THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF VETERINARY AND BIOMEDICAL SCIENCES

CHARACTERIZATION OF A NOVEL ENDOGENOUS RETROVIRUS ENVELOPE PROTEIN IN NORTH AMERICAN MULE DEER

JASON BANKERT
Spring 2012

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Immunology & Infectious Disease and Toxicology
with honors in Immunology & Infectious Disease

Reviewed and approved* by the following:

Mary Poss, PhD/DVM
Professor in Biology
and in Veterinary and Biomedical Sciences
Thesis Supervisor

James Howell, PhD
Program Coordinator IID/Tox and
Coordinator of Undergraduate Advising
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Endogenous retroviruses are present in many animals, including humans, and represent ancient invasions of the host germ line following infection events. The proviruses will be passed vertically if they are not deleterious to the host. We have described a new family of gammaretroviruses in mule deer called CrERVγ. These encode intact genes and are capable of expressing viral proteins in the cell. We have characterized and expressed the envelope glycoprotein from CrERVγ cytoplasmically and have compared it to the envelope glycoproteins of four other gammaretroviruses. The physical properties as well as structural characteristics and phylogenetic studies were analyzed to understand how CrERVγ relates to the other gammaretroviruses. We show that the envelope glycoprotein in CrERVγ shares common motifs found in all gammaretroviruses and also show patterns in N-glycosylation sites, secondary structures and physical properties that are both common to the other gammaretroviruses and also unique to CrERVγ.

TABLE OF CONTENTS

Abstract……………………………………………………………………………...i
List of Figures……………………………………………………………………iii
Acknowledgements………………………………………………………………iv

INTRODUCTION………………………………………………….....................1
Endogenous retroviruses…………………………………………………...1
CrERV………………………………………………………………………...2
Envelope glycoprotein……………………………………………………..2
Antigenicity of the envelope glycoprotein…………………………………....3
Hypothesis……………………………………………………………………4

MATERIALS AND METHODS………………………………..…………………4

RESULTS……………………………………………………………………………..9
Phylogeny ……………………………………………………………………9
Physical properties…………………………………………………………..10
Secondary structure…………………………………………………………12
Transmembrane determination……………………………………………...14
Motifs………………………………………………………………………..15
N-glycosylation……………………………………………………………...16
Transfection …………………………………………………………………17
Amino acid changes in the clone……………………………………………18

DISCUSSION………………………………………………………………………..19

CONCLUSION……………………………………………………………………….22

References……………………………………………………………………………23

LIST OF FIGURES

Figure 1: Phylogenetic analysis of CrERVγ, GaLV, FeLV, MoMuLV and BLV

Figure 2: Physical features analysis of CrERVγ, GaLV, FeLV, MoMuLV and BLV

Figure 3: Secondary structure analysis of CrERVγ, GaLV, FeLV, MoMuLV and BLV

Figure 4: Transmembrane determination in CrERVγ, GaLV, FeLV, MoMuLV and BLV

Figure 5: Motif analysis of CrERVγ, GaLV, FeLV, MoMuLV and BLV

Figure 6: N-glycosylation analysis of CrERVγ, GaLV, FeLV, MoMuLV and BLV

Figure 7: Transfection results

Figure 8: Amino acid changes in the clone

AKNOWLEDGEMENTS

I would like to thank God for giving me this opportunity and for guiding me through this process even when it seemed impossible. I would also like to thank Dr. Mary Poss for her unending patience and compassion and drive. She took me under her wing throughout this project. I would also like to like Dr. Oekyung Kim for doing the transfection in chicken cells and helping me with the immunostaining and other procedures. I would like to thank Dr. Daniel Elleder for his support and teaching and fellowship. I would also like to thank Yee Ling Chong and Gina Riggio for their support, help and teaching.

INTRODUCTION

Endogenous retroviruses integrate into the host genome as a provirus and if in the germ line, they can be passed to progeny via Mendelian inheritance so that all progeny of the initially infected individual contain the provirus in their genome (Weiss 2006). If deleterious to the offspring there will be no further evidence of the integration event. For those viruses that are neutral or have a beneficial effect on the host, the provirus becomes fixed into the genome and accumulates mutations that may eventually renders it non-infectious. Some endogenous retroviruses, such as Murine leukemia virus (MLV) related viruses, have remained viable for millions of years (Weiss, 2006). Endogenous retroviruses have gained particular attention due to their carcinogenesis and immunodeficiency potential. Studies with gammaretroviruses MLV and FeLV-B have shown that these viruses can insert into the genome next to proto-oncogenes and transform cells. Recombinant MLVs and recombinant FeLVs have been found to have malignant and immunodegenerative effects in mice and cats, respectively (Rosenberg & Jolicoeur 1997) (Weiss 2006) (Elder, et. al 1997). A MLV-related virus, called Xenotropic Murine Leukemia Virus-related Virus (XMRV), is proposed to be present in humans and has been isolated in prostate cancer tissues, in lymphocytes and in chronic fatigue patients, however, there are conflicting hypotheses about their pathogenic effects in humans (Qiu, et. al 2010) (Knox, et. al 2011). In addition, the immune response mounted against viral protein antigens, in particular neutralizing antibodies produced against the envelope protein, has been useful as markers for disease as well as in vaccine development (Qui, et. al 2010)(Stephensen, et. al 1977)(Langhammer, et. al 2006).

A novel endogenous gammaretrovirus (CrERVγ) has recently been characterized in North American mule deer (*Odocoileus hemionus*) located in Montana and surrounding states (Elleder, et. al 2012). We have characterized 7 total proviruses. PCR-based technology and southern blot have shown polymorphism in integration sites among these proviruses, which suggests that the integrations are relatively recent. One provirus, called In55, has open reading frames (ORFs) for all three viral proteins: gag, pol and env (Elleder, et. al 2012). This suggests that the provirus has the potential to produce functional viral proteins, such as the envelope glycoprotein in the host cells. Although In55 have intact ORFs, a viral particle has not been isolated. However we predict that an exogenous form of the virus could exist due to the presence of unique, relatively recent integration sites in the mule deer population. This project involved description of the properties of the envelope glycoprotein from CrERVγ and efforts to express it in a cell line.

The surface of a retrovirus is studded with the envelope glycoprotein. It is a transmembrane protein that plays an important role in fusion with and entry into the host cell. The envelope protein is associated with the host cell plasma membrane and can be visualized by microscopy using immunocytochemistry. It is a glycosylated oligomer. For instance, the FeLV envelope glycoprotein is made up of a complex of 4 to 6 copies of the outer membrane protein, gp70 molecules linked to the transmembrane protein, p15E via disulfide linkages (Youngren, et. al 1984). There are distinct domains on the viral envelope glycoprotein with each domain having a distinct function that allows the virus to bind and fuse with the host cell membrane. The Receptor Binding Domain (RBD), about 200-250 amino acids, located on the N-terminus is critical for host cell receptor binding and specificity. The RBD is conserved in gammaretroviruses but contains hypervariable regions: VRA, VRB, VRC that control receptor binding and specificity. The first loop in VRA has been determined to have critical residues involved in receptor recognition and binding (Tailor & Kabat, 1997). Gammaretroviruses contain a conserved motif in the RBD called the PHQ motif that has been, shown to be critical in membrane fusion. Of this motif, the H residue seems to be the most important for binding and fusion and a loss of this residue of the PHQ motif significantly diminishes infectivity (Rey, et. al, 2007) (Lavillette & Kabat 2004). The RBD is anchored to the C terminus C domain (Cdom). Cdom has been shown to be involved in receptor recognition and membrane fusion. (Rey, et. al 2007) (Barnett & Cunningham, 2001). Binding of the RBD to the host receptor causes a conformational change in the envelope glycoprotein, brought about by the proline-rich region that brings the fusion peptide of the transmembrane region into contact with the host cell membrane. In the ectoderm of the transmembrane region, C-terminal to the fusion peptide is the HTLV-1 like heptad repeat 1-heptad repeat 2 subdomain (HTLV-1-like_HR1-HR2**)**, which spans from the N-terminal heptad repeat (HR1) to the C-terminal heptad repeat (HR2). This region is present in most gammaretroviruses and includes an immunosuppressive region, a C(X)6 motif involved in intrasubunit disulfide bond formation and a chloride ion binding site (Kobe, et. al 1999). Studies have shown that the immunosuppressive region is highly conserved in all retroviruses (Denner 1998). In viral fusion, the three HR1 helices and three HR2 helices align in an antiparallel orientation (Marchler-Bauer, et al, 2011).

The three dimensional structure of each domain of the envelope glycoprotein is critical for its function. This is shown in assays involving mutagenesis and truncations in the Moloney murine leukemia virus (MoMLV*)*. In these studies, it was discovered that the receptor-binding region is amino acids 9 to 230 and within the region are two basic amino acids, Arg-83 and Arg-95, which play an

important role in interaction with the receptor (Bae, et. al 1997). In addition, post-translational modifications to the envelope glycoprotein are critical, namely N-linked glycosylations. These N-linked glycosylation sites have been found to be important in binding to the host cell receptor and also for translocation of the envelope glycoprotein to the membrane and proper folding of the protein (Knoper, et. al, 2009) (Battini, et. al 1994). In N-linked glycosylation, mannose-rich, dolichol-linked oligosaccharides are added to the asparagine residues with a recognition sequence: Asn-(X)-Ser (Thr) (Pinter & Honnen, 1987). After glycosylation, the envelope glycoprotein often undergoes a maturation process involving glycosidases and glycosyltransferases in the golgi apparatus. A second type of glycosylation takes place in the golgi, O-linked glycosylation. This involves the step-wise addition of sugars to clusters of serine and threonine residues on hydroxyl groups via glycosyltransferases. Also in the golgi, the envelope glycoprotein is cleaved into the extracellular surface subunit and transmembrane component (Pinter & Honnen, 1987).

The three-dimensional structure of the protein is also important for antigenicity. It is hypothesized that most antibodies bind near the receptor binding region, neutralizing the virus (Youngren, et. al 1984), however other regions could also be targeted by antibodies and antigenicity varies between different sites or regions of the envelope glycoprotein. In FeLV, an exogenous virus, antibodies were discovered to bind several regions on the g70 outer membrane protein and on the p15E transmembrane protein and neutralizing sites are thought to be on g70 (Elder et. al 1987) (Langhammer, et. al 2006) (Fiebig, et. al 2003). Structural changes in both the g70 and p15E peptides can result in masking of neutralizing epitopes in certain subtypes of FeLV. In addition, even though carbohydrate moieties are not the targets of neutralizing antibodies, altered glycosylation patterns can alter antibody recognition and binding (Elder, et. al 1987). The different regions on the envelope glycoprotein show different levels of antigenic variability. Some regions are highly polymorphic and display high antigenic variability, with isolates of the same subtype of virus showing different antigenic variation. This is seen in both FeLV and bovine leukemia virus (Youngren, et. al 1984) (Bruck, et. al 1982). This variability provides a protection mechanism from the host immune components. This is especially true for FeLV, which is spread through exogenous infection and is thus exposed to the host immune system (Elder, et. al 1987). Alternatively, other regions are quite conserved and could elicit antibody cross-reactivity between different species of gammaretroviruses. Studies have confirmed the presence of significant homologies in the envelope glycoprotein of different gammaretroviruses. For example, in the endogenous virus, porcine endogenous retrovirus (PERV) two epitopes have been discovered: E1 on the C-terminus and E2 on the

N-terminus, which are identical in all PERVs and are highly conserved among all gammaretroviruses (Fiebig, et. al 2003). An antigenic peptide derived from the C-terminus of p15E of FeLV was found to contain sequences similar to the E2 epitope and an antigenic peptide from the N-terminus near the fusion peptide of FeLV was found to contain sequences related to the E1 epitope, stressing the conserved nature of these epitopes. These regions were also found to be partly conserved in lentiviruses and antibodies specific for HIV-1 strains with similar epitopes as E2 are shown to neutralize several strains of HIV-1 (Fiebig, et. al 2003).

CrERVγ In55 is a novel endogenous retrovirus with no evidence of having an exogenous form or being infectious. It does have open reading frame for all major viral genes, gag, pol and env. This paper will focus on the env gene and it's relation with other gammaretroviral envelope glycoproteins. Comparisons of the envelope glycoprotein were done with other gammaretroviruses to identify hared features and properties. Because of the uninterrupted open reading frame, I also determined if the envelope glycoprotein could be expressed and properly processed in cell culture.

MATERIALS AND METHODS

PCR Amplification of In55. A PCR amplification of genomic DNA from mule deer sample # 373 was done using Ex Taq DNA polymerase. A 50 µL PCR reaction mixture was prepared as follows:

Reagent	Volume (µL)
10x Ex Taq Buffer	5.0
dNTPs	4.0
MDfor1 (forward primer)	2.0
In55d (reverse primer)	2.0
Ex Taq Polymerase	0.25
Water	36.25
Genomic DNA	0.50

The PCR program was run in the following order and conditions:

Temperature (°C)	Time	# of Cycles
95	2 mins	1
94	15 sec	32
61	30 sec	32
68	5 mins	32
68	10 mins	1

PCRAmplification of the 2-kb Env Insert. PCR amplification of a 2-kb insert containing the env gene with restriction sites for EcoRI and BglII was done with Ex Taq DNA polymerase; using the amplified In55 template. The following primers, designed using DNAStar Lasergene software (DNA Star 2012),were used to amplify the region of In55 containing the envelope glycoprotein gene. The primers were synthesized by the Pennsylvania State University Genomics Core Facility-University Park and were designed to contain the indicated restriction sites to facilitate cloning.

MDEcoRF: 5'-AGGGAATTCCAGAATGGAAGGCGAA-3'

MDaBgl2: 5'- GCAGATCTCTTTTGATTATGGGGAG-3'

Three 50µL nested PCR reaction mixtures were prepared as follows:

Reagent	Volume (µL)
10x Ex Taq Buffer	16.5
dNTPs	13.2
MDEcoRF (synthesized forward primer)	6.6
MDaBgl2R (synthesized reverse primer)	6.6
Ex Taq Polymerase	0.83
Water	119.63
Amplified In55 PCR product (44.5ng/ µL)	0.50

The PCR program was run in the following order and conditions:

Temperature (°C)	Time	# of Cycles
95	2 mins	1
95	15 sec	3
53	25 sec	3
68	2 mins	3
95	15 sec	25
56	25 sec	25
68	7 mins	25
60	20 mins	1

Amplified 2-kb env PCR products were resolved on a 1% agarose gel for ~4 hours at 50V. The resulting 2- kb bands containing env with restriction sites for EcoRI and BglII was cut and gel isolated using a Gel Isolation kit (Qiagen 2012).

Restriction digest reaction. A double digest reaction was performed with EcoR1 and BglII restriction endonucleases in a 10x EcORI buffer creating sticky ends for annealing of the 2-kb PCR product containing the envelope glycoprotein gene into the 6.4 kb expression vector (p3XFLAG-CMV-9) (Sigma Aldrich 2012). A 40-µL reaction mixture for the double digest reaction was setup as follows:

Reagent	Volume (µL)
10x EcoRI Buffer	4.0
EcoRI restriction endonuclease (20U/µL)	1.0
BglII restriction endonuclease (10U/µL)	1.5
2-kb insert containing env (76 ng/ µL)	30
Water	3.5

The double digest mixture was let sit to react for over 5 hrs. in a hot water bath at 37°C. The digested 2-kb env fragment was then purified using the Qiagen PCR Purification Kit (Qiagen 2012).

Ligation. The purified, digested 4-kb insert containing env (42 ng/µL) was ligated into the digested expression vector p3XFLAG-CMV-9) (13 ng/µL) (Sigma Aldrich 2012) using T4 DNA Ligase in the T4 DNA Ligase buffer. Three ligation mixtures were setup as: 1. Negative control (vector only); 2. 5:1 insert: vector ratio; 3. 3:1 insert: vector ratio.

Mixture #	Vector (µL)	Insert (µL)	Water (µL)	T4 DNA Ligase Buffer (µL)	T4 DNA Ligase (µL)
1 (negative)	1.0	-	7.6	1.0	0.40
2	1.0	1.4	6.2	1.0	0.40
3	1.0	0.87	6.7	1.0	0.40

Transformation. The ligated mixtures were transformed into 50µL of thawed DH5alpha cells. 3µL of each ligation mixture was added to 50µL of thawed DH5alpha cells. The cells were plated on Carbenacillin plates and incubated overnight at 37°C. The Qiagen Mini-Prep kit was used to isolate the vector from three positive colonies on each plate to total six mini-preps (Qiagen 2012).

Restriction digest screen. A double digest reaction with EcoRI and Bgl2R in 10x EcoRI buffer was performed to screen the positive colonies to determine if the insert and vector are present. The insert + FLAG tag peptide was cleaved at the EcoRI and Bgl2 sites, giving a distinct band at ~3kb when the digest reactions are resolved on 1% agarose gel. 10-µL digest reactions were set up as follows:

Reagent	Volume (µL)
10x EcoRI Buffer	1.0
EcoRI restriction endonuclease (20U/µL)	0.1
BglII restriction endonuclease (10U/µL)	0.1
DNA product from transformed DH5α cells, isolated via Mini-Prep	1.0
Water	7.8

Sequence Data: All sequences were sent to the Pennsylvania State University Genomics Core Facility-University Park.

Transfection. DF-1 (chicken fibroblast) cells were seeded at $1X10^5$ cells per well in 12 well plate. Cells were transfected with env plasmid using Lipofectamine 2000 (Invitrogen) in 50uL OptiMEM medium. DNA- Lipofectamine 2000 complexes (100 µl) were directly added to each well of the plates containing cells and incubated. One well contained the transfection control GFP; the negative control, the flag vector with no insert; the positive control, 704P; cells treated with the env expression vector. 1ug of the expression vector was used. The cells were fixed using 4% paraformaldehyde.

Immunostaining. The wells were washed 3x with PBS solution (300ul-400ul) to remove the paraformaldehyde. 1% BSA was used as the blocking solution. 0.3% TritonX-100 was used to permeabilize the cells. M2 anti-mouse monoclonal antibody (1 mg) was diluted 1:400 and used as the primary antibody. Anti-mouse IgG (2mg/mL) was diluted 1:400 and used as the secondary antibody. DAPZ was used to stain the nuclei. Cells were visualized using fluorescent microscopy.

Physical properties analysis. The length of the amino acid chain, molecular weight, pI index, the number of negatively charged residues, the number of positively charged residues, aliphatic index and the grand average of hydropathicity (GRAVY) in In55 compared to four other gammaretroviruses: FeLV, GaLV, MoMuLV and BLV was done using ProtParam software provided by the Swiss Institute of Bioinformatics (Gasteiger, et. al 2005)

Secondary structure analysis. Secondary structures: alpha, beta, coils and turns along the amino acid chains and hydrophobic regions of In55 and the four comparison gammaretroviruses was determined using DNA Star software (DNA Star, 2012).

Transmembrane determination. TMHMM software is provided by the Center for Biological Sequence Analysis (CBS) and is used to predict the location of transmembrane helices (CBS 2012). The data is reported in the probability of a transmembrane helix occurring at specific sites along the length of the amino acid sequence. The program predicts which regions are extracellular and which are intracellular based on these results.

Motif analysis, The sequence and location of the HTLV-1 HR1-HR2 region, immunosuppressive region, Cl binding site and the C(X)6 motif were determined using published amino acid sequence data for FeLV, MoMuLV, GaLV and BLV (Overbaugh, et. al, 1992)(Ting, et. al, 1998)(Dube, et. al, 2009)(Petropoulos & Chappey, 1997). The sequence and location of the proline-rich region and the PHQ motif was determined from Rey, et. al, 2007).

N-glycosylation site analysis. N-glycosylation sites were determined by NetNGlyc software from the Center for Biological Sequence Analysis (CBS). The sites of glycosylation, the specific Asn (X) Ser (Thr) residue, the potential of N-glycosylation and the predicted occurrence of N-glycosylation at the particular site are determined by this software (CBS 2012).

Phylogenetic analysis: The amino acid sequences from all viruses were aligned via MUSCLE and phylogenic analysis was performed on the aligned sequences using MEGA 5.05.

Accession numbers. All comparisons were done using the envelope glycoprotein amino acid sequences for CrERVγ (Accession number AEW89630.1), FeLV (Accession number AAA43050), GaLV (Accession number AAC96086.1), MoMuLV (Accession number AAL69911.1) and BLV (Accession number ACR15160.1).

RESULTS

Phylogenic Analysis

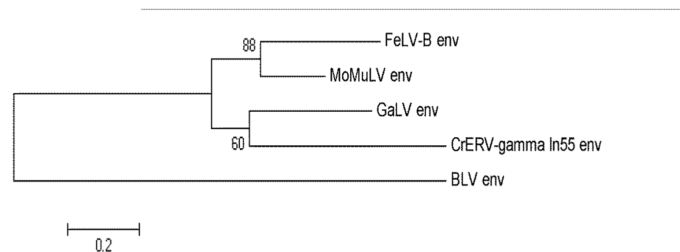


Figure 1: Phylogenetic analysis with bootstrapping of the envelope glycoprotein amino acid sequences using MEGA 5 software for CrERVγ In55, FeLV, GaLV, MoMuLV and BLV.

Phylogenetic analysis on the envelope glycoprotein from all viruses was performed using MEGA 5.05 software. It is shown that BLV is more distant than the other viruses, as it branches earlier in the lineage than In55 and the other gammaretroviruses. Then, CrERVγ In55 clusters with GaLV and MoMuLV is most closely related to FeLV. The physical and structural properties of the envelope proteins of these viruses were compared to see if they were determined by ancestral relationships.

Physical Properties Analysis

	CrERVγ In55	**GaLV**	**FeLV**	**MoMuLV**	**BLV**
Number of amino acids	647	670	642	665	515
Molecular weight (Da)	71990.4	74582.5	71079.6	73301.8	58580.5
pI index	7.50	8.66	8.98	8.62	8.78
Negatively charged residues (Asp + Glu)	62	57	46	48	34
Positively charged residues (Arg + Lys)	63	67	61	58	42
Aliphatic index	83.89	87.18	81.57	81.79	96.23
Grand average of hydropathicity (GRAVY)	-0.238	-0.261		-0.257	-0.120

Figure 2: Comparison of the number of amino acids, protein size and physical properties of the envelope glycoprotein in CrERVγ, FeLV, BLV, GaLV and MoMuLV.

Physical properties were determined using ProtParam software provided by the Swiss Institute of Bioinformatics Gasteiger, et. al 2005) (Figure 1). The length of the primary amino acid sequence and the molecular weights (measured in Daltons) of the envelope glycoprotein were determined for each virus. The length of all gammaretrovirus envelope glycoprotein is between 642-670 aa., except for that of BLV, which is only 515aa. In55 (647) is smaller than that of the most related virus, GaLV, (670). The number of negatively (Asp and Glu) and positively (Arg and Lys) charged residues, in the envelope glycoprotein was determined for each virus. In55 and GaLV have more negatively charged residues compared to the other viruses. The number of negatively charged residues can influence the subcellular location of the protein. The number of charged residues also impacts protein folding and interactions with other proteins and macromolecules. Meanwhile, the number of positively charged amino acids (Arg and Lys)

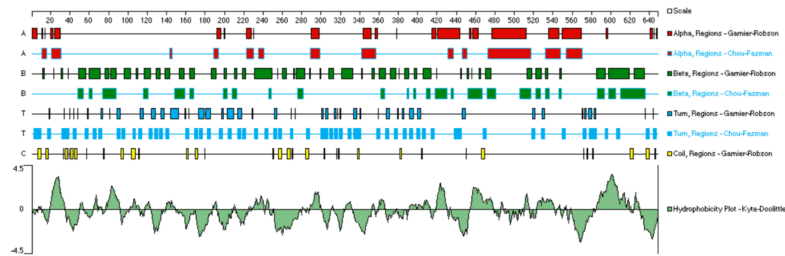
in In55 was similar to GaLV, FeLV and MoMuLV. The aliphatic index of the envelope glycoprotein for each virus was determined. These values measured the relative amount of space taken up by the aliphatic sides chains of alanine, valine, leucine and isoleucine and can be used to determine the thermostability of a protein (Ikai 1990). The aliphatic index of In55 is slightly less than GaLV at 87.18 and slightly higher than FeLV and MoMuLV, which have a value of 81.57 and 81.79, respectively, showing that it has slightly higher thermostability than FeLV or MoMuLVdue to the tightly packing of the aliphatic side chains. Meanwhile, In55 aliphatic index significantly less than BLV at 96.23.

The grand average of hydropathicity (GRAVY) values were calculated for the envelope glycoprotein for each virus. This value determines the solubility of the protein. A positive GRAVY value indicates that the protein is hydrophobic, while a negative GRAVY value indicates that the protein is hydrophilic (Kyte & Doolittle, 1982). All values are negative as predicted since the envelope glycoprotein is a transmembrane protein, with In55 being close to GaLV, FeLV and MoMuLV, however more related to MoMuLV with values of -0.238 and -0.234 respectively, compared to -0.257 in FeLV, -0.261 in GaLV and -0.120 in BLV.

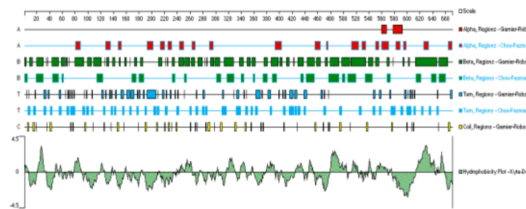
One key difference that distinguished In55 from the other gammaretroviruses is the pI index. The pI index is the isoelectric point of the envelope glycoproteins. In55 has an isoelectric point of 7.50, markedly lower compared to 8.98, 8.78, 8.66 and 8.62 for FeLV, BLV, GaLV and MoMuLV, respectively, due to having an increased number of acidic amino acids compared to the other gammaretroviruses. The isoelectric point of a protein has been shown to influence protein behavior and conformation in different pH values (Nobuo, 1971). The significantly reduced isoelectric point in 7.50 could predict structural differences or functional differences in In55 and the other gammaretroviruses.

Secondary Structure Analysis

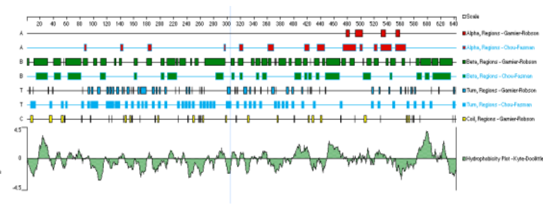
CrERVγ In55



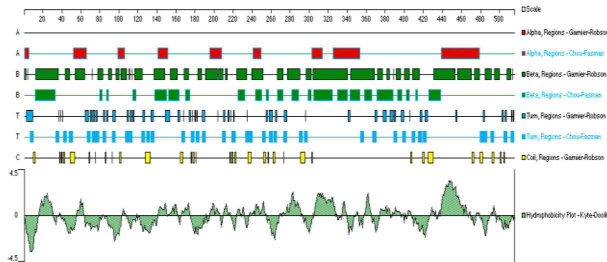
GaLV



FeLV



BLV



MoMuLV

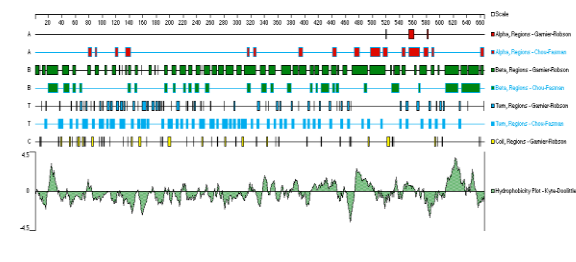


Figure 3: Secondary structure predictions in CrERVγ In55, FeLV, GaLV, MoMuLV, BLV showing predicted locations of alpha helices, beta regions, coils and turns. Also shows predicted areas of hydrophobicity along the amino acid chains.

Secondary structure of the envelope glycoprotein for each of the viruses was determined using DNA Star Protean software (DNA Star 2012) (Figure 2). The probable location of alpha, beta, turn and coil regions were determined. In55 shares significant structural conservation in the secondary structure of the envelope glycoprotein with these gammaretroviruses, especially in the transmembrane region, residues 400-500 and 540-560. This conservation is expected given the importance of the transmembrane region, such as viral fusion and structure (Marchler-Bauer, et al, 2011). In addition to predicting secondary structure, the DNA Star Protean software shows areas of hydrophobicity along the

protein sequence. As predicted, In55 has a hydrophobic (negative) peak in the transmembrane region, which is also present in the transmembrane of all the viruses. Also, there is a strong hydrophilic peak at the N-terminus, around position 40 in In55 which is also conserved in all viruses. In55 does differ from the other viruses at this region in that it is characterized by the presence of an alpha region, whereas all other viruses have a beta region at this region. This region is close to the PHQ motif, which has been found to be critical in receptor binding (Rey, et. al 2007), therefore this conserved, hydrophilic site may also have an impact on receptor binding.

Interestingly, there is a hydrophobic peak at the N-terminus of the In55 envelope glycoprotein that is also present in the other viruses. This hydrophobic area occurs between positions 10-20, the same region in each of the viruses, suggesting it is a conserved feature in gammaretroviruses. There is variation in the secondary structure in this region, however. In55 has alpha helices in this region, which is absent in the other gammaretroviruses, in which this region is characterized by beta regions, coils and turns with minimal or no alpha regions.

Another interesting region is a hydrophilic peak at about position 600 in the C-terminus of In55, characterized as a beta region with turn regions flanking on both sides. This exact structure appears at the same relative position in all the other viruses, although the turn regions are not as closely flanking BLV.

A region that is shared only by In55 and GaLV occurs at position 220 in which both of the viruses have an alpha region at this position A relatively unique feature in In55 is the presence of alpha and turn regions at the very N-terminus of the envelope glycoprotein which is only present in BLV and none of the other viruses. There are differences in the hydrophobicities in these regions between In55 and BLV, however. In55 is only slightly hydrophobic at this point as BLV is largely hydrophobic at this point.

Transmembrane Region Analysis

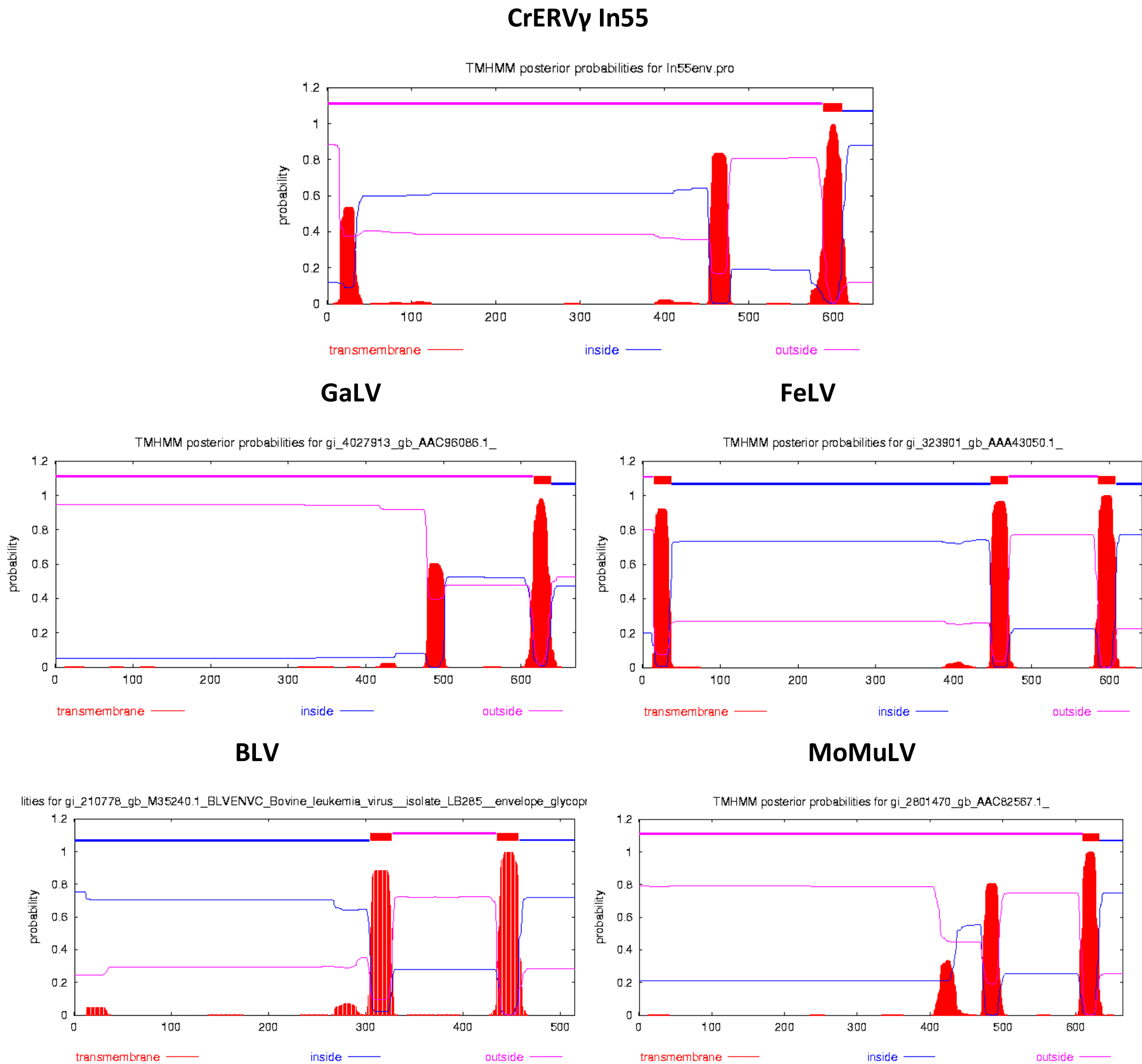


Figure 4: Prediction of transmembrane helices in proteins using TMHMM v 2.0 in the envelope glycoproteins of: CrERVγ In55, FeLV, GaLV, MoMuLV, BLV predicting the location of transmembrane helices along the entire amino acid sequence and the probability that a transmembrane helix appears at that region.

TMHMM software is used to predict the location of transmembrane helices (CBS 2012). The transmembrane helical bundle of In55 occurs at around positions 450 and 600. This is in the same area as the transmembrane helices in GaLV, FeLV, and MoMuLV and relatively closer to the C-terminus than BLV, which occurs at about 65 residues from C-terminus, whereas In55 TM region is situated about 50 residues from the C-terminus. In In55, there's a small peak predicted by THMM at around position 425,

N-terminal to the TM helices, which represents a small hydrophobic area, distinct from the transmembrane region. This region is also present at relatively the same position in all the viruses and is probably the fusion peptide, which is known to be N-terminal to the transmembrane region (Rae, et. al 2007) (Bae, et. al 997).

There is also a peak in the N-terminus of In55 at around residue 5 with a probability of being transmembrane of about 0.5. This also occurs in FeLV with a probability of around 0.9 and also present, to a lesser probability in BLV and minimally present in MoMuLV and GaLV. This contradicts the known structure of the envelope glycoprotein in which it is known that the transmembrane region does not occur this close to the N-terminus (Rey, et. al, 2007), so it is highly unlikely that this site is a transmembrane helix, but this does show that there is a hydrophobic region at this position and it is conserved in all viruses. This is in agreement with the secondary structure analysis which also showed the presence of a hydrophobic region in all of the viruses at the N-terminus.

Motif Analysis

	CrERVγ	GaLV	FeLV-B	MoMuLV	BLV
PHQ motif	44*	42	38	40	-
Proline-rich region (PRR)	290-354	292-356	291-355	292-356	116-180
HTLV-1-like_HR1-HR2	515-575	509-586	479-556	515-580	330-404
Immunosuppressive region	516-533	544-560	514-530	538-554	365-377, 380-383
CX(6)C motif	533-540	561-568	531-538	555-562	384-391

Figure 5: Location of key regions in motif along the entire length of the envelope glycoprotein in CrERVγ, FeLV, GaLV, MoMuLV and BLV. *Indicates that the PHQ motif is partially present in In55.

Common motifs reported in the envelope glycoprotein of gammaretroviruses were studied to determine their presence and location in In55 (Figure 4). One conserved motif is the PHQ motif, crucial for infectivity. It is shown in GaLV and MoMuLV at position 42 and 40, respectively and absent from BLV and in FeLV the important H residue is missing. It is only partially present in In55, having PHH instead of PHQ, which has also been reported in PERV (Lavillete & Kabat, 2004).Also conserved is the HTLV-1-like HR1-HR2 region, which includes the immunosuppressive regions, a Cl binding site, CX (6) motif and H1 and H2. Within this region is the highly conserved immunosuppressive region. This region

is present in In55 and all the other gammaretroviruses, which may be because of its importance in suppressing the immune response. In55 shares the same starting amino acid of this region, L, with all of the viruses except for BLV. The CX(6) motif is also present in the envelope glycoprotein of In55 and is conserved in all of the gammaretroviruses as well. In55 has an identical sequence with FeLV, GaLV and MoMuLV at this region except for a G residue instead of a K residue at position 537 making it distinct from the other three viruses, which have identical motifs.

N-glycosylation Site Analysis

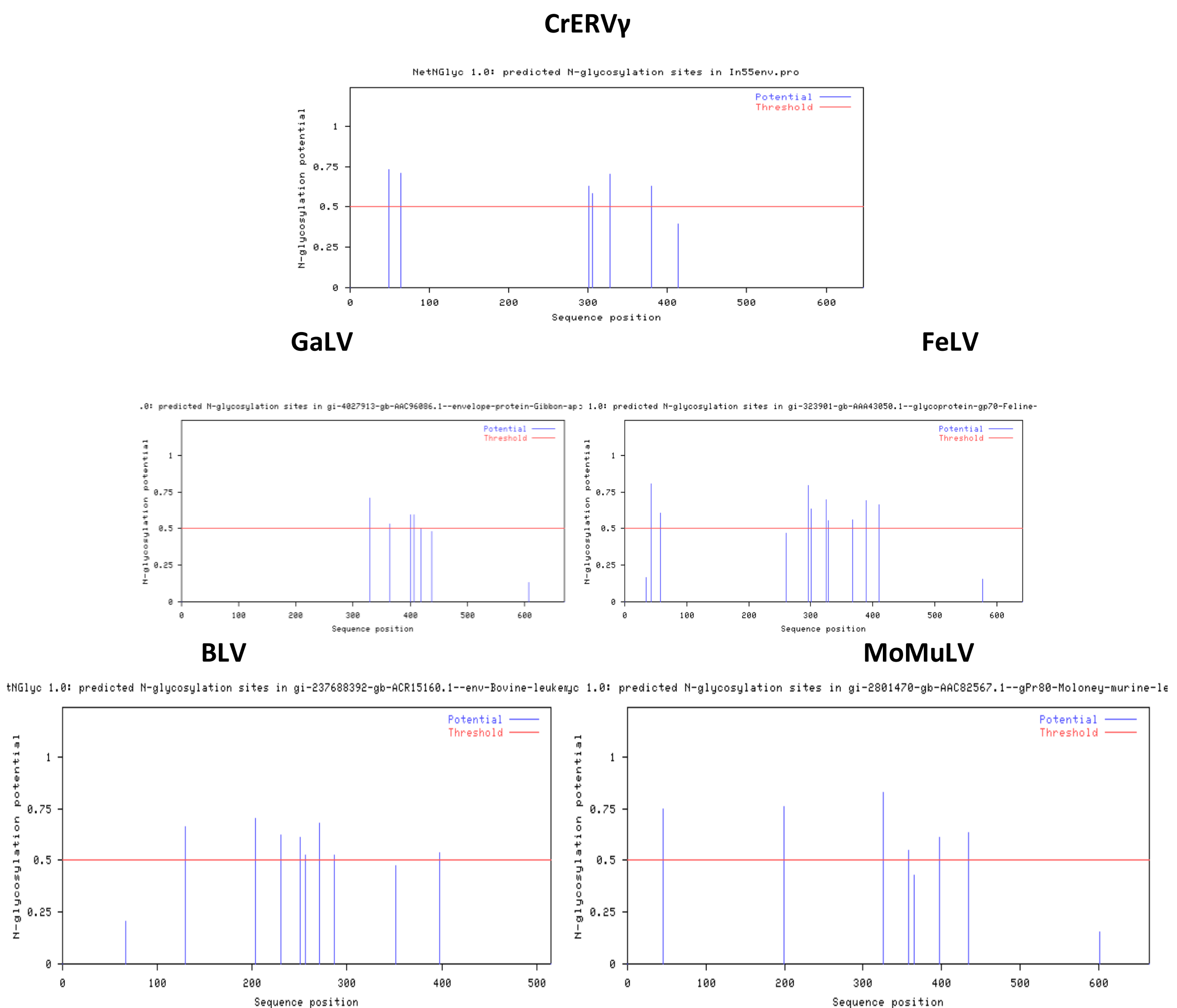


Figure 6: Predicted locations of N-glycosylations based off NetNGlyc software in CrERVγ, FeLV GaLV, MoMuLV, BLV.

The sites of N-glycosylation are visualized graphically (Figure 5) to see which region on which the glycosylation occurs in relation to the entire amino acid chain and subsequently determine the functional roles of these glycosylations. Only peaks above the threshold of 0.5 are significant. There appear to be distinct clusters of N-glycosylation sites along the amino acid chain. In 55 shows clusters of four N-glycosylation sites in the envelope glycoprotein between residues 300-400. There are also four N-glycosylation clusters in this region in GaLV and FeLV, five in MoMuLV and two in BLV. In In55 there are also two N-glycosylation sites at around residue 50. This is only shared by MoMuLV and FeLV, which have one and two N-glycosylations site, respectively. There are no N-glycosylation sites predicted to occur at this site in GaLV and BLV.

Transfection Results

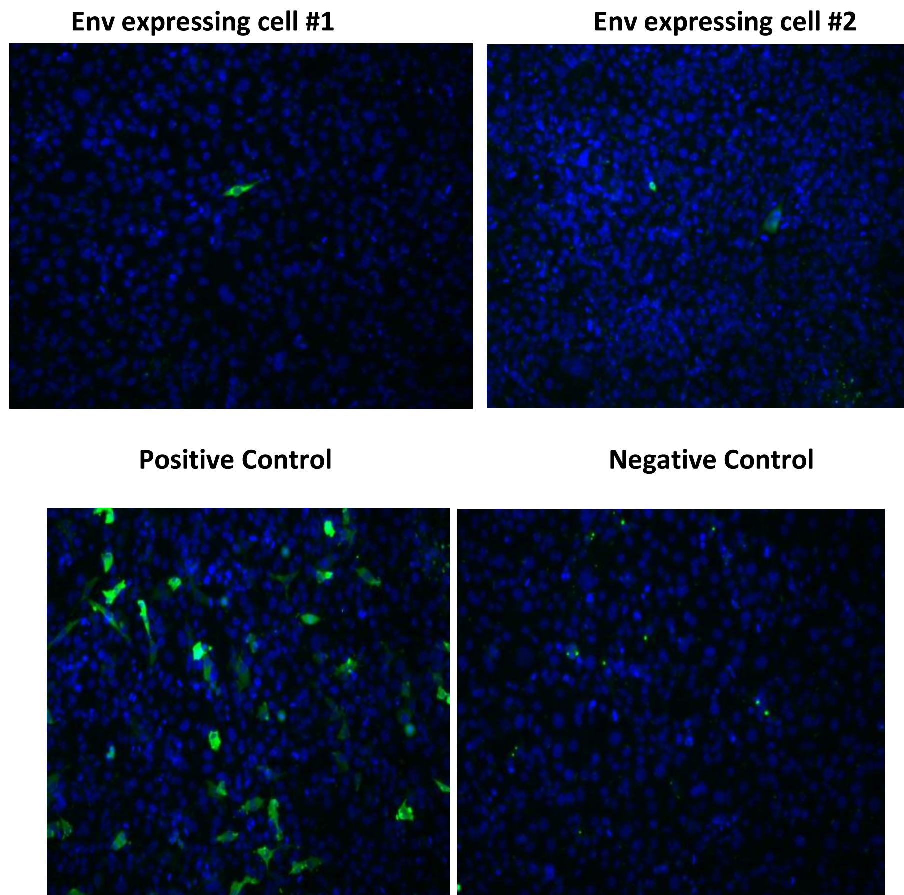


Figure 7: Fluorescent microscopy showing cytoplasmic expression of CrERVγ in chicken cells showing cells expressing env (top), the positive control: 704P (bottom left) and the negative control: empty FLAG vector (bottom right).

The envelope glycoprotein was expressed and visualized in chicken cells in vitro (Figure 8). Cells that are expressing the protein are stained green, due to staining of the FLAG-tag, while the nucleus is stained blue by DAPZ (Figure 8). There is one cell in both the top right and left pictures showing cytoplasmic expression of env. The staining at the bottom right side of the right picture is non-specific. The positive control vector used was 704P. Positive cells are stained green (FLAG-tag) with blue nuclei. The negative control cells are only stained blue because they were transfected with the empty vector.. The cells were successful transfected; however there was only cytoplasmic expression of the envelope protein.

Amino Acid Changes in the Clone:

Position on amino acid chain	**FLAGIn55E clone**	**In55**
95	S	P
301	S	N
354	M	K
476	Q	L

Figure 8: Amino acid changes in the amino acid sequence in the In55 envelope clone compared to the wild type In55 amino acid sequence

The amino acid sequence of the In55E clone was compared to that of the wild type In55 sequence using DNA Star software (DNA Star 2012) (Figure 9). There are significant amino acid substitutions in the clone that occur in the surface subunit (SU) and in the transmembrane region. These include the loss of a proline residue at position 95, important in folding and structure of a protein (Krieger, et. al 2005). At amino position 301, there's an asparagine to serine substitution. This position is an N-glycosylation site, so this substitution causes a loss of N-glycosylation. At position 354, there's a methionine to lysine substitution. At 476, there's a leucine to glutamine substitution, which results in the substitution of a hydrophilic amino acid into the transmembrane region. These substitutions could affect protein folding or intracellular translocation, preventing appropriate expression of the protein on the cell membrane.

DISCUSSION

CrERVγ In55 is a novel endogenous gammaretrovirus that has open reading frames in all major viral proteins, in particular env. This means that the provirus can actively express a functional envelope glycoprotein. To characterize this protein, physical and structural features and phylogenetic relatedness were compared with other known gammaretroviruses: FeLV, GaLV, MoMuLV and BLV.

Phylogenetic analysis (Figure 1) predicts that In55 is more closely related to GaLV and more distant from BLV, which branches off early from all the gammaretroviruses. Comparisons of the structural characteristics, motif commonalities, physical properties, commonalities in motifs and N-glycosylation patterns were used to interpret this phylogenetic analysis and show the relatedness of In55 to the other gammaretroviruses.

The secondary structure of the envelope glycoprotein is determined by the primary amino acid sequence which via H-bonding between the carboxyl and amino groups of the amino acids forms the secondary structures: alpha helices, beta sheets, coils and turns. These structures are crucial for protein function as specific intermolecular interactions dictate protein functionality. Thus, by comparing the secondary structures of proteins, functionality can be predicted. There is significant structural conservation in the secondary structure of the envelope glycoprotein between In55 and the other gammaretroviruses, especially in the transmembrane region (Figure 2). Interestingly, there are also hydrophobic regions at around residue 10 that are conserved (Figure 2). It is also in the leader sequence, so it could be involved in localization of the envelope glycoprotein (Rey, et. al, 2007). Meanwhile, there are predicted structural differences between In55 and GaLV and the other gammaretroviruses, such as the alpha region at position 220. There are also structural characteristics that make In55 unique. One is the presence of an alpha region at the very N-terminus of the protein that is only present in In55 and BLV and absent in the other gammaretroviruses (Figure 2), but BLV has a much higher hydrophobicity at this region. Another distinction is in the conserved N-terminus region at residue 10 (Figure 2). Although there is a hydrophobic region that is conserved in all of the gammaretroviruses at this point, In55 has an alpha region at this point whereas the other viruses have beta regions. Thus, secondary structural analysis demonstrates that In55 shares several regions in common with the other gammaretroviruses, including the transmembrane region and has regions that are unique to itself and its closest relative GaLV, such as the alpha region at position 220. In addition, In55 has unique structural characteristics that make it unique on its own.

There are also highly conserved motifs in all gammaretroviruses, such as the, PHQ motif, HTLV-1-like-HR1-HR2 region, immunosuppressive region and CX (6) C motif signifying their functional importance in the envelope glycoprotein. There were observed homologies in the HTLV-1 like HR1-HR2 region between In55 and the other viruses (Figure 5). This conserved nature of HTLV-1 HR1-HR2 may be indicative of its importance. It is located in the transmembrane region C-terminal to the fusion peptide and is necessary for viral fusion with the host cell membrane (Marchler-Bauer, et al, 2011). Another conserved region is the proline-rich region (Figure 5). This domain is located C-terminal to the receptor binding domain (RBD), between the surface subunit and transmembrane region interacting interface. It is thought that this region does not directly bind to the host cell receptor, but causes a conformational change in the envelope glycoprotein that's necessary for receptor binding via the RBD and subsequent fusion with the plasma membrane (Battini, et. al, 1994) (Lavillette, et. al 1998). These regions are shared by In55 and the other gammaretroviruses showing their structural relatedness and possible functional relatedness. Alternatively, there were some distinctions in In55 in some of these common motifs (Figure 5). For instance, all the viruses, except for BLV, contained the PHQ motif, indicating its importance in fusion of the virus with the host cell membrane, especially the H residue (Lavillette & Kabat, 2004) (Bae, et. al 1997). In55, however only had a partial PHQ motif, making it distinct from FeLV, GaLV and MoMuLV. Another distinction is in the CX(6) motif (Figure 5) in which In55 differs by a single amino acid from the other viruses. These data show that in addition to secondary structure, In55 contains the conserved motifs that are commonly present in the envelope glycoproteins of gammaretroviruses, yet also has slight differences in motifs, such as the C(X)6 motif and PHQ motif which distinguish it from the other viruses, including its closest relative, GaLV.

In 55 has physical features that are similar to GaLV, FeLV and MoMuLV, such as the GRAVY index and the aliphatic index, yet also has characteristics that distinguish itself from the other viruses. One key distinguishing physical feature in In55 is the pI index or the isoelectric point (Figure 2). This difference is due to an increase in the number of acidic amino acids in In55 compared to the other gammaretroviruses. This increase in acidic amino acids could explain the formation of distinct secondary structures as it causes alteration in the interactions between adjacent amino acids.

N-glycosylation occurs post-translationally and is critical for envelop glycoprotein function. After translation, the envelope glycoprotein goes through extensive maturation which involves glycotransferases and glycosidases which add and remove sugar resides, respectively. These sugar

residues are involved in proper protein folding, protein transport to the envelope and also involve in proper folding of the protein and also critical for membrane processing; therefore this site could have effects on proper processing and cleavage of the glycoprotein (Schaerer-Uthurralt, et. al 1994) (Painter & Honen, 1987). Interestingly, the results show that In55 has two major clusters of glycosylation sites (Figure 6). One occurs at the N-terminus at these sites are in the surface subunit (SU) and RBD and may have an impact on host cell receptor recognition and binding (Knoper, et. al 2009) (Figure 6). These N-glycosylation clusters are also present at this region in FeLV and MoMuLV but absent in GaLV and BLV. In addition, clusters of N-glycosylation occur at sites 300-400 in In55which is near the transmembrane region and could impact membrane fusion (Rey, et. al, 2007) (Figure 6). There are also N-glycosylation clusters that occur in all other viruses at this site. This shows again that In55 shares common Asn(X) Thr (Ser) signals in the 300-400 region with the other viruses, yet has unique N-glycosylation at the N-terminus that are only shared by FeLV and MoMuLV.

The similarities reported in the physical and chemical properties, as well as the conserved motifs and secondary structures among the envelope glycoproteins suggest that In55 shares several characteristics common to all gammaretroviruses, yet has distinct secondary structures only in common to GaLV.,. In addition, In55 has physical characteristics and structural features and motifs that make it unique from all the other viruses. These distinctions, especially in the pI index could have significant structural implications and also have effects on receptor interactions and antigenicity (Lavillete & Kabat, 2004).

In55 was successfully cloned and expressed in chicken cells, yet it was only present cytoplasmically and not on the plasma membrane (Figure 7). There are several reasons why there is cytoplasmic expression, but not the intended membrane expression of the protein. One likely explanation is that there were key missense mutations in the clone that caused a change in the primary amino acid sequence. Most of these changes occur in the surface subunit of the envelope glycoprotein and one in the transmembrane region (Figure 8). One amino acid substitution (position 301), occurs at an N-glycosylation site and thus would alter N-glycosylation patterns in the envelope glycoprotein (Figure 9). As mentioned earlier, N-glycosylation is critical for proper folding, processing and location of the protein (Knoper, et. al 2006). Alteration of this maybe could have caused misfolding of the protein, which resulted in it localizing in the cytoplasm instead of the membrane. Another key substitution occurs at position 95, which results in the substitution of a proline residue for a serine residue (Figure 9). Proline residues are often found in turn and loop regions of a protein and are critical for compaction and protein folding of the amino acid chain (Krieger, et. al 2005). Thus, loss of a proline residue at this site

could also cause protein misfolding, which may have prevented the envelope glycoprotein from translocating to the envelope. Another key substitution occurs in the transmembrane region (position 476), in which a hydrophobic leucine reside is substituted with a hydrophilic glutamine residue (Figure 9). The transmembrane region hydrophobic properties are crucial for interaction with the plasma membrane and insertion of the envelope glycoprotein into the membrane. Substitution of a hydrophobic residue with a hydrophilic residue could significantly alter these hydrophobic interactions between residues of the transmembrane region and with the membrane phospholipids, altering the envelope glycoprotein expression on the membrane.

CONCLUSIONS

CrERVγ In55 has open reading frames for all major viral proteins in the genome. In this paper, the envelope glycoprotein was characterized and compared with other gammaretroviruses. There are significant homologies in the amino acid sequences and secondary structures of these viruses, especially in the transmembrane region and regions involved in critical functions like host cell membrane recognition, binding fusion, immunosuppression or proline-rich regions. In addition, there are secondary features that are only shared with GaLV and additional structural and physical features that are unique to itself, distinguishing itself from the other gammaretroviruses. We have successfully expressed the envelope glycoprotein in cytoplasm, but it did not localize to the surface of the cell. Based on comparison with other gammaretroviruses, there's no reason why In55 could not be expressed on this surface. Given this, future work could be done to determine the antigenicity of the envelope glycoprotein exposed on the cell surface to circulating antibodies.

REFERENCES

1. Andrade, M.A.; S. I. O'Donoghue & Rost, B. (2001). Adaptation of protein surfaces to subcellular location. *J Mol Biol,* 276, 517-525.

2. Barklis, E., J. McDermott, S. Wilkens, E. Schabtach, M.F. Schmid, S. Fuller, S. Karanjia, Z. Love, Russell Jones, Y. Rui, X. Zhao and D. Thompson. (1997). Structural analysis of membrane-bound retrovirus capsid proteins. *The EMBO Journal*. 16: 1199–1213.

3. Battini, J-L., O. Danos & Heard, J-M. (1995). Receptor-Binding Domain of Murine Leukemia Virus Envelope Glycoproteins. J. Virol. 69(2): 713–719.

4. Bae, Y., S. M. Kingsman, & Kingsman, A. J. (1997). Functional Dissection of the Moloney Murine Leukemia Virus Envelope Protein gp70. *J. Virol*. 71(3); 2092–2099.

5. Barnett,A.L., J.M. Cunningham. (2001). Receptor binding transforms the surface subunit of the mammalian C-type retrovirus envelope protein from an inhibitor to an activator of fusion. *J. Virol*., 75 (19): 9096–9105

6. Boller, K., O. Jannsen, H. Schuldes, R. R. Tonjes, & Kurth, R. (1997). Characterization of the Antibody Response Specific for the Human Endogenous Retrovirus HTDV/HERV-K. *J. Virol*. 71(6): 4581–4588.

7. Chandhasin, C., P.N. Coan, I. Pandrea, C.K. Grant, P.A. Lobelle-Rich, A. Puetter & Levy,L.S. (2005). Unique long terminal repeat and surface glycoprotein gene sequences of feline leukemia virus as determinants of disease outcome. *J. Virol.* 79(9): 5278-5287.

8. Denner, J. (1998). Immunosuppression by retroviruses: implications for xenotransplantation. *Ann. N.Y. Acad. Sci.* 862, 75-86.

9. DNA Star (2012).Lasergene Core Suite software

10. Dube,S., L. Abbott, D.K. Dube, G. Dolcini, S. Gutierrez, C. Ceriani, M. Juliarena, J. Ferrer, R Perzova & Poiesz,B.J. (2009). The complete genomic sequence of an in vivo low replicating BLV strain. *Virol. J*. 6(120)

11. Fiebig, U., O. Stephan, R. Kurth & Denner, J. (2003). Neutralizing antibodies against conserved domains of p15E of porcine endogenous retroviruses: basis for a vaccine for xenotransplantation? *Virology*. 15; 307(2):406-13.

12. Gemeniano,M., O. Mpanju, D.R. Salomon, M.V. Eiden and Wilson,C.A. (2006). The infectivity and host range of the ecotropic porcine endogenous retrovirus, PERV-C, is modulated by residues in the C-terminal region of its surface envelope protein. *Virology.* 346(1):108-117.

13. Gasteiger E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel & Bairoch A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server;* (In) John M. Walker (ed): *The Proteomics Protocols Handbook*:571-607

14. Ikai, A. (1990). Thermostability and aliphatic index of globular proteins. *J. Biochem.* 88, 1895-1898

15. Kaulitz, D., U. Fiebig, M. Eschricht , C. Wurzbacher, R. Kurth & Denner J. (2010). Generation of neutralising antibodies against porcine endogenous retroviruses (PERVs). *Virology*. 411(1): 78-8.

16. Krieger F, Möglich A, Kiefhaber T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J Am Chem Soc.* 127(10):3346-52.

17. Kelley L.A. & Sternberg, M.J.E. (2009). Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols 4*, 363 – 371.

18. Knoper, R.C., J. Ferrarone,[1] Y.Yan, B.A. P. Lafont, & Kozak, C. A. (2009). Removal of either N-glycan site from the envelope receptor binding domain of Moloney and Friend but not AKV mouse ecotropic gammaretroviruses alters receptor usage. Virology. *Virology*. 391(2): 232–239.

19. Knox, K., D. Carrigan, G. Simmons, F. Teque, Y. Zhou, J. Hackett Jr., X. Qiu, K. Luk, G. Schochetman, Allyn Knox, A. M. Kogelnik & Levy, J. A. (2011). No Evidence of Murine-Like Gammaretroviruses in CFS Patients Previously Identified as XMRV-Infected. *Science. DOI: 10.1126/science.1204963.*

20. Kyte, J. & Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol Biol.* 157(1):105-32.

21. Langhammer S, Hübner J, Kurth R, Denner J. (2006). Antibodies neutralizing feline leukaemia virus (FeLV) in cats immunized with the transmembrane envelope protein p15E. *Immunology*. 117(2): 229-37.

22. Lavillette, D. & Kabat, D. (2004). Porcine Endogenous Retroviruses Infect Cells Lacking Cognate Receptors by an Alternative Pathway: Implications for Retrovirus Evolution and Xenotransplantation. J. Virol. 78(16): 8866-8877.

23. Lavillette, D., M. Maurice, C.Roche, S. J. Russell, M. Sitbon & Cosset, F-L. (1998). A Proline-Rich Motif Downstream of the Receptor Binding Domain Modulates Conformation and Fusogenicity of Murine Retroviral Envelopes. *J. Virol. 72(12): 9955-9965.*

24. Nobuo, U.l. (1971). Isoelectric points and conformation of proteins: I. Effect of urea on the behavior of some proteins in isoelectric focusing. *Biochimica et Biophysica Acta (BBA)* - Protein Structure. 229(3): 567–581.

25. Nakaya, Y., S. Hoshino, J.Yasuda & Miyazawa, T. (2011). Mapping of a neutralizing epitope in the surface envelope protein of porcine endogenous retrovirus subgroup B. J. *Gen. Virol*. 92: 940-944.

26. Nunberg, J.H., M.E. Williams & Innis, M.A. (1984). Nucleotide sequences of the envelope genes of two isolates of feline leukemia virus subgroup B. *J. Virol*. 49(2): 629-632.

27. Overbaugh,J., E.A. Hoover, J.I. Mullins, D.P. Burns, L. Rudensey, S.L. Quackenbush, V. Stallard & Donahue,P.R. (1992).Structure and pathogenicity of individual variants within an immunodeficiency disease-inducing isolate of FeLV. *Virology* 188(2): 558-569.

28. Pincus, SH., R Cole, R Ireland, F Mcatee, R Fujisawa & Portis, J. (1995). Protective Efficacy of Nonneutralizing Monoclonal Antibodies in Acute Infection with Murine Leukemia Virus. *J. Virol.* 69(11): 7152–7158

29. Pinter, A. & Honnen, W. J. (1987). O-linked glycosylation of retroviral envelope gene products. *J. Virol*. 63(3): 1016-1021.

30. Petropoulos,C.J.& Chappey, C. (1997). Appendix 2: Retroviral taxonomy, protein structure, sequences, and genetic maps. *RETROVIRUSES:* 757;

31. Qiagen Samples & Assays Technologies. (2012). Mini-Prep, Gel Isolation and PCR Purification Kits.

32. Qiu, X, P. Swanson, K. Luk, B. Tu, F. Villinger, J. D. Gupta, R. H. Silverman, E. A. Klein, S. Devare, G. Schochetman & JHackett, J. (2010). Characterization of antibodies elicited by XMRV infection and development of immunoassays useful for epidemiologic studies. *Retrovirology* 2010, 7:68.

33. Rosenberg N, Jolicoeur P (1997). Retroviral Pathogenesis. *Retroviruses*. 475-585.

34. Schlessinger, A. &Rost,B. (2005). Protein flexibility and rigidity predicted from sequence. *Proteins.* 61:115-126.

35. Sigma Aldrich. (2012). p3XFLAG-CMV™-10 Expression Vector E7658. <http://www.sigmaaldrich.com/catalog/ProductDetail.do?lang=en&N4=E7658|SIGMA&N5=SEARCH_CONCAT_PNO|BRAND_KEY&F=SPEC&cm_sp=Customer_Favorites-_-Detail_Page-_-Image-E7658>.

36. Stephens, C. R. & Waelbroeck, H. (1998). Codon Bias and Mutability in HIV Sequences. *J Mol Evol*. 48:390–397.

37. Stephenson, J.R, M. Essex, S. Hino, W. D. Hardy, Jr., & Aaronson, S.T. (1977). Feline oncornavirus-associated cell-membrane antigen (FOCMA): Distinction between FOCMA and the major virion glycoprotein. *Proc. Natl. Acad. Sci. USA* 74(3): 1219-1223.

38. Stothard, P (2000). The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques.* 28:1102-1104.

39. Tailor, S. C. & Kabat, D. (1997). Variable regions A and B the envelope glycoprotein of feline leukemia virus interact with discrete receptor domains subgroup B and amphotropic murine leukemia glycoproteins of feline leukemia virus. *J. Virol.* 71(12):9383.

40. Ting,Y.T., C.A. Wilson, K.B. Farrell, G.J. Chaudry & Eiden,M.V. (1998). Simian sarcoma-associated virus fails to infect Chinese hamster cells despite the presence of functional gibbon ape leukemia virus receptors. *J. Virol*. 72(12): 9453-9458.

41. Weiss, R. A. (2006). The discovery of endogenous retroviruses. *Retrovirology*. 3: 67.

ACADEMIC VITA of Jason G. Bankert

Jason G. Bankert
909 E. Atlantic Ave
Altoona, PA 16602
jgb5062@psu.edu

Education: Bachelor of Science Degree in Immunology & Infectious Disease, Penn State University, Spring 2012
Bachelor of Science Degree in Toxicology, Penn State University, Spring 2012
Minor in Global Health
Thesis Title: Characterization of a Novel Endogenous Retrovirus Envelope Protein in North American Mule Deer
Thesis Supervisor: Mary Poss, PhD, DVM

Related Experience:
Volunteer work at the James E. VanZandt VA Medical Center in infection control
Job shadowing in infectious disease
Global Medical Brigades trip to Honduras
Global Health Minor Fieldwork Experience in South Africa

Awards/ Honors:
Schreyer Honors Scholar
Alpha Lambda Delta Honor Society for Freshmen
Penn State Altoona Honors Program
Dean's List

Presentations/Activities:
Co-organizer of the Run to the Rescue fundraiser for the g90 project
Penn State Ag Advocates
Penn State Altoona Lion Ambassadors
Penn State Altoona Student Government Association