanderplasschen, A., Hollinshead, M., Smith, G.L. (1998). Intracellular and extracellular vaccinia virions enter cells by different mechanisms. *The Journal of General Virology.* **79 (Pt 4)**: 877-887.

259. Vázquez, M.I., Esteban, M. (1999). Identification of functional domains in the 14-kilodalton envelope protein (A27L) of vaccinia virus. *Journal of Virology.* **73**(11): 9098-9109.

260. Wallengren, K., Risco, C., Krijnse-Locker, J., Esteban, M., Rodriguez, D. (2001). The A17L gene product of vaccinia virus is exposed on the surface of IMV. *Virology.* **290**(1): 143-152.

261. Warbrick, E. (2000). The puzzle of PCNA's many partners. *Bioessays.* **22**(11): 997-1006.

262. Wilcock, D., Duncan, S.A., Traktman, P., Zhang, W.H., Smith, G.L. (1999). The vaccinia virus A4OR gene product is a nonstructural, type II membrane glycoprotein that is expressed at the cell surface. *The Journal of General Virology.* **80 (Pt 8)**: 2137-2148.

263. Wilson, V.G., Rangasamy, D. (2001). Viral interaction with the host cell sumoylation system. *Virus Research.* **81**(1-2): 17-27.

264. Wolffe, E.J., Katz, E., Weisberg, A., Moss, B. (1997). The A34R glycoprotein gene is required for induction of specialized actin-containing microvilli and efficient cell-to-cell transmission of vaccinia virus. *Journal of Virology.* **71**(5): 3904-3915.

265. Wolffe, E.J., Moore, D.M., Peters, P.J., Moss, B. (1996). Vaccinia virus A17L open reading frame encodes an essential component of nascent viral membranes that is required to initiate morphogenesis. *Journal of Virology.* **70**(5): 2797-2808.

266. Wolffe, E.J., Weisberg, A.S., Moss, B. (1998). Role for the vaccinia virus A36R outer envelope protein in the formation of virus-tipped actin-containing microvilli and cell-to-cell virus spread. *Virology.* **244**(1): 20-6..

267. Wood, R.D. (1996). DNA repair in eukaryotes. *Annual Review of Biochemistry.* **65**: 135-167.

268. Wright, C.F., Coroneos, A.M. (1993). Purification of the late transcription system of vaccinia virus: identification of a novel transcription factor. *Journal of Virology.* **67**(12): 7264-7270.

269. Wright, C.F., Oswald, B.W., Dellis, S. (2001). Vaccinia virus late transcription is activated in vitro by cellular heterogeneous nuclear ribonucleoproteins. *The Journal of Biological Chemistry.* **276**(44): 40680-40686.

270. Xiang, Y., Latner, D.R., Niles, E.G., Condit, R.C. (2000). Transcription elongation activity of the vaccinia virus J3 protein in vivo is independent of poly(A) polymerase stimulation. *Virology.* **269**(2): 356-369.

271. Xiang, Y., Moss, B. (1999). IL-18 binding and inhibition of interferon gamma induction by human poxvirus-encoded proteins. *Proceedings of the National Academy of Science of the United States of America.* **96**(20): 11537-11542.

272. Xiang, Y., Moss, B. (1999). Identification of human and mouse homologs of the MC51L-53L-54L family of secreted glycoproteins encoded by the Molluscum contagiosum poxvirus. *Virology.* **257**(2): 297-302.

273. Xiang, Y., Simpson, D.A., Spiegel, J., Zhou, A., Silverman, R.H., Condit, R.C. (1998). The vaccinia virus A18R DNA helicase is a postreplicative negative transcription elongation factor. *Journal of Virology.* **72**(9): 7012-7023.

274. Yao, X.D., Evans, D.H. (2001). Effects of DNA structure and homology length on vaccinia virus recombination. *Journal of Virology.* **75**(15): 6923-6932.

275. Yeh, W.W., Moss, B., Wolffe, E.J. (2000). The vaccinia virus A9L gene encodes a membrane protein required for an early step in virion morphogenesis. *Journal of Virology.* **74**(20): 9701-9711.

276. Yoder, J.D., Chen, T.S., Gagnier, C.R., Vemulapalli, S., Maier, C.S., Hruby, D.E. (2006). Pox proteomics: mass spectrometry analysis and identification of Vaccinia virion proteins. *Virology Journal.* **3**(1): 10.

277. Zdobnov, E.M., Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics.* **17**(9): 847-848.

278. Zhang, Y., Keck, J.G., Moss, B. (1992). Transcription of viral late genes is dependent on expression of the viral intermediate gene G8R in cells infected

with an inducible conditional-lethal mutant vaccinia virus. *Journal of Virology*. **66**(11): 6470-6479.

279. Zhang, Y., Keck, J.G., Moss, B. (1992). Transcription of viral late genes is dependent on expression of the viral intermediate gene G8R in cells infected with an inducible conditional-lethal mutant vaccinia virus. *Journal of Virology*. **66**(11): 6470-6479.

280. Zhang, Y.F., Moss, B. (1991). Vaccinia virus morphogenesis is interrupted when expression of the gene encoding an 11-kilodalton phosphorylated protein is prevented by the Escherichia coli lac repressor. *Journal of Virology*. **65**(11): 6101-6110.

281. Zuccola, H.J., Filman, D.J., Coen, D.M., Hogle, J.M. (2000). The crystal structure of an unusual processivity factor, herpes simplex virus UL42, bound to the C terminus of its cognate polymerase. *Molecular Cell*. **5**(2): 267-278.