Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations | Graduate School

1994

Expression, Purification and Characterization of Hepatitis C Virus Core Protein from E.Coli Using a Chemically Synthesized Gene (1), and Cloning, Expression, Purification and Characterization of the Major Core Protein (P26) from Equine Infectious Anaemia Virus

Ashley James Birkett

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

 Part of the Biochemistry Commons

© The Author

Downloaded from
https://scholarscompass.vcu.edu/etd/4367

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

www.manaraa.com

Virginia Commonwealth University
School of Medicine

This is to certify that the dissertation prepared by Ashley James Birkett entitled "Expression, Purification and Characterization of Hepatitis C Virus Core Protein from *E.coli* using a Chemically Synthesized Gene (1) and Cloning, Expression, Purification and Characterization of the Major Core Protein (p26) from Equine Infectious Anaemia Virus" has been approved by his committee as satisfactory completion of the dissertation requirement for the degree of Doctor of Philosophy.

[redacted]

Darrell L. Peterson, Ph.D., Director of Dissertation

[redacted]

LaVerne G. Schirch, Ph.D., School of Medicine

[redacted]

H. Tonie Wright, Ph.D., School of Medicine

[redacted]

Mark S. Rosenkrantz, Ph.D., School of Medicine

[redacted]

Anthony D. Carter, Ph.D., School of Medicine

[redacted]

Robert K. Yu, Ph.D., Department Chairman

[redacted]

Hermes A. Kontos, M.D., Ph.D., Dean, School of Medicine

[redacted]

William L. Dewey, Ph.D., Dean, School of Graduate Studies

August 18, 1994
Date

EXPRESSION, PURIFICATION AND CHARACTERIZATION OF HEPATITIS C VIRUS CORE PROTEIN FROM *E.COLI* USING A CHEMICALLY SYNTHESIZED GENE (1), AND CLONING, EXPRESSION, PURIFICATION AND CHARACTERIZATION OF THE MAJOR CORE PROTEIN (P26) FROM EQUINE INFECTIOUS ANAEMIA VIRUS

A dissertation submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at Virginia Commonwealth University.

by

Ashley James Birkett
B.Sc. Applied Biological Sciences,
Bristol Polytechnic, Bristol, England,
July, 1990.

Director: Dr. Darrell L. Peterson,
Professor of Biochemistry,
Department of Biochemistry
and Molecular Biophysics.

Virginia Commonwealth University
Richmond, Virginia
August, 1994.

© Ashley James Birkett, 1994

All Rights Reserved

DEDICATION

To my family, and to Cindy, for their love, patience, and continual encouragement over the years.

ACKNOWLEDGEMENTS

First and foremost I would like to thank Dr. Darrell Peterson for giving me the opportunity to work in his laboratory. His tremendous enthusiasm toward basic research, and his continual willingness to teach and discuss ideas, made my graduate education a most rewarding and enjoyable experience.

I would like to thank the other members of Dr. Peterson's laboratory, Ling Jin, Xin Wei, and Julian Gomez, not only for their technical help, but for providing such a pleasant working environment. Further, I wish to thank Dr. Francisco Gavilanes (University of Madrid) and the members of his laboratory for their collaboration on the biophysical studies, and Dr. Tonie Wright for his helpful discussion of the crystallization portion of this project.

I would also like to thank Dr. Verne Schirch for providing me with the initial opportunity to visit the United States as an undergraduate 'exchange student' in 1988-89. My thanks also goes out to Dr. Jan Chlebowski for initiating and maintaining the student-exchange program with Bristol Polytechnic, and also for his continual willingness to provide assistance.

Elke and Siggy, your generosity, love, and friendship will always be appreciated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ALT	alanine aminotransferase
bp	base pair(s)
dNTP	deoxynucleotide triphosphate
EDTA	ethylenediaminetetra acetic acid
ER	endoplasmic reticulum
ELISA	enzyme linked immunosorbent assay
Gp	glycoprotein
GST	glutathione-S-transferase
HAV	hepatitis A virus
HBV	hepatitis B virus
HCV	hepatitis C virus
HCC	hepatocellular carcinoma
Ig	immunoglobulin
IPTG	isopropyl-β-thiogalactoside
LB	Lura-Bertani broth
Mg	milligram
MoAb	monoclonal antibody
Nm	nanometres
NS	non-structural
Nt	nucleotide(s)
NTA	nitriloamine

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PT-NANBH	post-transfusion non-A, non-B hepatitis
RIBA	recombinant immunoblot assay
RT	reverse transcription
SDS	sodium dodecyl sulfate
UTR	untranslated region

ABSTRACT

EXPRESSION, PURIFICATION AND CHARACTERIZATION OF HEPATITIS C VIRUS CORE PROTEIN FROM *E.COLI* USING A CHEMICALLY SYNTHESIZED GENE (1); AND CLONING, EXPRESSION, PURIFICATION AND CHARACTERIZATION OF THE MAJOR CORE PROTEIN (P26) FROM EQUINE INFECTIOUS ANAEMIA VIRUS (2)

By Ashley James Birkett, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 1994.

Major Director: Darrell L. Peterson, Ph.D.
Professor of Biochemistry.
Department of Biochemistry and
Molecular Biophysics.

The hepatitis C virus (HCV) core gene has been chemically synthesized and used to direct the expression of core protein in *Escherichia coli* (*E.coli*). When cloned downstream of the inducible T7 promotor the core gene directed the expression of a soluble protein with a molecular weight of 22 kDa. Under native conditions the protein behaved in a manner consistent with the formation of a multimeric structure, which may represent assembled core particles. Further, when examined by electron microscopy a heterogeneous mixture of nucleocapsid-like particles were visible. Core protein was specifically

recognized by antibodies present in HCV infected serum, suggesting that this protein may be useful as a diagnostic tool for detecting HCV infection. To circumvent the difficulty in purifying this protein in significant quantities a core-polyhistidine fusion gene was constructed. This protein was readily purified, using nickel chelation chromatography, under denaturing conditions. Studies are currently on-going to ascertain the conditions required for refolding this protein, and the suitability of this protein to serve as a diagnostic tool.

In a second study, the major core protein (p26) of the lentivirus equine infectious anaemia (EIAV) was expressed in *E.coli* and purified to >95% homogeneity. Circular dichroism spectroscopy revealed that p26 exhibits the following assignment of secondary structural elements; 40% alpha helix, 22% beta sheet, 10% beta turn, and 28% random coil. It has been determined that p26 contains a single free cysteine residue (Cys48), and an intramolecular disulfide bond between cysteine residues 198 and 213. Data acquired by circular dichroism spectroscopy and fluorescence spectroscopy indicate that this disulfide bond plays a critical role in maintaining the structure of p26. Crystals of p26 were successfully grown at a protein concentration of 8mg/ml in 0.1M cacodylate, 0.7M sodium acetate, pH 6.5, at 4 °C. A sandwich ELISA was developed, using recombinant p26, to detect anti-p26

antibodies in horse sera. The assay successfully identified the 19 positive sera in a blind panel of 30 samples supplied by the National Veterinary Service Labs (Ames, Iowa). Accordingly, the assay appears to meet the criteria required of a commercial diagnostic reagent for determining EIAV infection in horses.

I. EXPRESSION, PURIFICATION AND CHARACTERIZATION OF HEPATITIS C VIRUS CORE PROTEIN FROM *E.COLI* USING A CHEMICALLY SYNTHESIZED GENE

INTRODUCTION

Hepatitis C Virus

Hepatitis C virus (HCV) was identified in 1989 by researchers at Chiron as being the major cause of post-transfusion non-A, non-B hepatitis (PT-NANBH) (Choo *et al.*, 1989). This ended a 15 years search for an ellusive virus that, in the late 1970's and early 1980's, reportedly infected up to 10% of the nation's blood supply (Aach *et al.*, 1981). A 1990 report suggested that 60 to >90% of PT-NANBH observed in the U.S., Europe and Japan was caused by hepatitis C (Choo *et al.*, 1990), which amounts to as many as 0.3 million cases per annum in the U.S. alone (Davis *et al.*, 1990). The recent development of reliable diagnostic assays has effectively eradicated hepatitis C from the nation's blood supply; however, the virus continues to be transmitted by other routes such as intravenous drug use and sexual contact.

Acute HCV infections are typically asymptomatic and generally less severe than those caused by either hepatitis A (HAV) or hepatitis B virus (HBV). However, it is the fact that approximately three-quarters of infections progress to a chronic state which has caused such a health concern. Chronic

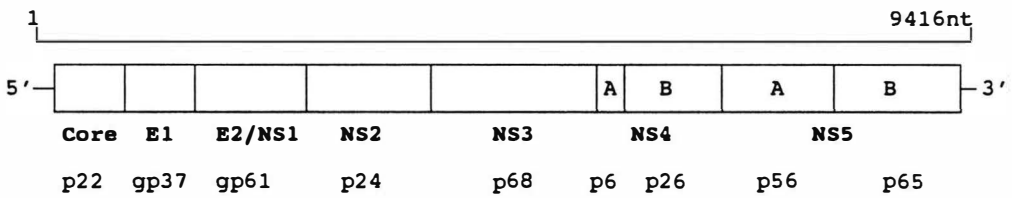
infections of HCV appear to persist for over 30 years, and are associated with the development of hepatocellular carcinoma (Colombo *et al.*, 1989; Saito *et al.*, 1990; Nishioka *et al.*, 1991; Ohkoshi *et al.*, 1990).

Genomic Organization

Hepatitis C virus is an enveloped virus with a diameter of 30-60nm and has a positive strand RNA genome. Several reports have highlighted the similarity, in terms of both sequence and genomic organization, of this virus to two families of viruses - the pestiviruses and the flaviviruses (Miller and Purcell, 1990). The latter are a family of small animal viruses which are categorized according to their homology to the yellow fever virus (flavus meaning 'yellow' in latin), the prototype family member which was first described almost one century ago. This group of viruses now contains over 50 animal viruses, all of which have a single-stranded, positive RNA genome (for a review, see Chambers *et al.*, 1990). The genome, which is typically 10-11kB in length, is transcribed directly into a single, genomic length mRNA. The mRNA is subsequently translated into a single polyprotein of approximately 3000 amino acids, which is processed by proteases to yield functional viral proteins.

The hepatitis C viral genome has a coding region which is flanked by both 5' and 3' untranslated regions (UTR). The 5' UTR is 341bp long and highly conserved, suggesting that it

Figure 1: Schematic Representation of the HCV Genome.


1
9416nt
5′
A
B
A
B
3′
Core
E1
E2/NS1
NS2
NS3
NS4
NS5
p22
gp37
gp61
p24
p68
p6
p26
p56
p65

may perform an as yet undefined regulatory function. The shorter 3' UTR (23-54bp) is less conserved and it's putative role in the lifecycle of the virus has not yet been identified. The coding sequence of the viral genome is translated into either a 3010 or 3011 amino acid polyprotein, which is processed by both cellular and viral proteases to generate mature or functional viral proteins (Figure 1).

The hepatitis C viral proteins have been categorized into two groups; the structural proteins (core, C; and envelope, E1 and E2/NS1), and the non-structural proteins (NS2-5). The structural proteins are derived from the amino terminus of the viral polyprotein and form the basic framework of the virus, whereas the non-structural proteins are derived from the carboxyl-terminus and exhibit various enzymatic activities which are involved in virus maturation and replication.

Non-Structural Proteins

The most intensely studied of the non-structural proteins to date has been NS3, which has recently been shown to possess serine protease activity (Tomei *et al.*, 1993; Bartenschlager *et al.*, 1993). NS3 specifically cleaves the viral precursor at dibasic amino acid sequences. This viral protease contains a 'catalytic triad' of amino acids (histidine, serine and aspartic acid), a motif which is characteristic of serine proteases. The essential function of these three amino acids is highlighted by the fact that they are conserved in all

viral isolates sequenced to date, and mutation of the catalytic serine residue (to alanine), has been shown to abolish enzymatic activity (Tomei *et al.*, 1993; Bartenschlager *et al.*, 1993). Since the function of this enzyme is to cleave the non-structural proteins of the viral polyprotein, an essential feature of the viral lifecycle, NS3 has been identified as a potential target of anti-viral therapy. Further, this 69kDa protein also possesses a nucleoside triphosphate binding domain which is speculated to be involved in a helicase activity (Takamizawa *et al.*, 1991).

Non-structural-5 protein is a 116kDa protein which is postulated to be involved in viral replication. It contains a 'GDD' consensus sequence which is conserved among RNA dependent, RNA polymerases (Takamizawa *et al.*, 1991). Recently it has been reported that NS5 codes for two distinct proteins, termed A and B, with molecular weights of 56 and 65 kDa (Tomei *et al.*, 1993). The individual functions of these two proteins is still unclear at this time.

NS2 and NS4 are very hydrophobic proteins with molecular weights of 24 and 53kDa respectively. Their hydrophobicity has led to the prediction that they serve a function which requires them to be 'membrane bound', although this has not been substantiated by experimental data. Like NS5, NS4 has recently been shown to be processed to yield two proteins of 6 and 26 kDa (Tomei *et al.*, 1993).

Structural Proteins

There are now recognized to be 3 different structural proteins associated with the hepatitis C virus; core (C), and two envelope proteins (E1, and E2/NS1). Both of the envelope proteins are glycosylated.

While the non-structural proteins of HCV are processed by NS3, the virally encoded serine protease, the structural proteins appear to be processed by proteases which are endogenous to the host. The cleavage of core and E1-E2/NS1 was postulated to be mediated by a signal peptidase located in the lumen of the endoplasmic reticulum (ER) (Hijikata *et al.*, 1991), and this has recently been supported by experimental data (Santolini *et al.*, 1994). Accordingly, the signal sequence for E1 is presumed to be located at the carboxyl-terminus of core. Processed core protein is predicted to be 191 amino acids in length (21-22kDa), and a species of approximately this size has been identified by immunoblotting following expression of the viral genome in a variety of eukaryotic expression systems (Chiba *et al.*, 1991; Grakoui *et al.*, 1993; Harada *et al.*, 1991; Hijikada *et al.*, 1991; Kumar *et al.*, 1992).

Diagnosing HCV Infection

In just a few years rapid advancements have been made in curtailing the spread of this virus. This is largely as a result of the determination of the sequence of the viral

genome, which enabled the development of reliable assays for screening donated blood products. The effectiveness of these diagnostic assays has steadily increased since the introduction of the original c-100-3 assay. This assay was introduced in 1989 and screened for antibodies to a non-structural region of the virus (Kuo *et al.*, 1989). Current diagnostic assays detect antibodies to several structural and non-structural proteins and are extremely effective in identifying infected individuals.

The first diagnostic assay for detecting HCV infection arose from the first HCV DNA clone (5-1-1), which was extended to cover 363 viral amino acids and fused to superoxide dismutase (SOD) (Kuo *et al.*, 1989). The introduction of this assay was revolutionary in it's ability to screen sera for the major causative agent of PT-NANBH when it became commercially available in 1990. However, it soon became apparent that antibodies to this region of the viral polyprotein were not universal in HCV infection. Most patients only developed anti-c-100-3 antibodies after being infected for several months, and a significant proportion of patients, around 25%, failed to develop a humoral response to this antigen (Chiba *et al.*, 1991; Harada *et al.*, 1991). False positives were also a problem with this assay, particularly among individuals with alcoholic liver disease and autoimmune chronic active hepatitis (McFarlane *et al.*, 1990).

Once the entire HCV genome had been sequenced and

tentatively characterized it was determined that the viral sequence contained within the c-100-3 clone was derived from NS4, a relatively non-immunogenic protein. In recent years the inclusion of viral epitopes derived from the core and NS3 regions of the viral polyprotein has markedly improved the sensitivity and specificity of HCV diagnostic tests. One of the most widely used diagnostic assay at present is an immunoblot assay known as RIBA-2, a test which incorporates four recombinant proteins; 5-1-1, c-100-3, C33c and c22. C33c is an NS3 fusion protein, and c22 is a core fusion protein. Patients with antibodies to just one of the four viral antigens are considered 'indeterminate' for hepatitis C infection, while those reacting to two or more are considered positive. Since the viral epitopes are fused to superoxide dismutase (SOD), an SOD control is included in the test to confirm that the reactivities detected are indeed specific for the viral sequences.

The inclusion of NS3 and core epitopes in the second generation HCV diagnostic assays has greatly increased their sensitivity to the acute phase of hepatitis C infection (Puoit *et al.*, 1992; Van der Poel *et al.*, 1991). In addition to appearing earlier in infection than antibodies to c-100-3, antibodies to NS3 and core are also detected in patients suffering from NANBH who had previously tested negative for HCV infection.

Various diagnostic ELISA's are also available for

diagnosing HCV infection (Chemello *et al.*, 1993). They incorporate the same viral antigens as the recombinant immunoblot assays, and, while they are easier to perform, they are unable to characterize antibody responses to individual proteins.

Confirmation of the presence of viral nucleic acid in sera has traditionally been determined by hybridization using northern blots. However, the hepatitis C virus circulates at such low levels in the serum that this technique is not sensitive enough when considering infection with this virus (Alter and Hoofnagle, 1984). A combined reverse transcription and polymerase chain reaction (RT-PCR) has been employed successfully for detecting infection by the hepatitis C virus (Cristiano *et al.*, 1991; Schlauder *et al.*, 1992). Since 10% of patients with NANBH are negative for all HCV antibodies currently detected by RIBA-2 (Chemello *et al.*, 1993), PCR provides an alternative method for detecting infection. Viral RNA is usually detectable 1-3 weeks after the patient presents symptoms of infection (Puoti *et al.*, 1992). Further, since PCR detects the presence of viral RNA instead of antibodies to viral proteins, it is an excellent method for detecting an active HCV infection. Accordingly, PCR is typically used to evaluate the success of interferon therapy, in addition to differentiating between an active and a past-resolved infection. In a small group of patients where no antibody responses are detectable, RT-PCR is the only marker of

infection.

In the RT-PCR test oligonucleotide primers are selected to coincide with highly conserved regions of the viral genome, typically the 5′ UTR (Imberti *et al.*, 1991). The inherent sensitivity of this procedure, which is one of its advantages, is also one of it's shortfalls since cross-contamination of samples is a major concern. Nevertheless, despite fears of contamination, as well as its high expense, RT-PCR has proved to be an important tool for detecting HCV infection.

Vaccines and Therapy For HCV Infection

Although there is not a commercially available vaccine for hepatitis C at the present time, trials using recombinant envelope proteins (gp33 and gp72), conducted in chimpanzees, are underway. In the only trial published to date, five of seven chimpanzees 'immunized' with E1 and E2 exhibited protection when challenged with the same viral subtype (Choo *et al.*, 1994).

Therapy for hepatitis C virus infection has largely focussed on alpha-interferon, which has been somewhat successful in treating chronically infected patients (Davis *et al.*, 1989; DiBisceglie *et al.*, 1989). During interferon therapy patients exhibit a decrease in the levels of liver enzymes circulating in the serum, signifying a decrease in the amount of liver damage being inflicted. However, around 70% of patients relapse shortly after the cessation of

administration of the drug.

While studies are ongoing in evaluating the usefulness of interferon therapy, it is becoming evident that other factors, in addition to the typical parameters such as patient profile and treatment schedule, play a role in the effectiveness of this treatment. An interesting finding, particularly from a biochemical viewpoint, is that the genotype of the virus may play an important role in affecting the outcome of interferon therapy (Tsubota *et al.*, 1993). This raises an intriguing question concerning the relationship between clearance of the virus and the immune responsiveness to different B and T-cell sites, since it would appear that certain genotypes of the hepatitis C virus are more resistant to an immune system which has been 'boosted' by interferon treatment than others.

Viral Subtypes

Four different HCV genotypes (I, II, III and IV) have been categorized to date. Genotypes I and II are classified as subtype 1, and genotypes III and IV are classified as subtype 2 (Machida *et al.*, 1992). The majority of amino acid variation between these genotypes is clustered around two hypervariable regions (HVR's) located in the envelope sequences. HVR-1 covers amino acids 1-25 of E2/NS1, and HVR-2 spans amino acids 91-99 of E2/NS1 (Okada *et al.*, 1992).

A clinical study published in 1992 presented evidence

that patients are more likely to respond to interferon therapy if they are infected with a single viral isolate than if they are infected with multiple viral genotypes (Okada *et al.*, 1992). This observation reaffirms the fact that there is indeed a differential viral sensitivity to interferon treatment. It has also been shown that the 27 amino acid stretch that encompasses HVR-1 mutates at a rate of approximately 1 amino acid per month, and that the newest mutant soon becomes the predominant genotype present in the host (Kato *et al.*, 1992). This is presumably due to the capability of the 'new mutant' to evade the immune system of the host, at least temporarily.

Properties of Hepatitis C Virus Core Protein

Structure and Function

HCV core protein has been successfully expressed in several eukaryotic systems, including baculovirus (Hsu *et al.*, 1993), and monkey cos cells (Harada *et al.*, 1991); as well as *in vitro*, using a rabbit reticulocyte lysate (Hijikata *et al.*, 1991). The benefit of these systems is that they are capable of processing the structural polyprotein in an analogous manner to that seen during infection. These studies permitted the mapping of the amino terminus of core protein to the 5' end of the viral genome, directly after the 5' UTR. Further, the carboxyl-terminus of core was mapped by sequencing the amino terminus of E1, which is cleaved from

core during polyprotein maturation in the endoplasmic reticulum. Accordingly, core protein has been postulated to be 191 amino acids in length (Takeuchi *et al.*, 1990), assuming that cleavage from E1 involves a single proteolytic cleavage event. However, it is conceivable that core and E could be cleaved by the excision of a short 5-10 amino acid peptide, which would still generate core protein exhibiting an apparent molecular weight of approximately 21kDa.

Consistent with computer aided predictions, core protein does not appear to undergo any post-translational modifications such as glycosylation (Harada *et al.*, 1991; Hsu *et al.*, 1993). This observation has prompted the suggestion that core protein could be expressed effectively in a prokaryotic system such as *E.coli,* and may have antigenic properties similar to those of the protein expressed in baculovirus (Chiba *et al.*, 1991). Indeed, a single report has described the successful expression of core protein in *E.coli* (Muraiso *et al.*, 1990). This protein was shown to react specifically with antibodies from HCV infected individuals by immunoblotting.

Despite the successful expression of core protein, using several different expression systems, the protein has not been purified in sufficient quantities to permit it's structure and functional characteristics to be examined.

Immunogenicity

Unlike the envelope sequences, the core sequences of hepatitis C virus are remarkably well conserved. While the envelope sequence undergoes a rapid mutation, particularly at the HVR's, the core protein does not. It is widely accepted that there are only two different core sequences; the first, found in subtype 1, and the second associated with subtype 2. The major site of variability has been mapped to a region near the middle of the 22kDa protein, where the two subtypes differ by 7 amino acids over a stretch that spans just 11 in total (Figure 2).

B-cell epitopes within the core protein sequence have been tentatively mapped using both synthetic peptides and fusion proteins. Using GST-core fusion proteins, the major B-cell epitope(s) was mapped to the region encompassing amino acids 21-41 of core (Nasoff *et al.*, 1991). Similarly, residues 39-74 have been implicated in harboring a B-cell epitope using synthetic peptides (Okamoto *et al.*, 1990). Recently Feinstone and co-workers have more precisely mapped the amino-terminal epitope to amino acids 33-50 (Akatsuka *et al.*, 1993). In combination, these data suggest that the amino-terminal B-cell epitope, if it is indeed a single epitope, is located within a 9 amino acid sequence between amino acids 33-41. The same group has successfully identified two further B-cell sites on core. These epitopes, which have only been observed using synthetic peptides (Nasoff *et al.*, 1991), have

been mapped to amino acids 165-174 (Akatsuka *et al.*, 1993).

It has recently been reported that core possesses two HLA class-1 restricted epitopes (Kozeil *et al.*, 1993). These sequences are situated at the amino-terminal region of the protein, and their presence indicates that core protein is a target for HCV-specific cytotoxic T lymphocytes. These cells are an integral part of the immune system, and are recognized as having particular importance in destroying virally infected cells.

Core Protein in Diagnosing HCV Infection

The successful cloning and expression of core protein, in a variety of different expression systems, prompted the study of humoral responses to this protein using immunoblots. A study published in 1991 used core protein expressed in baculovirus to develop an ELISA to measure anti-core antibodies in two hepatitis C patients (Chiba *et al.*, 1991). The first patient showed a typical clinical course of chronic NANBH exhibiting multiple peaks of ALT, a marker of liver cell damage caused by hepatitis (Figure 3A). It was determined that anti-core antibodies were detectable 7 weeks post-infection in this particular patient, which was before the first peak of ALT. Alternatively, anti-c-100-3 antibodies were not detectable until 57 weeks post-infection, almost a full year after the anti-core activity was detectable.

IPKARRPEGRTWAQPGY ... Subtype 1 (Genotypes I and II)

---D--ST-KS-GK--- ... Subtype 2 (Genotypes III and IV)

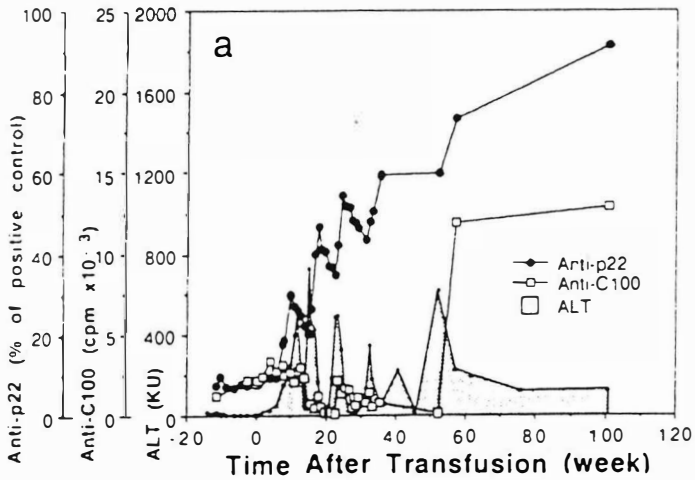
Figure 2: Amino Acid Subtype Differences in Core.

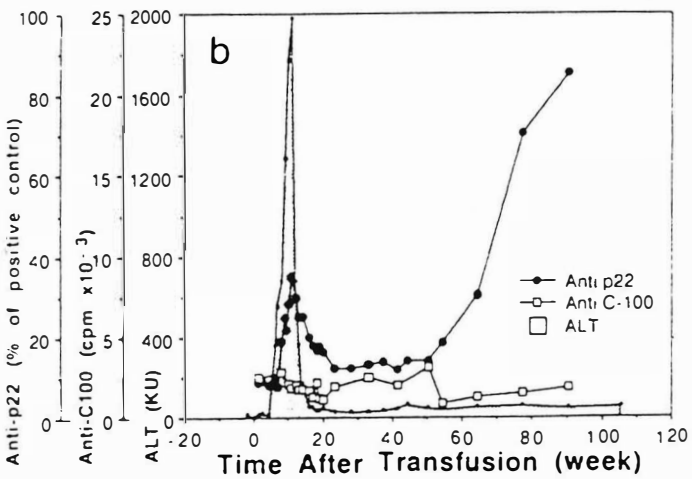
In the second patient, who also exhibited a typical, yet different, clinical course of NANBH, the results were similar -anti-core antibodies were detected at about 10 weeks post-infection (Figure 3B). This was just after the elevation in the serum ALT levels, but before the single ALT peak exhibited by this patient. Despite maintaining low, but nevertheless abnormal levels of ALT for 4 years, the patient remained anti-c-100-3 negative (Chiba *et al.*, 1991). Examination of 23 other NANBH patients showed that while 13 patients seroconverted to anti-core positive within 1-3 months, only 4 patients seroconverted to anti-c-100-3 positive during the same period.

These results highlight the usefulness of core protein as a diagnostic tool for detecting HCV infection. Predictably, the inclusion of core epitopes in the commercially available diagnostic assays has dramatically increased their effectiveness, not only for detecting infection at an earlier stage, but also detecting infection in patients who did not develop anti-c-100-3 antibodies. Further, it has been suggested that detecting infection at an earlier stage is beneficial for interferon therapy. Those patients receiving interferon treatment in the acute stage of infection are more likely to clear the virus before it progresses to a chronic infection (Davis *et al.*, 1989; Di Bisceglie *et al.*, 1989).

The ideal method for detecting anti-core antibodies would be with an ELISA assay, where recombinant core protein would

Figure 3: Detection of Anti-p22 in Hepatitis C Patients. The first case (a) shows a typical course of chronic hepatitis C with multiple peaks of ALT. The second case (b) showed another typical course of chronic hepatitis C, having a single ALT peak but maintaining low but significant abnormal ALT value for years. ALT levels and anti-c100-3 levels are shown in Karmen units (KU) and cpm in a radioimmunoassay, respectively. Anti-p22 activity is shown as a percentage of the positive control to minimize the difference of A_{410} values in different plates. The positive control sample was from donor blood that was positive for anti-c100-3 and induced hepatitis C in the patient. **(Reproduced from Chiba *et al.*, 1991)**


a
Anti-p22 (% of positive control)
Anti-C100 (cpm x10⁻³)
ALT (KU)
Anti-p22
Anti-C100
ALT
Time After Transfusion (week)


b
Anti-p22 (% of positive control)
Anti-C100 (cpm x10⁻³)
ALT (KU)
Anti p22
Anti C-100
ALT
Time After Transfusion (week)

be used to capture antibodies. Ideally, this recombinant antigen would mimic the protein expressed during HCV infection, ensuring maximum recognition by circulating antibodies. A source of recombinant core protein would also permit structural studies of the protein.

SPECIFIC AIMS

1. Chemically synthesize the hepatitis C core gene using overlapping synthetic oligonucleotides.

2. Express hepatitis C core protein using an *E.coli* expression system and develop a purification protocol for the protein.

3. Study the structural and immunological characteristics of the purified protein.

BACKGROUND

Chemical Gene Synthesis

The chemical synthesis of DNA, in the form of single stranded oligonucleotides, has enabled complete genes to be synthesized in an entirely chemical manner. This procedure has many potential applications, the most obvious being the synthesis of genes when a cDNA template is unattainable. Further, the recent acknowledgement of the problem caused by differential codon usage between species in the overexpression of cloned genes means that genes can now be tailor made to suit the codon usage frequency of the species in which the gene is to be expressed, instead of the one from which it is derived. A future application of chemical gene synthesis may be the synthesis of novel genes to direct the expression of man-made proteins or peptides exhibiting novel enzymatic or functional activities.

Currently there are two procedures for chemically synthesizing genes. The first requires the chemical synthesis of the entire gene, i.e. both strands. In this procedure the oligonucleotides are first annealed, and then nicks are repaired by treatment with DNA ligase (Agarwai *et al.*, 1970;

Khorana, 1979). This product can then be cloned directly, or amplified by PCR prior to cloning.

The alternate procedure is likely to be more cost effective as it does not involve complete chemical synthesis of the gene. Instead, pairs of overlapping oligonucleotides are annealed, and then filled-in, using DNA polymerase, to generate a double stranded gene fragment which can be cloned (Rossi *et al.*, 1982; Scapulla *et al.*, 1982). This procedure is less dependent upon the secondary structure of the oligonucleotides, and is generally considered to be more efficient. Current technologies dictate that the maximum fragment length that can be generated is approximately 200bp, since the present upper limit for oligonucleotide synthesis is around 100bp. Therefore, if the final gene product is greater than 200bp, gene fragments must be connected by restriction sites so that they can be assembled into a final, full-length product. While this does impose some degree of inflexibility to the procedure, the degenerate nature of the genetic code, tied to the fact that there are an ever increasing number of restriction enzymes/sites from which to choose, means that this is unlikely to be an unsurmountable problem.

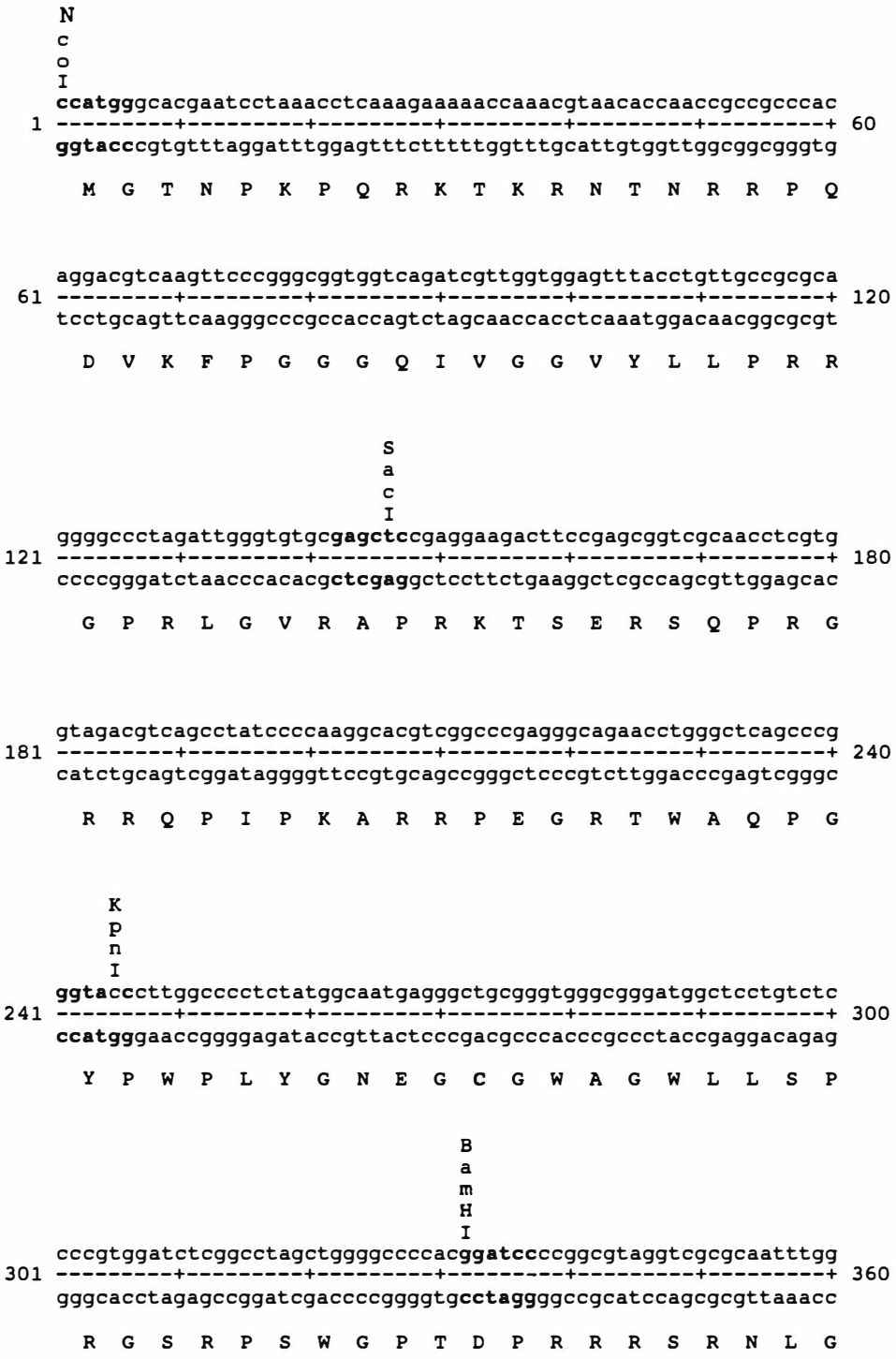
After careful assessment of each procedure it was decided that the second procedure, where overlapping oligonucleotides are filled in using polymerase, would be implemented for the chemical synthesis of the hepatitis C core gene.

Redesigning of the HCV Core Gene to Facilitate it's Chemical Synthesis

Minor alterations were made to the core sequence, subtype 1, to facilitate its chemical synthesis and successful expression in *E.coli*. Firstly the nucleotide sequence was modified at six sites in order to create new and unique restriction sites. This facilitated the synthesis of the gene as five independent fragments, of between 70 and 150 base pairs in length, connected by restriction sites. The alterations in the nucleotide sequence were achieved with only a single change in amino acid sequence, a glycine replacing serine at residue 2. This change was necessary in order to introduce an NcoI restriction site at the 5′ end of the gene for cloning purposes. A termination codon was introduced at the predicted 3′ end of the gene after the final residue, Cys184. This was necessary since the viral genome is translated as a single polyprotein *in vivo* and therefore does not contain a termination codon at the 3′ end of the core gene. Amino acid codons were selected based upon the frequency of usage in *E.coli*, since this would be the species in which expression of the protein would be carried out. Care was taken to avoid the inclusion of codons which are used only rarely by *E.coli*, as it has been widely reported that such 'rare codons' often obviate high level expression of exogenous genes .

The restriction sites which were used to connect the five

Figure 4: Nucleotide and Amino Acid Sequences of the Synthetic HCV Core Gene. The six 'engineered' restriction sites are shown in bold type.


N
c
o
I
ccatgggcacgaatcctaaacctcaaagaaaaaccaaacgtaacaccaaccgccgcccac
ggtacccgtgtttaggatttggagtttcttttggtttgcattgtggttggcggcgggtg
M G T N P K P Q R K T K R N T N R R P Q
aggacgtcaagttcccgggcggtggtcagatcgttggtggagtttacctgttgccgcgca
tcctgcagttcaagggcccgccaccagtctagcaaccacctcaaatggacaacggcgcgt
D V K F P G G G Q I V G G V Y L L P R R
S
a
c
I
ggggccctagattgggtgtgcgagctccgaggaagacttccgagcggtcgcaacctcgtg
ccccgggatctaacccacacgctcgaggctccttctgaaggctcgccagcgttggagcac
G P R L G V R A P R K T S E R S Q P R G
gtagacgtcagcctatccccaaggcacgtcggcccgagggcagaacctgggctcagcccg
catctgcagtcggataggggttccgtgcagccgggctcccgtcttggacccgagtcgggc
R R Q P I P K A R R P E G R T W A Q P G
K
p
n
I
ggtacccttggcccctctatggcaatgagggctgcgggtgggcgggatggctcctgtctc
ccatgggaaccggggagataccgttactcccgacgcccacccgccctaccgaggacagag
Y P W P L Y G N E G C G W A G W L L S P
B
a
m
H
I
cccgtggatctcggcctagctggggccccacggatccccggcgtaggtcgcgcaatttgg
gggcacctagagccggatcgaccccggggtgcctaggggccgcatccagcgcgttaaacc
R G S R P S W G P T D P R R R S R N L G

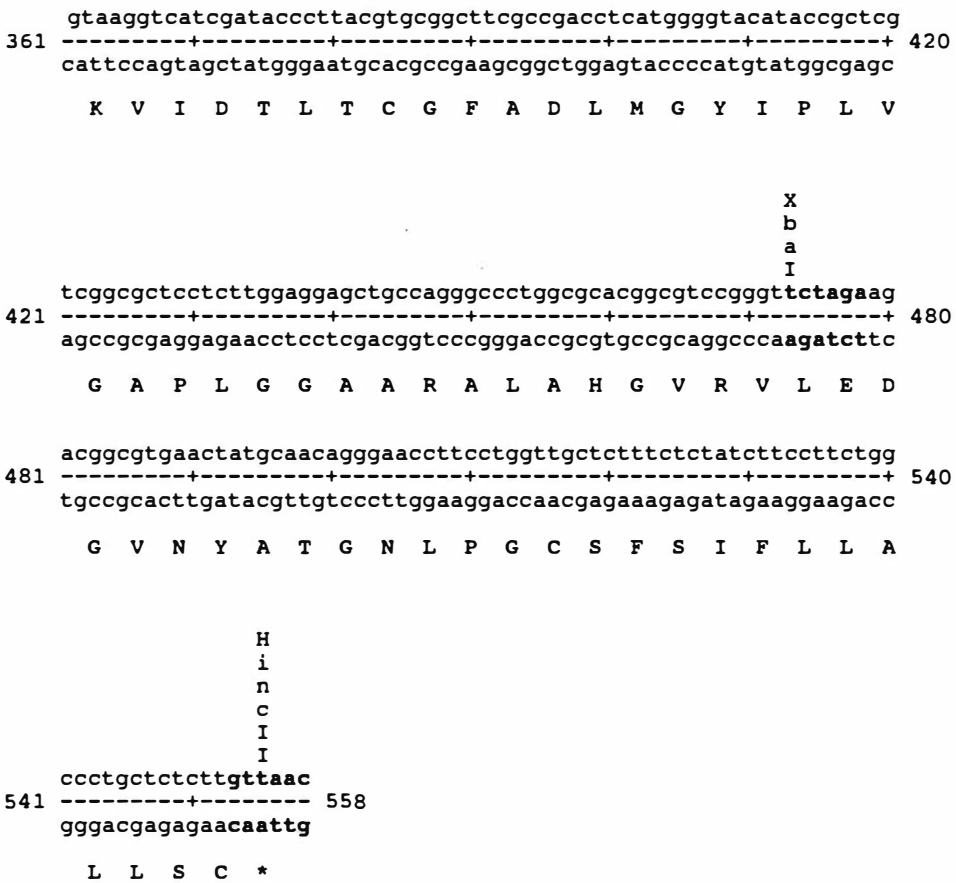

gtaaggtcatcgataccttacgtgcggcttcgccgacctcatggggtacataccgctcg
361 ---------+---------+---------+---------+---------+---------+ 420
cattccagtagctatgggaatgcacgccgaagcggctggagtaccccatgtatggcgagc
K V I D T L T C G F A D L M G Y I P L V
X
b
a
I
tcggcgctcctcttggaggagctgccagggccctggcgcacggcgtccgggttctagaag
421 ---------+---------+---------+---------+---------+---------+ 480
agccgcgaggagaacctcctcgacggtcccgggaccgcgtgccgcaggcccaagatcttc
G A P L G G A A R A L A H G V R V L E D
acggcgtgaactatgcaacagggaaccttcctggttgctctttctctatcttccttctgg
481 ---------+---------+---------+---------+---------+---------+ 540
tgccgcacttgatacgttgtcccttggaaggaccaacgagaaagagatagaaggaagacc
G V N Y A T G N L P G C S F S I F L L A
H
i
n
c
I
I
ccctgctctcttgttaac
541 ---------+-------- 558
gggacgagagaacaattg
L L S C *

Figure 5: Nucleotide Sequence of the Five Pairs of Oligonucleotides used to Construct the Synthetic HCV Core Gene. Oligonucleotides are written 5′ to 3′, and the regions of overlap within pairs are indicated by bold type.

Primer 1a
ccatgggcacgaatcctaaacctcaaagaaaaaccaaacgtaacaccaaccgccg
cccacaggac**gtcaagttcccgggcggtgg**

Primer 1b
gagctcgccacacccaatctagggcccctgcgcggcaacaggtaaactccaccaa
cgatctga**ccaccgcccgggaacttgac**

Primer 2a
gagctccgaggaagacttccgagcggtcgcaacctcgtggtagac**gtcagcctat**
ccccaaggca

Primer 2b
ggtacccgggctgagcccaggttctgccctcgggccgacg**tgccttggggatagg**
ctgac

Primer 3a
ggtacccttggcccctctatggcaatgagggctgcgggtgg**gcgggatgggcggg**
atggctc

Primer 3b
ggatccgtggggccccagctaggccgagatccacggggagacagg**agccatcccg**
cccacccgc

Primer 4a
ggatccccggcgtaggtcgcgcaatttgggtaaggtcatcgataccttacgtg
cggcttcg**ccgacctcatggggtacata**

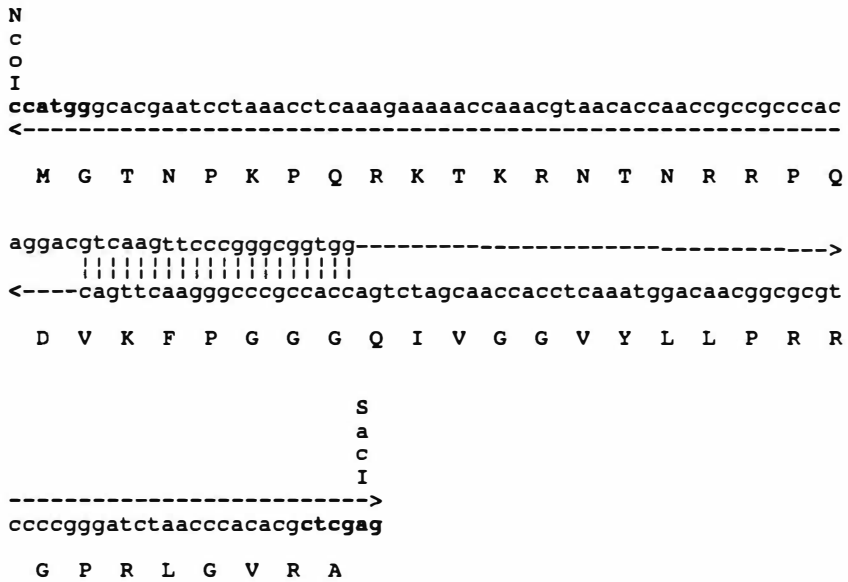
Primer 4b
tctagaacccggacgccgtgcgccagggccctggcagctcctccaagaggagcg
ccgacgagcgg**tatgtaccccatgaggtcgg**

Primer 5a
tctagaagacggcgtgaactatgcaacaggga**accttcctggttgctctttc**

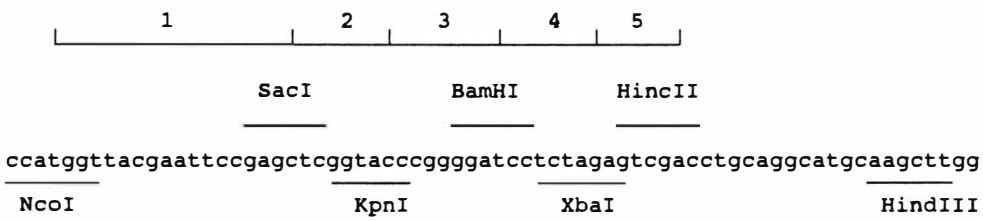
Primer 5b
gttaacaagagagcagggccagaaggaagatagag**gaaagagcaaccaggaaggt**

Figure 6: Cloning Strategy for HCV Core Gene Fragments. **Panel A -** Schematic representation of the 'annealing' and 'filling-in' reaction for the synthesis of fragment 1 using oligonucleotides 1a and 1b. **Panel B -** Nucleotide sequence and relevent restriction sites in the multiple cloning site of pUC18N to show the sites of insertion of each of the five DNA fragments.

A


N
c
o
I
ccatggcacgaatcctaaacctcaaagaaaaaccaaacgtaacaccaaccgccgcccac
<----------
M G T N P K P Q R K T K R N T N R R P Q
aggacgtcaagttcccgggcggtgg----------
<----cagttcaagggcccgccaccagtctagcaaccacctcaaatggacaacggcgcgt
D V K F P G G G Q I V G G V Y L L P R R
S
a
c
I
---------->
ccccgggatctaacccacacgctcgag
G P R L G V R A

B


1
2
3
4
5
SacI
BamHI
HincII
ccatggttacgaattccgagctcggtacccggggatcctctagagtcgacctgcaggcatgcaagcttgg
NcoI
KpnI
XbaI
HindIII

gene fragments were selected to correspond to sites in the multiple cloning site of pUC18N. This vector is essentially pUC18 except that it contains an NcoI restriction site immediately downstream of the multiple cloning site (Zheng *et al.*, 1992). The order of the restriction sites made it possible to sequentially move down the pUC18N multiple cloning site and insert fragment 1, then fragment 2, and so on until all 5 fragments had been cloned.

Predicted Features of HCV Core Protein

Successful expression of the synthetic HCV core gene would yield a protein of 184 amino acids in length, and having a predicted molecular size of 20,600 Daltons. The amino acid sequence has no potential glycosylation sites.

Probably the most striking feature of this protein is it's extreme basicity, which is predicted to be 11.94. The protein has 24 arginine residues (13%) and 7 lysine residues (3.8%), amounting to 31 basic residues out of a total of 184 (16.8%).

Expression of Viral Core Proteins in *E.coli*

E.coli was selected for the expression of HCV core protein for several reasons. First and foremost viral core proteins are typically not glycosylated, or targets of other post-translational modification, rendering a Prokaryotic

system such as *E.coli* an ideal choice for expression. In addition to giving high levels of expression, *E.coli* is undoubtedly the best understood system for expression of exogenous genes, and the DNA manipulations involved in creating expression vectors are relatively simple. Further, the many different expression plasmids and cell strains provides a multitude of different expression systems. Finally, the culture media required for growing bacteria is relatively cheap, an important consideration when making large scale fermentations.

Several viral core proteins have been successfully purified following their over-expression in *E.coli*. Notably the HBV core protein (Stahl et al., 1982) and HIV core protein (Ehrlich et al., 1992), both of which assemble into core particles.

MATERIALS AND METHODS

Chemical Synthesis of the HCV Core Gene

Synthesis and Preparation of Oligonucleotides

Oligonucleotides were synthesized by the 'MCV-VCU Nucleic Acid Core Facility'. To prepare the oligonucleotides for use, they were first 'de-blocked' by heating to 55°C, overnight (or 85°C, for 1 hour), in tightly sealed tubes. The detritylated primers were then dried under vacuum, resuspended in 1ml of distilled water, and extracted three times with 0.5ml of ethyl ether. Finally the oligonucleotides were dried and resuspended in 0.6ml of distilled water. The concentration was determined by measuring the absorbance at 260nm, and dividing this figure by the extinction coefficient for the oligonucleotide (A=16,000,G=12,000,C=7,000,T=9,600 $M^{-1}cm^{-1}$).

Annealing, Filling-In, and PCR Amplification of Gene Fragments

Ten oligonucleotide primers were synthesized as five pairs (Figure 5), where each pair corresponded to one fragment of the gene, and had a 20 base pair overlap at their 3' end such that they could be annealed. Once annealed, a pair of primers created a single fragment which stretched from one

restriction site to the next but was only double stranded in the middle where the 20 base pair overlaps had annealed. Following annealing at 42°C, for 2 minutes, the DNA was made double stranded by five cycles of 'filling-in' at 72°C for 30 seconds, and 'melting' at 94°C for 1 minute. This does not achieve amplification of the DNA, but it does maximize the amount of full length, double stranded DNA produced (Figure 6). The reaction volume was 100μl, and conditions for oligonucleotide extension were as follows: 250nM of each oligonucleotide, 400μM dNTP's, 1x Vent buffer, 1 unit Vent polymerase (all New England Biolabs). This strategy proved to be successful for the cloning of fragments 2,3 and 5. However, fragments 1 and 4 could not be cloned using this procedure. It was inferred that these two fragments, which incorporate longer oligonucleotides than fragments 2,3 and 5, possibly contained less full length product following the 'filling-in' step. To circumvent this possibility, pairs of short oligonucleotides (20-mers) which were complementary to the 3' ends of the filled-in fragment (Figure 7), were used to amplify the fragment prior to cloning. Because the 3' ends of the fragments are only present after successful primer extension, the PCR amplification was specific for full-length product. 2μl of the fill-in reaction was used as the template for this PCR reaction. The reaction conditions were as follows: 400μM dNTP's, 1μM of each 20-mer primer, 1 unit of Vent polymerase in 1x Vent buffer.

5′ g g**cc atg** g**g**c acg aat cct aaa cct (F1-fwd, NcoI)

5′ c**g**g **agc tc**g cac acc caa cct gg (F1-rev, SacI)

5′ ac**g gat cc**c cgg cgt agg tc (F4-fwd, BamHI)

5′ ct**t cta ga**a ccc gga cgc cg (F4-rev, XbaI)

Figure 7: Sequences of the Oligonucleotide Primers used to Amplify HCV Core Gene Fragments 1 and 4 (F1 and F4). Restriction sites are highlighted with bold type.

Cloning of Double Stranded Gene Fragments

DNA was extracted directly from PCR reactions using a commercially available kit ('Mermaid', BIO101), digested with the appropriate pair of restriction enzymes (NEB), and cloned into the multiple cloning site of pUC18N. 'Mini-prep' plasmid isolations were performed using the alkaline lysis method (Sambrook *et al.*, 1989), and plasmid stocks of positive clones were made using a commercially available plasmid isolation kit (Qiagen).

DNA Sequencing

The sequence of the final construct, pUC18N-C184, was confirmed by automated DNA sequencing performed by the 'MCV-VCU Nucleic Acid Core Facility.'

Expression and Purification of Core Protein

Expression in E.coli

The entire core gene was excised from pUC18N-C184 using the restriction enzymes NcoI and HindIII. This 570bp fragment was inserted into pET3d (Novagen) at the NcoI/HindIII sites, creating pET3d-HCVC (Figure 8). Digestion of pET3d with these enzymes excises the T7 terminator; however, this was expected to have a minimal effect on gene expression. The expression plasmid (pET3d-HCVC) was transformed into the *E.coli* strain HMS174(DE3)(Novagen) and grown at 37°C in M9 media (Sambrook *et al.*, 1989) supplemented with 50μg/ml ampicillin. When the

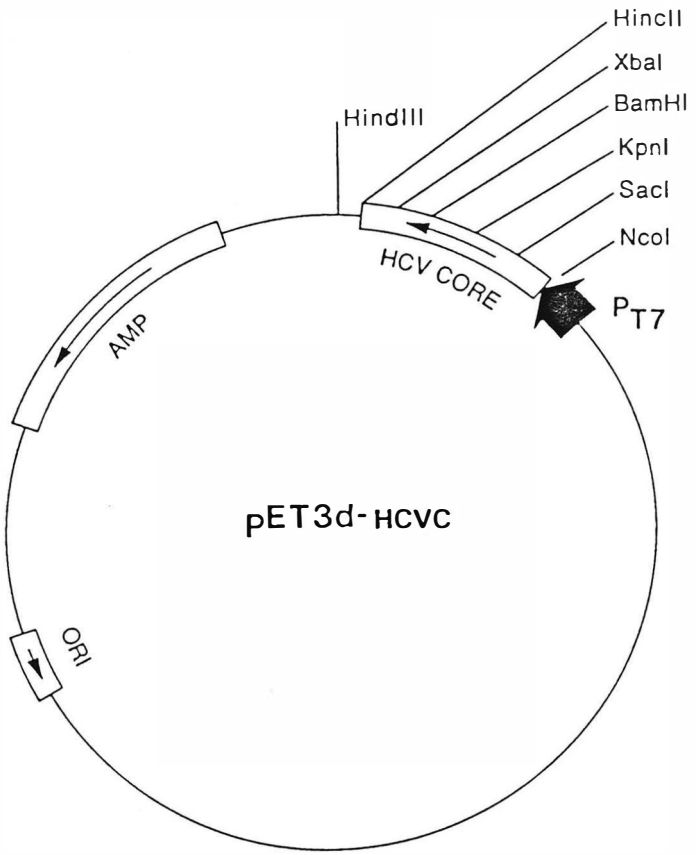


Figure 8: PET3d-HCVC expression plasmid.

cell density reached an A_{600} of 0.6-0.8 the cells were induced by the addition of 0.4mM IPTG and grown for an additional 3 hours. Cells were harvested by centrifugation at 5000xg for 10 minutes. Dry cell pellets were stored at -70°C until needed.

Purification of Core Protein

Cell pellets were suspended in one-tenth of the original culture volume of 50mM Na-phosphate, pH 6.8. The cells were lysed by a single passage through a french pressure cell, and centrifuged at 18,000xg for 30 minutes to pellet cellular debris. The soluble cell lysate was applied to a hydroxyapatite column which had been pre-equilibrated with 10mM sodium phosphate, pH6.8. The column was washed with 150mM sodium phosphate, pH 6.8, until the absorbance (280nm) of the eluent fell below 0.1. The protein was eluted from the column by washing with 250mM sodium phosphate, pH 6.8. The protein in this fraction was concentrated by precipitation with 25% polyethylene-glycol 8000 and recovered by centrifugation at 10,000xg for 30 minutes. The precipitate was resuspended in a minimal volume of 10mM tris-Cl, pH7.8, and subjected to gel filtration chromatography using a sepharose 4B column (Pharmacia). Fractions were tested for core protein by immmunobotting with a monoclonal anti-core antibody (410G5H4).

SDS-PAGE

Analytical polyacrylamide gel electrophoresis was performed according to the method of O'Farrell (O'Farrell, 1975). 4% acrylamide stacking gel and 15% acrylamide resolving gel were employed.

Characterization of Core Protein

Ultra-Violet Spectroscopy

Ultra-violet spectroscopy was performed using a Hewlett Packard 8451A diode array spectrophotometer.

Gel Filtration Chromatography

Gel filtration was carried out using a 'Superose 6' column attached to a Pharmacia FPLC system. The buffer was 10mM tris-Cl, pH8.0, and the flow rate was 0.4ml/min.

Electron Microscopy

Purified HCV core protein (2mg/ml) was applied to the grid and negatively stained with 1% phosphotungstic acid. The protein was examined using a Philips EM400 elecrton microscope at a screen magnification of x 130,000.

Immunological Studies of Core

Western Blotting

Following the separation of protein samples on 15% SDS-

PAGE gels, the protein was blotted onto PVDF membrane (Millipore), according to the method of Burnette (1981). Transfer was achieved using the following buffer composition; 192mM glycine, 20% methanol, 1g/L SDS, 10mM tris-Cl, pH 8.5, for 1 hour at 200mA. Membranes were blocked in a 4% solution of BSA (Sigma) in PBS for 30 minutes. Next the membranes were incubated with an anti-core monoclonal antibody in PBS supplemented with 1% BSA overnight. The following morning membranes were washed and incubated with a 1:3000 dilution of rabbit anti-mouse peroxidase antibody (Sigma) for 1 hour. Finally, the membranes were washed extensively and developed using 8ml PBS, 2ml chloronapthol (Sigma) in methanol (3mg/ml), and 8μl hydrogen peroxide (30%) (Fisher).

Monoclonal Antibodies to HCV Core

The two anti-core monoclonal antibodies (410G5H4 & 46E7F6) were a gift from Ortho Diagnostics (Raritan, NJ). They were both raised against a synthetic peptide, corresponding to amino acids 29-43 of the HCV polyprotein, having the following sequence, QIVGGVYLLPRRGPR.

Serology of HCV Positive Sera

A series of well-characterized HCV positive serum samples were also a gift from Ortho Diagnostics. Serum samples 4384, GC2230792, and GC2230800 were reactive to all 4 proteins (5-1-1, c100-3, c22-3 and c33c) on RIBA-2, sample

4523 was reactive to c33 and c22-3 only, and samples 12-080 8906 and GJ61329 were reactive to c22-3 only.

ELISA to Detect Anti-Core Antibodies

Purified HCV core protein, coated to the wells of microtitre plates, was used in an ELISA to capture anti-core antibodies. The assay was tested with human serum, known to have anti-HCV core antibodies, and with two monoclonal antibodies in order to determine whether they could bind to core protein under native conditions.

Plates were coated overnight at room temperature at a protein concentration of 25μg/ml in bicarbonate buffer, pH 9.2. Next morning the plates were blocked with washing buffer (1hr), dried, and stored at -20°C until needed.

Monoclonal anti-core antibodies were serially diluted in dilution buffer (Ortho Diagnostics), and 50μl of each dilution was added to the microtitre wells. Following a 30 minute incubation the wells were washed and 50μl of a 1:3000 dilution of peroxidase conjugated anti-mouse antibody (Sigma) was added and incubated for 10 minutes. Following extensive washing, 50μl of TM Blue substrate (TSI) was added to each well and allowed to develop for 5 minutes. The reaction was stopped by the addition of 100μl of 1N sulfuric acid and the plates were read at a wavelength 450nm.

To detect human anti-core antibodies the procedure was similar to that described above, except for the following

points. Serum was diluted 200 and 1000-fold prior to testing, and an anti-human peroxidase conjugated antibody (1:3,500)(Ortho Diagnostics) was used to detect bound antibody.

Construction of Plasmids for the Expression of Histidine-Tagged Core Proteins

In an attempt to develop a more rapid and reproducible purification procedure for hepatitis C core protein, a series of vectors which express core protein as a poly-histidine tagged protein were constructed. Specifically, the 'pQE' family of vectors (Qiagen) were tested for their ability to direct the expression of core protein. The polyhistidine (6x His) can be incorporated at either the amino or carboxyl-terminus of proteins which, via the strong affinity of poly-histidine sequences for nickel, permits rapid purification of proteins using nickel chelation chromatography (Hochuli, 1990).

Construction of pQE30-HCVC and pQE60-HCVC

The plasmid pQE30 directs the high-level expression of cloned proteins, complete with an amino-terminal poly-histidine affinity tag, in *E.coli*. To construct the plasmid pQE30-HCVC it was necessary to introduce a BglII restriction site at the 5' end of the core gene; BamHI was not an alternative because the gene already contains an internal

BamHI restriction site (Figure 9). BglII was chosen because it is complementary to BamHI, by virtue of their common 'over-hang' sequence (GATC), and it is also located immediately downstream of the 6 histidine codons. Although a variety of other cloning sites were available, cloning into the BamHI site of the plasmid reduced the number of extraneous amino acids inserted between the histidine tag and the start of the core gene to just two (arginine and serine). The HindIII site at the 3' end of the core gene in the original pUC18N-C184 clone, which was used as the template for PCR, was suitable for the 3' end of the gene since pQE30 contains a HindIII restriction site in it's multiple cloning site. Therefore, a perfect match M13 reverse sequencing primer was a suitable choice to serve as the reverse sequencing primer.

While the plasmid pQE30 is suitable for constructing expression vectors which include an amino-terminal histidine tag, pQE60 allows the construction of expression vectors which insert a carboxyl-terminal tag onto cloned proteins (QiaExpressionist, 2nd Edition). To construct the plasmid pQE60-HCVC it was necessary to design a mismatch primer which removed the stop codon from the original clone, as well as insert an 'in-frame' BglII restriction site at the 3' end of the gene. This allowed the poly-histidine codons to be read, 'in-frame', directly after the gene. In this instance BglII was chosen over BamHI for two reasons. Firstly, the presence of the aforementioned internal BamHI in the core gene, and

secondly, the insertion of the gene at the BglII restriction site of the plasmid reduces the number of extraneous amino acids between the 3′ end of the core gene and the six histidine codons. Like pQE30-HCVC, it was not possible to avoid the insertion of an arginine and a serine between the core gene and the histidines. The NcoI restriction site in pQE60 was used for insertion of the 5′ end of the gene. Thus it was possible to use the 'perfect match' NcoI primer used during the chemical synthesis of the core gene as the forward primer (Figure 7).

Both modified HCV core genes were amplified using PCR, and cloned into their respective vectors. *E.coli*, strain M15, was used for all steps using pQE vectors.

Construction of pET3d-HCVC(C-His)

To construct this plasmid the HCV core gene, complete with it's carboxyl-terminal histidine tag, was excised from pQE60-HCVC as an NcoI-HindIII fragment. This DNA was cloned directly into the plasmid pET3d, which had itself been prepared by digestion with the same two restriction enzymes.

Construction of pET3d-HCVC(N-His)

In order to construct this plasmid it was necessary to simultaneously introduce an NcoI restriction site upstream of the amino-terminal histidines, as well as delete the original NcoI site at the start of the core gene. This was achieved

Figure 9: Construction of Core-Histidine Fusions.
Primer sequences and templates used for the construction of plasmids harboring the HCV core gene with amino and carboxyl-terminal histidine tags.

pQE60-HCVC

Template: pUC18N-HCVC

Fwd: 5′ G GCC ATG GGC ACG AAT CCT AAA CCT 3′

Rev: 5′ CGG AGA TCT ACA AGA GAG CAG GGC CAG 3′

pQE30-HCVC

Template: pUC18N-HCVC

Fwd: 5′ GC AGA TCT ACG AAT CCT AAA CCT CAA 3′

Rev: Standard M13/pUC sequencing primer (-47)

pET3d-HCVC(C-His)

No PCR necessary. HCVC gene (NcoI/HindIII) from pQE60-HCVC cloned directly into pET3d.

pET3d-HCVC(N-His)

Template: pUC18N-HCVC

Fwd: 5′ G GCC ATG GGC CAT CAC CAT CAC CAT CAC GGC ACG AAT CCT AAA CCT 3′

Rev: Standard M13/pUC sequencing primer (-47)

using the primer shown in figure 9. The initiation codon (ATG) at the 5′ end of the gene was also deleted. This was done to prevent it serving as a site of initiation, in which case it would have initiated translation of a truncated protein lacking the six histidine residues. The M13 reverse sequencing primer was used as the reverse primer, and the template was pUC18N-C184. The PCR product was digested with the enzymes NcoI/HindIII and ligated into pET3d which had been prepared by digesting it with the same pair of enzymes.

Expression of Histidine-Tagged Core Proteins

pQE60-HCVC and pQE30-HCVC were transformed into *E.coli*, strain M15; while the pET constructs, pET3d-HCVC(C-his) and pET3d-HCVC(N-his), were used to transform *E.coli* strain HMS174(DE3). Single colonies were used to inoculate 3ml cultures of M9 media, supplemented with 50μg/ml ampicillin (pET/HMS174(DE3)); or 3mls of LB media, supplemented with 50ug/ml ampicillin and 25μg/ml kanamycin (pQE/M15). The cultures were grown at 37°C overnight, and then used to inoculate 50mls of the same media. The 50ml cultures were grown at the same temperature until their optical density (600nm) measured 0.6-8.0. The cells were induced by the addition of 1mM IPTG and grown for a further 3 hours. To test for expression of the histidine tagged HCV core fusion protein, 1ml of cell culture was pelleted and the cells

suspended in 100μl of SDS-PAGE sample buffer, boiled, and analyzed by both SDS-PAGE and immunoblotting using a monoclonal anti-HCV core antibody.

Purification of Histidine-Tagged Core Proteins

The most efficient expressor of HCV core protein bearing a poly-histidine tag (pET3d-HCVC(C-his) was grown and induced in the manner described previously. This protein accumulated as insoluble aggregates following high-level expression, therefore it was necessary to develop a denaturing purification scheme in order to purify this protein. Two different schemes, utilizing either 8M urea, or 6M guanidine-HCL, were used in an attempt to purify the HCV core protein bearing a carboxyl-terminal histidine tag.

The cell pellet from one liter of cell culture was resuspended in 100ml of 100mM Na-phosphate, 10mM tris-cl, pH 8.0, supplemented with either 6M GuHCl (A) or 8M urea (B). Cells were lysed by a single passage through a French pressure cell. The lysate was spun at 16,000xg for 20 minutes to pellet insoluble material. Soluble protein was mixed with 10ml of a 50% slurry of Ni-NTA resin (Qiagen), and stirred gently for 1 hour. Unbound protein was washed from the column by washing with buffer B, supplemented with 10mM imidazole. The histidine-tagged core protein was then eluted from the column with buffer B, supplemented with 100mM EDTA.

Generation of Polyclonal Antisera to Core Protein

Briefly, 100μl of core(C-his) protein (1mg/ml) was diluted to 500μl with PBS and mixed 1:1 with Freund's complete adjuvent. Following brief sonication, the homogenous emulsion was administered to a single New Zealand white rabbit by making multiple injection, subcutaneously, on the back of the rabbit. One month later the rabbit was 'boosted' by injecting antigen in a similar manner, except that the antigen was mixed 1:1 with Freund's incomplete adjuvent. The 'boost' was repeated one month later. Ten days later the serum was tested for the presence of anti-core antibodies using immunoblots and the rabbit sacrificed.

RESULTS

Chemical Synthesis of the HCV Core Gene

The HCV core gene was successfully synthesized, in an entirely chemical manner, using overlapping oligonucleotides. Gene fragments 2, 3 and 5 were successfully cloned by simply annealing, filling-in, digesting and ligating the fragments with pUC18N. However, attempts to clone fragments 1 and 4 by this method failed. This was possibly due to the fact that these fragments incorporated longer synthetic oligonucleotides causing them to have more secondary structure, and/or a higher proportion of truncated species. To overcome this problem short oligonucleotide primers (20-mers), which were complimentary to the 3' end of the 'filled-in fragment', were used to amplify the dsDNA prior to cloning. This approach proved successful for the cloning of the remaining core fragments (1 and 4).

The sequence of the synthesized gene, assembled in pUC18N, was confirmed by automated DNA sequencing (data not shown).

Expression and Purification of HCV Core Protein

The HCV core gene was excised from pUC18N-C184 as an NcoI/HindIII fragment, and cloned into the pET3d expression vector at the same restriction sites. Upon growth and induction of this clone with IPTG core protein was successfully expressed as a 22kDa protein, with a second major species of 18kDa, as determined by reducing SDS-PAGE analysis and immunoblotting (Figure 10). The presence of multiple species is presumably due to premature translation termination, and/or proteolysis. If the latter is the case, we predict that since cells which are boiled directly in SDS-PAGE sample buffer prior to immunoblotting, the proteolysis is due to an intracellular protease. Further, the fact that the cleaved forms of core still blot with the core monoclonal antibody, which is directed to the amino-terminus of the protein, suggests that it is the carboxyl-terminus which is susceptible to proteolysis.

HCV core protein proved extremely difficult to purify, probably in part due to the heterogenic nature of the expressed protein. Despite being expressed at relatively high levels (5-10mg/L), only 100-500 micrograms of protein was typically purified per liter of cell culture.

When analyzed by reducing SDS-PAGE, the purified protein ran as two species with molecular weights of 22 and 16kDa (Figure 11). However, in the absence of reducing agent the protein migrated as a macromolecular structure which failed to

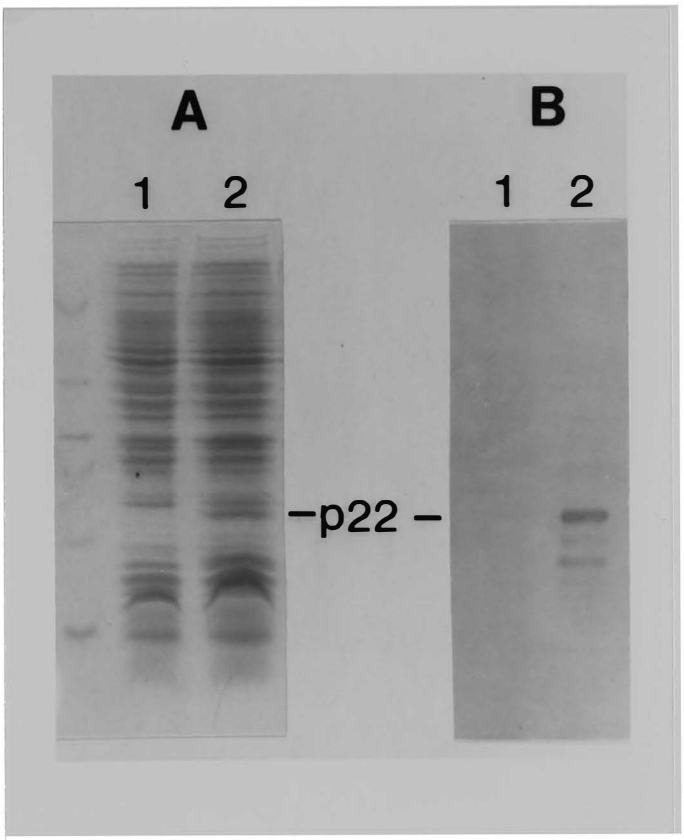


Figure 10: Expression of 22kDa HCV Core Protein Following the Induction of Plasmid pET3d-HCVC with IPTG. SDS-PAGE (A) and immunoblotting (B) with a HCV core specific monoclonal antibody (410G5H4) (Lanes A1 & B1 - uninduced; lanes A2 & B2 - induced).

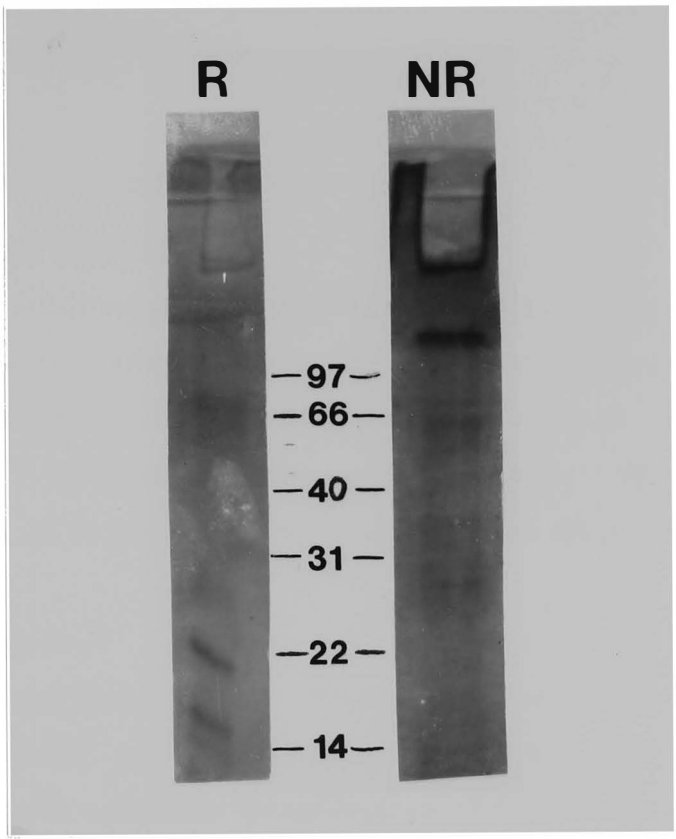


Figure 11: SDS-PAGE Analysis of Purified HCV Core Protein Under Reducing (R) and Non-Reducing (NR) Conditions. Reduced protein migrates as a doublet of 22 and 16 kDa, and unreduced protein accumulate at the top of the resolving gel.

enter the resolving portion of the gel (Figure 11). These observations suggest that, under native conditions, HCV core protein is assembled into a multimeric, disulfide-linked macromolecular structure.

Physical Characterization

Ultra-Violet Spectroscopy

When analyzed by ultra-violet spectroscopy, HCV core protein did not exhibit a profile characteristic of a typical protein (Figure 12). Instead of a maximum at 280nm, core exhibited a maximum at approximately 260nm. This could be indicative of an association of the protein with nucleic acid.

Gel Filtration

Upon analysis by gel filtration chromatography, core protein eluted at a position consistent with it being assembled into a multimeric structure of high molecular weight (Figure 13). Recombinant HBV core protein is known to have a molecular weight in excess of 10^3 kDa (Stahl *et al.*, 1982), and this experiment shows that recombinant HCV core protein has a molecular size in this range. The broad peak exhibited by HCV core protein is indicative of heterogeneity in the molecular size of the purified protein.

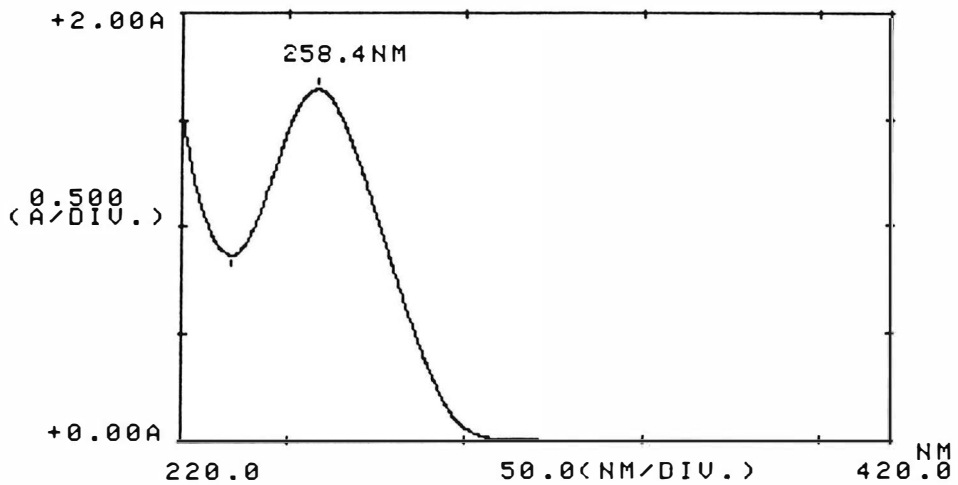


Figure 12: Ultra-Violet Spectrum of Core Protein.

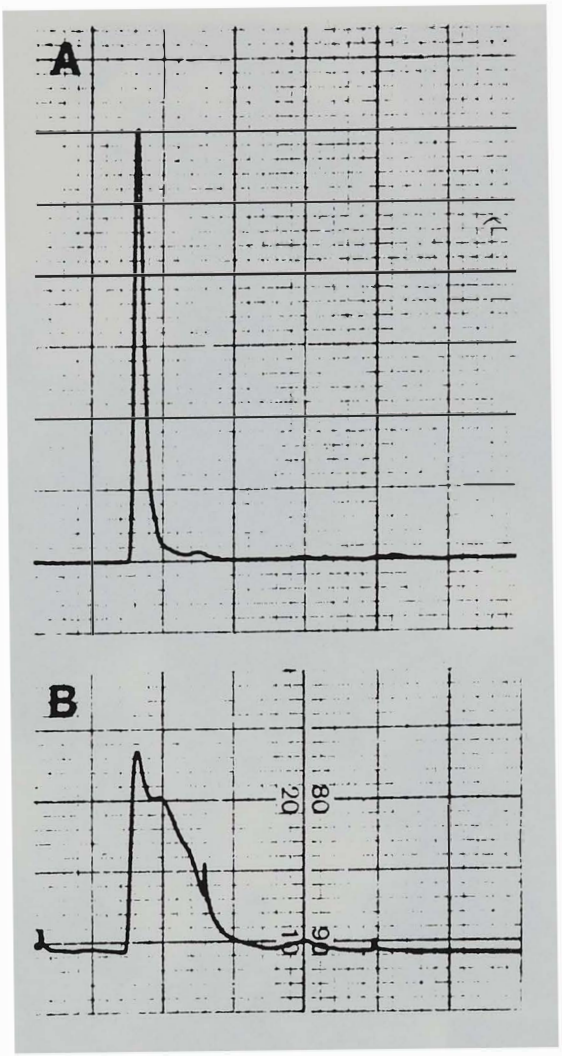


Figure 13: Gel Filtration Analysis of Core Protein. Panel A, recombinant HBV core protein; Panel B, recombinant HCV core protein.

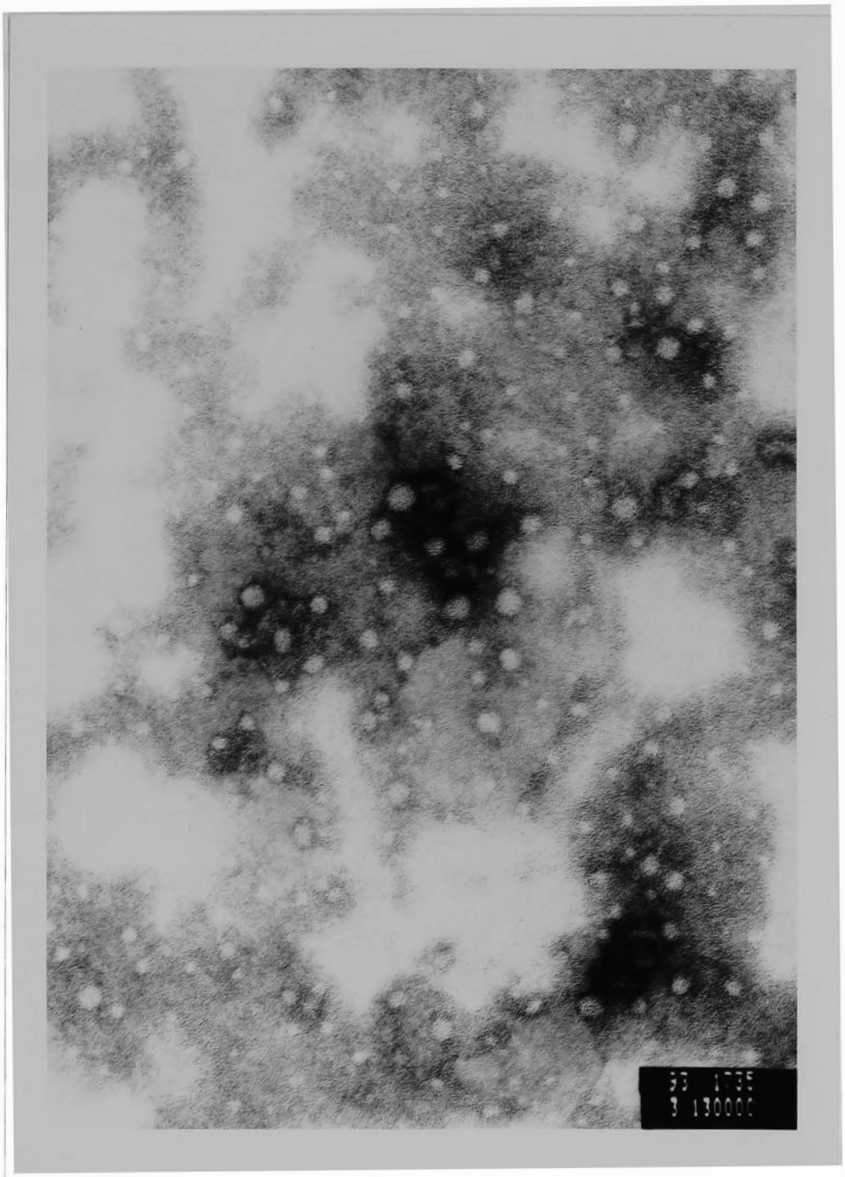


Figure 14: Electron Micrograph of Purified Core Protein. The protein was negatively staining with tungstic acid. A heterogenous mixture of nucleocapsid-like particles with diameters ranging from 12-20nm are visible.

Electron Microscopy

Examination of purified HCV core protein by electron microscopy revealed a heterogeneous mixture of nucleocapsid particles with diameters in the range of 12-20nm (Figure 14).

Immunological Characterization of Core

Recognition of Core Protein by Two Monoclonal Antibodies

Immunoblotting confirmed that the two monoclonal anti-core antibodies were able to specifically bind core protein expressed in *E.coli* (data not shown). In order to determine whether or not the antibodies could bind native core protein, an ELISA was developed and the ability of core protein to capture the monoclonal antibodies was tested. The results, shown in figure 15, highlight the fact that both monoclonal antibodies have a relatively high affinity for native core protein, purified from *E.coli*. Even at antibody dilutions of 10^5, antigen-antibody associations can still be detected.

These observations suggest that the amino acid sequence to which the monoclonal antibodies were directed is accessible at the surface of the protein. Further, it demonstrates that core protein was successfully coated to the wells of the microtitre plates under the 'coating conditions' described under 'materials and methods'.

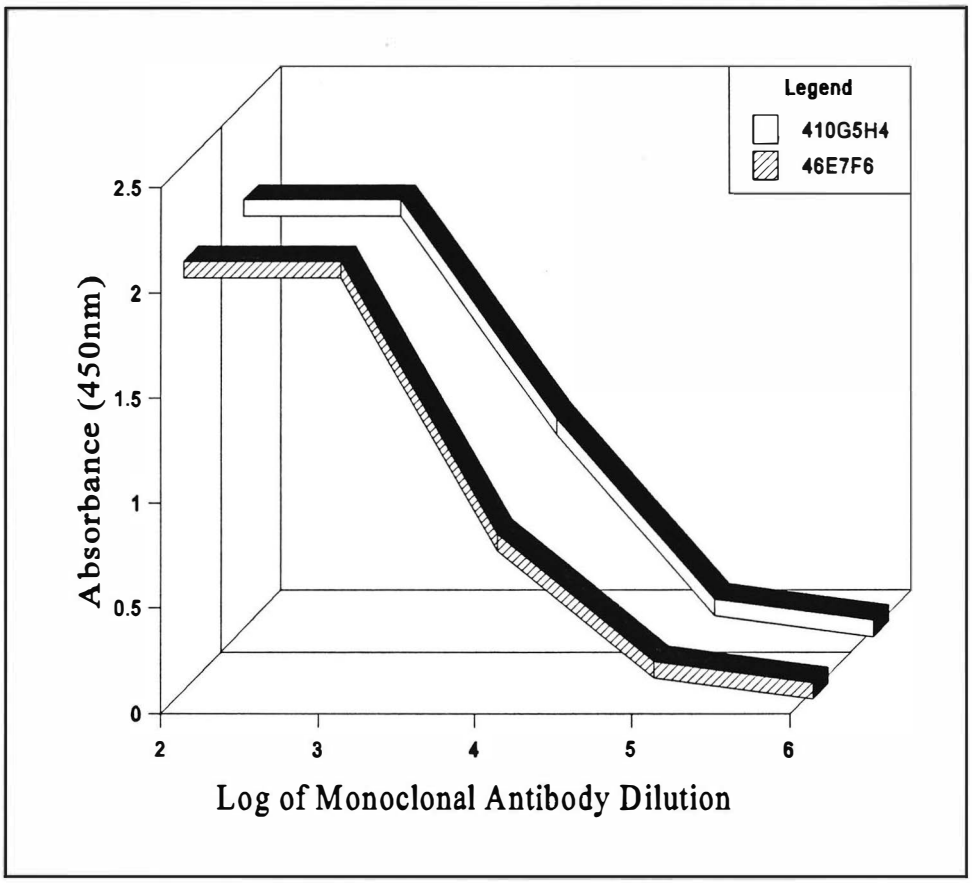


Figure 15: Recognition of Native Core Protein by Two Monoclonal Anti-Core Antibodies.

ELISA For Detecting Anti-HCV Core Antibodies in Patients' Sera

A series of well characterized HCV positive serum samples were examined for the presence of anti-core antibodies using the newly developed ELISA. Samples were tested at dilutions of 1:200 and 1:1000. The results, shown in table 1, confirm that anti-core antibodies were detected in all 6 sera tested, at both dilutions.

Histidine-Tagged HCV Core Protein

Cloning and Expression

In an attempt to purify milligram quantities of HCV core protein in a reproducible and rapid manner, a new expression vector was constructed. Polyhistidine-core fusions, with the polyhistidine residues localized at the carboxyl or amino-terminus, were created using a commercially available kit (QiaExpress, Qiagen Inc.). It was envisaged that expression of the protein using this system would; i) afford a rapid and simple purification of the protein using NTA-resin (Qiagen); and ii) allow the purification of only full length core protein by virtue of the fact that truncated core protein would have lost the polyhistidine tag, and therefore not stick to the NTA resin.

Polyhistidine-core expression vectors, constructed using the commercially available pQE expression system, proved to be extremely poor for expressing poly-histidine tagged core protein. Both amino (pQE30-HCVC) and carboxyl-terminal

Table 1: Detection of Anti-Core Antibodies in HCV Infected Sera.
All samples, except NHS (normal human serum) were previously shown to be positive for HCV infection by a variety of commercial tests.

SAMPLE ID	HCV CORE REACTIVITY 1:200	1:1000
NHS	0.12	0.00
4523	2.00+	2.00+
4384	2.00+	2.00+
12-080 8906	2.00+	1.25
GJ61329	2.00+	2.00+
GC 2230792	2.00+	2.00+
GC2230800	2.00+	1.62

histidine fused constructs (pQE60-HCVC) exhibited levels of protein expression which were barely detectable by immunoblot analysis (Figure 16, lanes 2 & 3).

In an attempt to increase expression levels the histidine tagged core genes were transferred to the pET-3d expression plasmid, a system which had previously been shown to successfully express native core protein at relatively high levels (Figure 10). In pET3d-HCVC(N-his) the original initiator methionine residue for core was deleted in order to prevent initiation from this point, which would result in core protein lacking the amino-terminal polyhistidine fusion. This was achieved by synthesizing a mis-match primer which inserted an NcoI site 5′ of the six histidine codons, and deleted the original NcoI site. Using these vectors, both the amino-terminal and carboxyl-terminal histidine tagged proteins were expressed at levels of between 5 and 10mg/L. Core protein bearing an amino-terminal tag, like the original untagged protein, was proteolyzed at it's carboxyl-terminus (Figure 16, lane 6). The protein was soluble, however it did not readily bind to the NTA-resin under native conditions. This observation, and the fact that this protein did bind to the resin under denaturing conditions (data not shown), is suggestive that the amino terminus of the protein is not exposed under native conditions.

Surprisingly, core protein bearing a carboxyl-terminal histidine tag, unlike its un-tagged counterpart and the amino-

terminal tagged protein, was insoluble and accumulated as inclusion bodies following expression. This presumably accounts for the observation that, unlike the native protein, this core-histidine fusion was not cleaved during expression (Figure 16, lane 5). This observation further supports the premise that the truncations observed with pET3d-HCVC are due to proteolysis by an intracellular protease, and not by premature translation termination. Presumably the expressed protein is not accessible to intracellular proteases when it is insoluble in the cytoplasm.

Purification of Core (C-His) Using Affinity Chromatography

Since both amino and carboxyl-histidine-tagged core proteins required denaturing conditions to expose the histidine tag, it was decided that the protein tagged at the carboxyl-terminus would be the best to pursue. This protein was expressed as a single species (Figure 16, lane 5; Figure 17), unlike the amino-terminal tagged protein, which was proteolyzed (Figure 16, lane 6). A denaturing purification scheme was developed to accomodate the fact that the carboxyl-terminus tagged protein was expressed in an insoluble form. While 8M urea was successful in solubilizing the protein, it failed to stick to the nickel column in this buffer (data not shown). Urea is considered a 'weak' protein denaturant, and this is presumably why the histidine tag is insufficiently exposed to interact with the nickel column under these

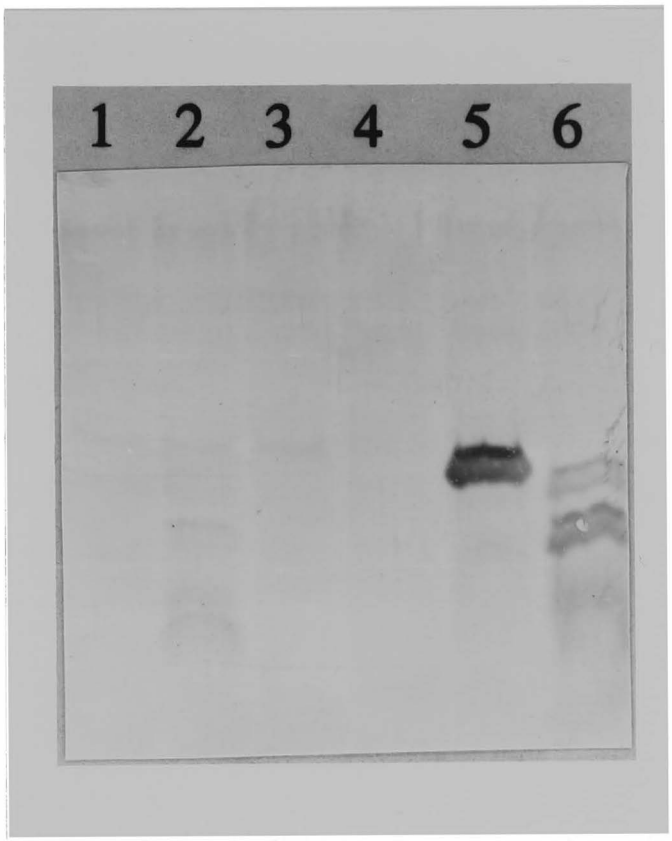


Figure 16 : Immunoblot to show the Expression of Histidine-Tagged Core Protein Using pET and pQE Expression Plasmids. Lane 1, M15[pRep4/pQE30] uninduced; Lane 2, M15[pRep4/pQE30] induced; Lane 3, M15[pRep4/pQE60] induced; Lane 4, HMS174(DE3) [pET3d-HCVC(C-His)] uninduced; Lane 5, HMS174(DE3) [pET3d-HCVC(N-His)] induced; Lane 6, HMS174(DE3) [pET3d-HCVC(C-His)] induced.

conditions. However, when the cells were lysed in a buffer containing 6M guanidine hydrochloride, a 'strong' protein denaturant, the protein stuck to the nickel column. Despite extensive washing of the column, and the use of an imidazole gradient to elute the protein, a small amount of contaminating proteins co-eluted with the core protein (Figure 18). The protein was determined to be approximately 90% pure according to SDS-PAGE.

Polyclonal Anti-Core(C-his) Antisera

Anti-sera raised against purified core(C-his) specifically recognized both histidine tagged and native HCV core protein when tested using immunoblots (Figure 19).

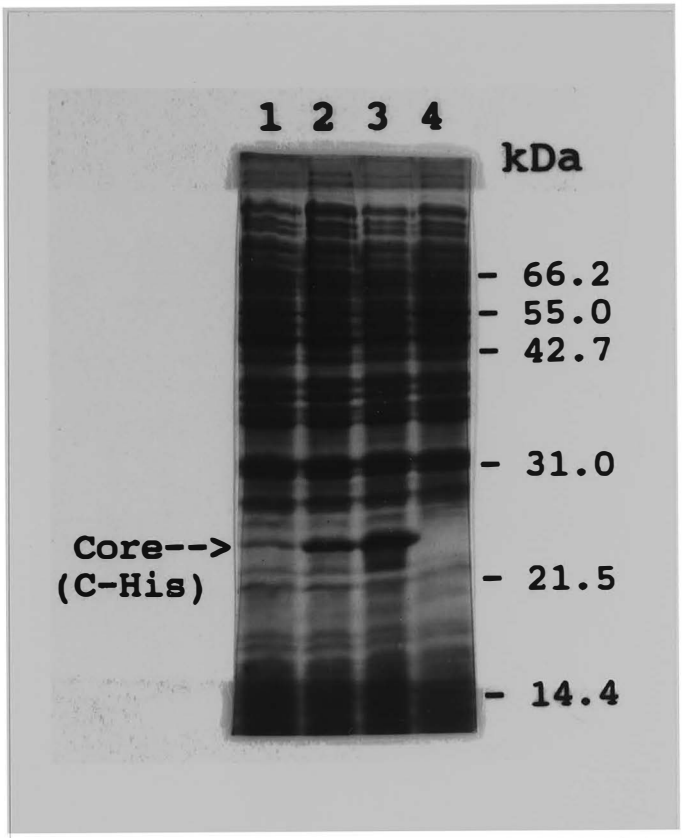


Figure 17: SDS-PAGE to Show the Expression of HCV Core (C-his).
Lane 1, 1 hr induction; lane 2, 2 hour induction; lane 3, 4 hour induction; lane 4, pre-induction.

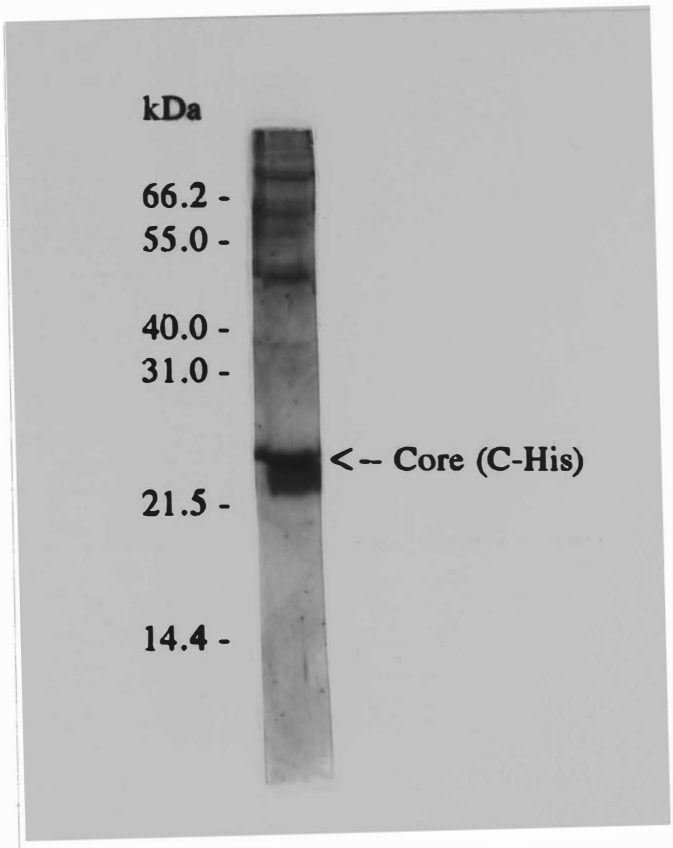


Figure 18 : SDS-PAGE of Core (C-His) Purified by Affinity Chromatography.

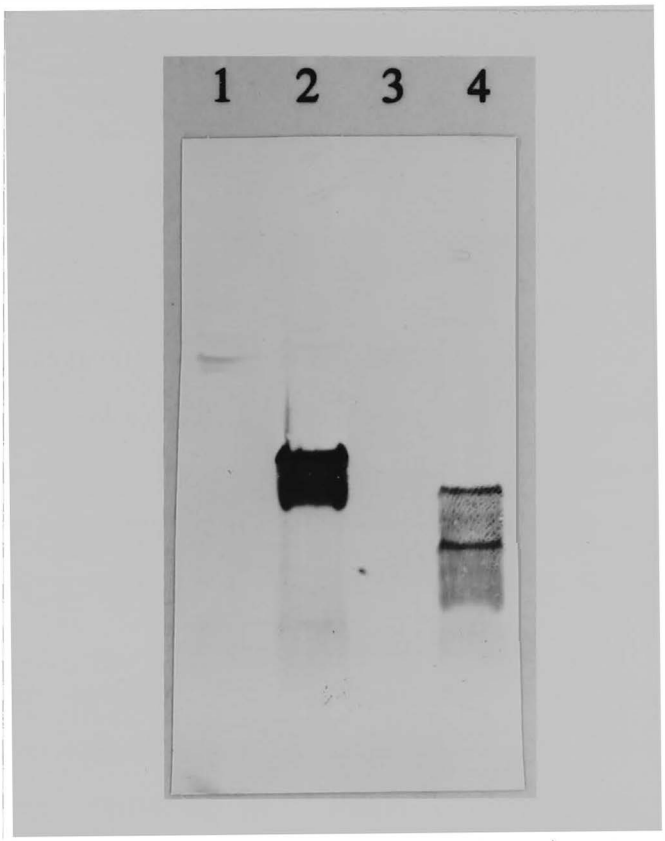


Figure 19 : Immunoblot to show the Recognition of Core(C-His) and Native Core By Polyclonal Anti-Core(C-His) Antisera. Lanes 1 & 2, pET3d-HCVC(C-His) in HMS174(DE3); lanes 3 & 4, pETHCVC in HMS174(DE3). Whole *E.coli* cell lysates , either preinduced (lanes 1 & 3), or following induction with IPTG (lanes 2 & 4), were used.

DISCUSSION

The hepatitis C virus has recently emerged as a health concern of worldwide significance. In the past 5 years over 90% of the cases of non-A, non-B hepatitis have been shown to be attributable to this virus. The majority of individuals who become infected with this virus develop chronic infections, which have been shown to be associated with the development of primary liver cancer (Colombo *et al.*, 1989; Saito *et al.*, 1990; Nishioka *et al.*, 1991; Ohkoshi *et al.*, 1990).

At the present time there is not a vaccine which specifically targets the hepatitis C virus, although trials are underway (Choo *et al.*, 1994). Therefore, there is a great need for rapidly and reliably diagnosing HCV infected individuals, both to reduce transmission of the virus, and to enable infected individuals to begin therapy. Treatment with alpha-interferon, particularly when administered during the acute stage of infection, has been shown to be successful in treating some patients (Davis *et al.*, 1989; DiBisceglie *et al.*, 1989).

The capsid, or core protein, which is encoded by the amino-terminal region of the viral polyprotein, has emerged as

one of the most immunogenic proteins encoded by the virus (Hsu *et al.*, 1993; Harada *et al.*, 1991; Chiba *et al.*, 1991). Accordingly, core fusion proteins are an important component of diagnostic ELISA's and RIBA's. The focus of this study was to develop an expression system for generating milligram quantities of full-length core protein. It was envisaged that this would permit the structural, functional and immunological characteristics of this protein to be investigated, as well as providing an antigen for capturing antibodies in a diagnostic ELISA.

The hepatitis C virus core gene was chemically synthesized using overlapping oligonucleotides (Rossi *et al.*, 1982; Scapulla *et al.*, 1982), and the synthetic gene was successfully expressed using a T7 polymerase based expression system. During the course of this study a new method for the chemical synthesis of genes was reported (Khudyakov *et al.*, 1993). Coincidentally, the first gene synthesized using this procedure, the 'Exchangeable Template Reaction', was also the hepatitis C core gene. In this reaction, pairs of oligonucleotides are annealed, filled-in with polymerase, and then digested at their 3' terminus with a restriction enzyme that creates a 3' overhang. This overhang allows the annealing of the next oligonucleotide. Following a second filling-in step, the product is digested again to allow the annealing of the next primer. In the original report the restriction enzyme 'BstX1' was used at each step. The product of this

digestion is a 4bp overhang, which was apparently sufficient to permit the annealing of the next primer. One clear shortfall in this procedure is it's inability to facilitate the exchange of regions of the gene for future studies. This may have important implications in the study of different core subtypes, which will merely require the replacement of one of the gene fragments using our construct.

Despite being expressed in a soluble form at levels of 5-10mg/L, HCV core protein proved extremely difficult to purify. This may be in part due to the apparent heterogeneity of the protein which is evident in immunoblots of total cell lysates (Figure 10). Despite the fact that core proved difficult to purify, the protein was successfully purified in sufficient quantities to begin structural and immunological characterization of it.

Probably the most interesting finding of this study is that hepatitis C core protein appears to assemble into a multimeric, disulfide-linked, particulate structure. To our knowledge, this is the first time that spontaneous assembly of hepatitis C core protein has been reported using protein acquired from a recombinant source. The particles appear somewhat heterogeneous when examined by both gel filtration chromatography (Figure 13), and electron microscopy (Figure 14), exhibiting an apparent diameter of 12-20nm. Accordingly, it is difficult to ascertain whether the core protein is truely assembled into core particles, or is subject to non-

specific aggregation.

A single report published in 1992 described the isolation and visualization of 33-nm nucleocapsid structures, associated with viral-like particles with a diameter of 55-nm, from hepatitis C infected serum (Takahashi *et al.*, 1992). This is somewhat larger than earlier reports of viral-like particles of 39-46nm derived from tissue culture (Jacob *et al.*, 1990). Further, the core protein associated with the 33-nm nucleocapsid particles isolated from serum was estimated by SDS-PAGE to have a molecular size of 26kDa. This is significantly larger than the 22kD species observed after expression of the viral genome in a variety of Eukaryotic expression systems, including baculovirus (Hsu *et al.*, 1993), monkey cos cells (Harada *et al.*, 1991) and *in vitro* translation (Hijikata *et al.*, 1991). The reason for this apparent discrepancy has yet to be resolved.

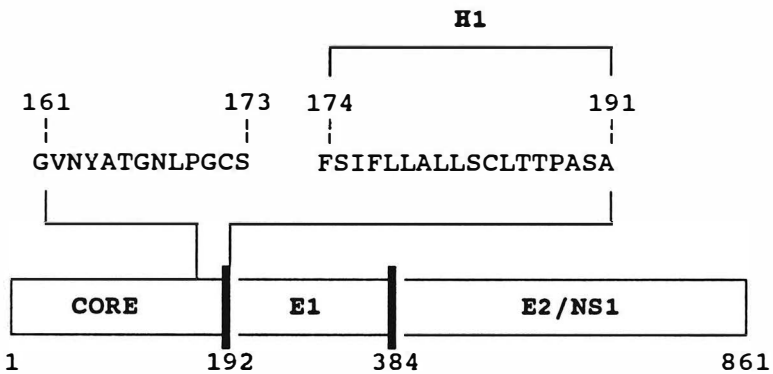
One possible explanation for the apparent instability and heterogeneity of core protein expressed in *E.coli* could be that the carboxyl-terminus of core has not been accurately determined *in vivo*. It has previously been reported that core arises after cleavage of the core-E1-E2 precursor by a host signal peptidase in the endoplasmic reticulum, and that the amino-terminus of E1 starts at amino acid 192, and E2 starts at 384. Interestingly, both of these 'cleavage sites' are preceded by two hydrophobic sequences (174-191 and 371-383), that may act as signal sequences to direct the insertion of

the viral glycoproteins into the membrane of the ER (Hijikata *et al.*, 1991).

A recent report by Santolini and coworkers (Santolini *et al.*, 1994) suggests that the hydrophobic region at the carboxyl-terminus of core (H1), and predicted to be the membrane insertion sequence of E1 (Hijikata *et al.*, 1991), may in fact represent the membrane insertion sequence of core (Figure 20). It was demonstrated using an *in vitro* translation system that H1 was cleaved from core, but only in the presence of microsomal membranes and the signal recognition particle. This would suggest that mature core protein is 173 and not 191 amino acids long. Further, this 173 amino acid protein, which was processed *in vitro,* was shown to co-migrate with core protein expressed in cell culture when examined by SDS-PAGE (Santolini *et al.*, 1994).

If this is indeed the case, and mature core protein is shorter than initially anticipated, then it raises many interesting points regarding the expression of core protein in *E.coli*. It is possible that the presence of the hydrophobic signal sequence may destabilize core protein, and/or render it more susceptible to intracellular proteases. Further, it may function to prevent correct folding of the protein, or particle assembly, until the protein has been successfully exported to the endoplasmic reticulum and the signal sequence removed. Alternatively, the protein may fold correctly but the putative signal sequence may be exposed at the surface of

Figure 20: Schematic Representation of the Signal Sequence (H1) Involved in the Intracellular Targeting of Core and E1. Adapted from Santolini *et al.*, 1994.


H1
161
173
174
191
GVNYATGNLPGCS
FSIFLLALLSCLTTPASA
CORE
E1
E2/NS1
1
192
384
861

the protein, and therefore be readily accessible to intracellular proteases. This would be in accordance with our observation that the carboxyl-terminus of core is susceptible to proteolysis during expression in *E.coli*.

Whatever the effect of the carboxyl-terminal signal sequence on the stability and folding of core it will be simple to modify our existing *E.coli* expression plasmid such that it directs the expression of core protein of 173 amino acids. Truncation of our clone, such that it directs the expression of a 173 amino acid species, will result in the removal of 11 amino acids from the carboxyl-terminus (FSIFLLALLSC). This sequence is clearly very hydrophobic in nature, and it will be interesting to ascertain if this clone synthesizes a more stable, homogeneous, and therefore more easily purifyable form of the hepatitis C core protein.

It has recently been reported that different HCV isolates have the potential to direct the expression of different forms of core protein (Lo *et al.*, 1994). A lysine residue at position 9 (HCV-1) leads to the synthesis of p16, while an arginine at this position (HCV-RH) leads to the synthesis of p21. It remains unclear as to how this conservative substitution exerts such a dramatic effect on the size of the processed protein. However, this observation does highlight the fact that mechanisms governing the synthesis of core protein may be more complicated than originally thought. The variant constructed in this study has an arginine at position

9, so therefore would be predicted to be 22kDa in size. It will be interesting to determine if the substitution of this amino acid to lysine affects core particle formation. Our *E.coli* expression system should allow investigation of this.

The ultra-violet spectroscopy profile of core suggests that HCV core protein is able to bind nucleic acid (Figure 12). It is likely that the nucleic acid binds to the basic regions of the protein, effectively neutralizing its extreme basicity (pI=11.94). This may also help to explain why core protein appears to behave as if it has a much more acidic isoelectric point than predicted by it's amino acid sequence, adhering to an anion exchange column at neutral pH (data not shown). It is widely accepted that the affinity between core protein and nucleic acid is involved in the formation of the nucleoprotein complex during assembly of the virus.

In addition to it's involvement in nucleoprotein complex formation, a recent report has implicated HCV core protein in regulating gene expression. Specifically, it has been shown that the amino-terminus of core (1-122), which is also the most basic, binds to HBV RNA and inhibits the assembly of HBV core particles (Shih *et al.*, 1993). The significance of this observation is still unresolved, but it has been speculated that HCV core protein may play a regulatory role in the gene expression of infected cells (Santolini *et al.*,1994). It will require more experimental evidence before this proposal gains wide acceptance.

The association of viral core proteins with nucleic acid following expression in *E.coli* has been observed in our laboratory, as well as several others, with the hepatitis B virus core protein. The carboxyl-terminus of this protein contains 16 arginine residues out of a total of 30 amino acids (53% arginine). When expressed in *E.coli*, core protein lacking this 'arginine-rich tail' loses its ability to bind nucleic acid, suggesting that the nucleic acid binds exclusively to this region of the protein. In hepatitis C core protein the basic amino acids are not clustered toward the carboxyl-terminus in the manner that they are in hepatitis B core protein, instead they are scattered throughout the protein. Overall the protein contains 16.8% basic residues and they are generally clustered toward the amino terminal two-thirds of the protein. Within this region, 22 (18.2%) of the residues are arginine, and 7 (5.8%) are lysine, meaning that in combination 24% (29/121) of the amino acids in this region are basic. The yellow fever virus, which is the prototype member of the *Flaviviridae*, also has an extremely basic core protein. In this protein 27% of the amino acids are lysine or arginine, highlighting the conserved nature of this basicity in viral core proteins.

Previous reports have shown that the detection of antibodies directed at the hepatitis C core protein are an important marker for detecting HCV infection (Puoti *et al.*, 1992; Van der Poel *et al.*, 1991). In addition to appearing

early in infection, the conserved nature of the protein means that subtype variability between different viral isolates is not a major concern when addressing core proteins. Only two major core subtypes have been described to date (Figure 2).

In this study we expressed core protein from subtype 1; however, the construction of the other core gene, subtype 2, will be simple to achieve using our chemically synthesized gene. The region of variability between the two core subtypes is located on fragment 2 of our chemically synthesized gene. Simple replacement of this fragment with a fragment bearing the amino acid changes that specify subtype 2 will be simple to achieve (Figure 4). A detailed comparison of the properties of the two core subtypes will assist in the determination of the role, if any, of this subtypic variability in the core protein.

Two anti-core monoclonal antibodies were shown to have a high affinity for native core protein expressed in *E.coli* (Figure 15). This indicates that the amino acid sequences to which the antibodies were raised, are exposed at the surface of core protein. Further, the fact that the monoclonal antibodies were raised against synthetic peptides suggests that the epitopes that they recognize are not conformationally-dependent. These findings highlight a possible use for these antibodies in immunohistochemistry. Studies are currently underway, in collaboration with the Department of Pathology at the 'Medical College of Virginia',

to study the cellular distribution of core protein in liver tissue from patients infected with the hepatitis C virus.

It has previously been shown that a strong T-cell response to core protein is correlated with a benign course of HCV infection (Botarelli *et al.*, 1993). While 73% of healthy seropositive individuals responded to core protein, only 11.5% of patients with chronic infections were responsive at the T-cell level. This indicates that the patients who develop chronic infections are those whose immune systems fail to mount an adequate T-cell response to core protein. Despite the recent identification of two HLA class-1 restricted epitopes on HCV core protein (Kozeil *et al.*, 1993), it remains to be determined whether these sites are directly involved in the differentiation between acute and chronic infections. The fact that class-1 restricted epitopes are the target of cytotoxic T-lymphocytes, which destroy virally infected cells, means that it is likely that these sites play a role in viral clearance during acute hepatitis C infection.

T-cell responses are mediated by the presentation of fragments of antigen in association with 'self' MHC molecules. The fact that antigen is presented in the form of linear peptide fragments means that linear, or denatured protein, is perfectly suitable for carrying out such studies. For this reason, we believe the purified histidine-tagged core protein, which can be reproducibly purified in milligram quantities, will be a useful tool for investigating T-cell responses to

the full-length core protein. These studies are currently underway in collaboration with Dr. David Milich at 'Scripps Research Institute.' Previous studies of the T-cell sites on core have been conducted using SOD-core fusions which only include the amino terminal region of core (Botarelli *et al.*, 1993). Therefore, it is probable that other undefined T-cell sites are present at the carboxyl-terminus of the protein. Characterization of this interaction between the host immune system and HCV core protein at the T-cell level will have an important role in the development of immunomodulatory agents and in vaccine development.

In contrast to T-cell responses, B-cell response are observed in almost all chronically infected individuals (Harada *et al.*, 1991; Chiba *et al.*, 1991; Hosein *et al.*, 1991; Muraiso *et al.*, Nasoff *et al.*, 1991). Since a little over 10% of chronically infected patients develop T-cell responses, antibody responses to core must be occurring in an apparent absence of a detectable T_h-cell response. This mechanism of 'T-cell help' has previously been reported in hepatitis B virus infection and involves a process termed 'intermolecular help' (Milich *et al.*, 1987). In short, T-cell responses to other viral proteins are able to provide the necessary T-cell help to B-cells and enable them to be stimulated to secrete antibody. It is possible that a similar mechanism occurs in HCV to allow universal B-cell responses in the apparent absence of T-cell help.

The most important epitopes for detecting hepatitis C infection at the present time are localized within the core and NS3 proteins. Accordingly, a European group has recently reported the characterization of core-NS3 fusion protein expressed and purified from *E.coli* (Osbourne *et al.*, 1993). The chimera contains the amino-terminal 105 amino acids of core, the region which is widely accepted as being the most immunogenic, fused to residues 1186-1364 of NS3. The protein has been shown to be effective in differentiating between hepatitis C infected individuals and healthy donors. However, while this approach is useful for detecting infected serum samples, it has a limited value in studying the individual serology of patients.

Recent observations by Feinstone and co-workers suggest that B-cell epitopes are not limited to the amino-terminal region of core. This highlights the necessity for incorporating full-length core protein for capturing anti-core antibodies in diagnostic assays. The newly described B-cell sites are located at amino acids 133-142 and 165-174 (Akatsuka *et al.*, 1993); sequences which are not included in recently developed antigens designed for capturing anti-core antibody (Osbourne *et al.*, 1993; Chen *et al.*, 1992). Likewise, the appreciation that core protein probably harbors B-cell sites which are conformationally dependent also suggests that full length, native core protein will have many benefits over core fusions and synthetic peptides for

diagnosing HCV infection. It has recently been shown that in patients infected with HIV, over 50% of the B-cell epitopes on p24, the HIV core protein, are conformationally dependent determinants (Haist *et al.*, 1993). This signifies the importance of an ability to mimic the structure of the native viral antigens when developing diagnostic assays, and also when studying the interaction between viral proteins and the immune system of the host. In this regard, our soluble *E.coli* derived core protein will serve as a useful tool for studying conformational epitopes.

The mapping of B-cell determinants on core provides support for the model of Santolini and coworkers (1994) regarding the carboxyl-terminal processing of core. So far no B-cell sites have been mapped to the amino acids 174-191, the region of core which is proposed to be cleaved during core protein maturation. This is despite the fact that the terminal regions of proteins are often surface exposed, suggesting that this sequence may indeed be cleaved off prior to nucleocapsid assembly. The recent identification of an epitope within amino acids 165-174 (Akatsuka *et al.*, 1993) suggests that amino acid 173 could indeed represent the carboxyl-terminus of the mature protein. However, it should also be noted that the amino acids encompassing 174-191 are generally hydrophobic in nature. Therefore, it is conceivable that this sequence may be present on mature core protein, but due to it's low immunogenicity is not the target of an antibody response.

Our preliminary ELISA results suggest that core protein expressed and purified from *E.coli* will be an excellent source of recombinant protein for a number of different applications. The ability of core protein to bind antibodies from HCV infected sera make it particularly useful for use in a diagnostic assay. We are currently working on improving the purification scheme for this protein so that we can purify large quantities of protein (>100mg). Having larger amounts of protein at our disposal will also facilitate the continuation of structural and immunological studies.

This task of purifying larger quantities of core protein is being approached in two different ways. The first approach involves the expression of core as a carboxyl-terminal histidine tagged fusion. This protein can be purified rapidly, and in a reproducible manner using nickel chelation chromatography. We have successfully expressed and purified this protein, under denaturing conditions, and are now in the process of developing a procedure for successfully refolding this protein such that it is soluble in the absence of denaturing agents. Osbourne and coworkers have successfully refolded an NS3-core fusion protein, for use as a diagnostic agent, using the method of Cleland (Cleland *et al.*, 1992). Accordingly, it is reasonable to assume that this procedure will serve as a useful starting point from which to begin attempts to refold full-length core protein.

The second approach for generating more core protein is

to remain with the original expression system (pET3d-HCVC), but to develop an efficient ELISA for monitoring the purification of the protein. It is envisaged that a rapid assay for the protein could limit the degree of proteolysis of core by *E.coli* proteases. The original purifications were achieved primarily using SDS-PAGE gels and immunoblotting, both of which are very time consuming procedures. We have generated polyclonal antisera to the histidine tagged core protein and shown that this antisera specifically recognizes both histidine-tagged, as well as untagged core protein (Figure 19). We are currently investigating the possibility of using the purified IgG to develop a sandwich ELISA for detecting core protein. The availability of a rapid and sensitive ELISA will be valuable asset for developing improved purification procedures for core in the future.

The expression and purification of full-length HCV proteins not only allows one to study the structure/function aspects of individual viral proteins, and develop diagnostic ELISA's, but also allows the dissection of the immune response as it relates to individual proteins. This is already possible when considering hepatitis B infection, where the detection of specific antigens (e.g. 'e' , core, and surface antigen) as well as antibodies to specific viral antigens (e.g. anti-'e', anti-core, and anti-surface antigen) allows a more detailed and therapeutically useful serology to be generated for each individual (Hoofnagle and Di Bisceglie, 1991). In hepatitis B

it is now possible to differentiate between chronic and acute infections by detailing the immune response of the patient to individual proteins.

At the present time there is little evidence for a correlation between antibody production to certain proteins and the status of the hepatitis C infection in individual patients. However, a transient IgM response to core has been observed, and proposed as a potential marker of acute or active HCV infection (Chen *et al.*, 1992). The development of immunoassays which specifically identify antibodies to individual proteins, such as core, will likely form the basis of future diagnostic assays.

CONCLUSIONS

Hepatitis C core protein, over-expressed and purified from *E.coli*, appears to be a useful tool for understanding the basic biochemistry of this virus, and also has potential applications in the medical field. Biochemically, recombinant core protein appears to exhibit several of the features that would be predicted for a viral core protein which is involved in nucleocapsid assembly. These include an affinity for nucleic acid, and spontaneous assembly into a multimeric, particulate structure. Accordingly, this recombinant protein would appear to be a useful tool for studying the structure and function of hepatitis C core protein.

Preliminary results suggest that this antigen is extremely effective in capturing anti-core antibodies from HCV infected sera. This is likely to be in part due to the ability of recombinant core protein to bind conformationally dependent antibodies in addition to those recognizing linear determinants. It is expected that this observation will ultimately be harnessed in the development of a marketable diagnostic ELISA. The detection of anti-core antibodies will have important implications, not only for reliably detecting

HCV infections, but also for studying the immune responsiveness of patients to core protein. Current HCV ELISA's incorporate epitopes from several different viral proteins, and are therefore unable to provide information regarding antibody responses to individual proteins. An ability to more precisely categorize the immune status of individual patients may help to uncover as yet hidden information regarding the different clinical courses of HCV infection.

II. CLONING, EXPRESSION, PURIFICATION AND CHARACTERIZATION OF THE MAJOR CORE PROTEIN (P26) FROM EQUINE INFECTIOUS ANAEMIA VIRUS

INTRODUCTION

THE LENTIVIRIDAE

The *Lentiviridae* are a family of animal viruses which are a subfamily of the *Retroviridae* (Charman *et al.*, 1976; Penny 1988; Stephens *et al.*, 1986). The most intensively studied member of this family of viruses has been the human immunodeficiency virus (HIV), which was classified as a lentivirus in the mid-1980's (Chiu *et al.*, 1985). Other members of this family of viruses are the equine infectious anaemia virus (EIAV/horses), caprine arthritis-encephalitis virus (CAEV/goat), bovine immunodeficiency virus (BIV/cow), feline immunodeficiency virus (FIV/cat), visna virus (sheep), and simian immunodeficiency virus (SIV/primate). These non-oncogenic, chronic disease causing lentiviruses have identical morphology, and are characterized by genetic and/or antigenic variation during persistent infection (Hahn *et al.*, 1986; Issel *et al.*, 1986; Montelaro *et al.*, 1986; Salinovich *et al.*, 1986; Hussain *et al.*, 1987; Payne *et al.*, 1987). The biologic and genetic relationship between the non-human lentiviruses and HIV-1 renders them potentially useful models for understanding the pathogenesis of HIV-1, and for

evaluating methods of effective treatment and control of viral infection (Letvin, 1990).

Equine infectious anaemia virus (EIAV), which was one of the first viruses identified in nature in 1904, is responsible for causing a chronic, debilitating disease in horses (Issel and Coggins, 1979). EIAV infection has been reported worldwide and is recognized as a livestock pathogen of significant economic importance to the horse industry. The early stages of infection are characterized by repeated, acute episodes of fever, weight loss and anaemia, with intervening periods of clinical quiescence. The progression to a chronic infection is typically manifested by arthritis, encephalitis, progressive pneumonia and slow neurological diseases (Issel and Coggins, 1979).

Genomic Organization of Lentiviruses

Members of the *Lentiviridae* share many common features, including a large RNA genome of 9 kb or more in length, and exhibit conserved genomic organization (Table 2). Like other complex retroviruses, lentiviruses exhibit a more complex genome than the other class of retroviruses - the so-called 'simple retroviruses'. Simple retroviruses, such as feline leukaemia virus (FeLV) and Rous sarcoma virus (RSV), are entirely dependent upon host cell transcription factors to regulate gene expression. In contrast, the complex

Table 2: Characteristics of Common Lentiviruses.

(Reproduced from, 'Pathogenesis of human immunodeficiency virus infection', by Jay A. Levy, Ph.D. Microbiological Reviews, March, 1993)

Clinical

Association with a disease with a long incubation period
Association with immune suppression
Involvement of hematopoietic system
Involvement of the CNS
Association with arthritis and autoimmunity

Biological

Host species specific
Exogenous and nononcogenic
Cytopathic effect in certain infected cells, e.g., syncytia (multinucleated cells)
Infection of macrophages, usually noncytopathic
Accumulation of unintegrated circular and linear forms of viral cDNA in infected cells
Latent or persistent infection in some infected cells
Morphology of virus particle by electron microscopy: cone shaped nucleoid

Molecular

Large genome (>9 kb)
Truncated *gag* gene: several processed Gag proteins
Highly glycosylated envelope protein
Polymorphism, particularly in the envelope region
Novel central open reading frame in the viral genome that separates the *pol* and *env* regions
Presence of accessory/regulatory genes

retroviruses encode several viral transactivators which enable them to exert more control over the expression of their viral genes. The synthesis of viral specific transactivators, encoded by early, regulatory mRNA's, enables complex retroviruses to adopt a replication cycle which can be divided into two different temporal phases. The 'early', or 'regulatory genes', encode a nuclear regulatory protein (called Tat in HIV), which transactivates 'long terminal repeat' (LTR) driven gene expression. This activation results in an increase in the levels of a second regulatory protein (called Rev in HIV), whose function is to activate the expression of the viral structural proteins (for review, see Levy, 1993).

Although the regulatory mechanisms of EIAV have not been studied as intensively as they have in HIV, recent reports suggest that EIAV encodes a highly homologous transcription transactivator (Dorn *et al.*, 1990), which appears to be mechanistically similar to HIV-1 Tat (Carroll *et al.*, 1991). Further, other aspects of the EIAV replication cycle, such as the identification of a functional candidate *rev* gene product (Stephens *et al.*, 1990) would appear to further confirm the evolutionary relationship between HIV-1 and the various animal lentiviruses.

The two late/structural mRNA's, which are transcribed in response to transactivation by Rev, include a singly spliced

(Env) and unspliced (Gag-Pol) mRNA. The former gives rise to the envelope glycoproteins, gp90 and gp45 in EIAV (gp120 and gp41 in HIV-1). Translation of the latter mRNA (Gag-Pol) generates two precursor polyproteins, also called Gag and Pol. Ribosomal frameshifting during the translation of the 3′ end of *gag* (Figure 21) accounts for the 20-fold higher relative abundance of the Gag precursor protein over Pol. Both precursors are subsequently processed by the viral protease to yield several functional proteins. The Pol precursor, which is 82kDa in EIAV, is cleaved into the following products: reverse transcriptase (RT); protease (PR); and the integrase (IN), which is involved in the integration of the virus into the genome of the host to form a provirus.

Lentiviral Core Proteins

Structure and Function

The *gag* (or core) open reading frame is located at the 5′ end of the genome, immediately following the 5′ long - terminal repeat (LTR) (Figure 21), and is the first gene to be translated. Consistent with other lentiviruses, four different EIAV Gag/core proteins have been identified; p26 (major structural protein), p15 (matrix protein), p11 (nucleic acid binding protein) and p9 (unknown function)(Henderson *et al.*, 1987). Previous studies have shown that all four core proteins are derived from the processing of a single 55kDa precursor (Montelaro *et al.*, 1982; Hussain *et al.*, 1988). The Gag

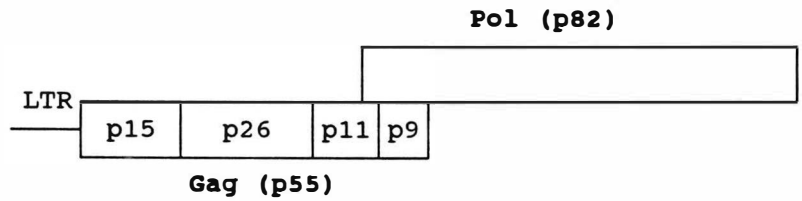


Figure 21: Schematic Representation of the Gag and Pol Regions of the Lentiviral Genome. LTR, long-terminal repeat. The molecular weights of the precursor proteins are shown in parenthesis, and the sizes of the four individual Gag proteins refer to the equine infectious anemia virus.

precursor polyprotein appears to be cleaved by a 10kDa protease, which is derived from the Pol precursor, to generate the four major non-glycosylated internal proteins of EIAV.

Retroviral Gag proteins retain the ability to direct the formation of virion-like particles, even when expressed in the absence of all other virus encoded proteins (Shiota *et* al., 1990; Delchambre *et al.*, 1989; McGuire *et al.*, 1994; Wills, 1989). Further, mutations in the Gag proteins have been demonstrated to cause retroviruses to lose their ability to bud from the plasma membrane of the host cell. Hence they are frequently referred to as the 'particle making machinery' of the virus.

A recent report concerning the processing of the 55kDa EIAV Gag precursor in a recombinant vaccinia virus system demonstrated that sub-viral core particles could be produced when recombinant vaccinia contained only p55. However, in the presence of the viral protease, the sub-viral particles, although morphologically indistinguishable, contained processed p26 protein (McGuire *et al.*, 1994). An inability to detect processed p26 in the absence of the viral protease indicates that the major structural protein (p26) is cleaved from the Gag precursor by the viral protease (McGuire *et al.*, 1994). The capability of subviral particles to form in the absence of proteolytic processing of the Gag precursor has led to the speculation that proteolytic cleavage of the Gag precursor may occur after particle assembly. This is supported

by similar observations which have been made for HIV-1 (Gowda *et al.*, 1989; Karacostas *et al.*, 1989).

The matrix protein, which is 15 kDa in EIAV (17 kDa in HIV-1), is located at the amino terminus of the Gag precursor (Figure 21) and ultimately forms an intimate association with the cytoplasmic face of the plasma membrane. Although this protein is not directly associated with other Gag proteins in mature virions, it is believed to play a role in directing the Gag precursor protein to the site of assembly on the plasma membrane. In HIV-1 the matrix protein has been shown to be essential for infectious particle formation (Bryant and Ratner, 1990; Gottlinger *et al.*, 1989)

The nucleocapsid protein of EIAV (p11), like that of all retroviruses, contains 'Cys-His' boxes ($C-X_2-C-X_4-H-X_4-C$) which appear to play a crucial role in both the packaging and dimerization of the viral RNA genome, but not in particle formation (Meric *et al.*, 1986; Meric *et al.*, 1988; Dupraz *et al.*, 1990; Gorelick *et al.*, 1990).

Subviral Particle Formation

In HIV, myristoylation of the p55-Gag precursor has been demonstrated to be essential for virion infectivity (Bryant and Ratner, 1990; Gottlinger *et al.*, 1989). The primary function of the myristic acid appears to be in targeting p55-Gag to the plasma membrane. This has lead several groups to suggest that membrane association may involve a specific

receptor located on the cytoplasmic face of the plasma membrane, analogous to p60v-src of Rous sarcoma virus (Resh, 1989; Wills and Craven, 1991). However, EIAV, like several retroviruses, is able to replicate without the need for myristoylation (Henderson *et al.*, 1987), indicating that other recognition mechanisms probably exist.

Recently it has been reported that in addition to the myristic acid moiety, a domain within the matrix (MA) region of HIV-1, downstream of the myristoylation site, is also required for a stable association between p55-Gag and the plasma membrane (Spearman *et al.*, 1994). Once in contact with the plasma membrane, the p55-Gag protein associates with the envelope proteins which are transported from the endoplasmic reticulum via the secretory pathway. The subsequent assembly of virions occurs simultaneously with the budding and release of immature viral particles.

During maturation the spherical, electron dense structure seen in immature particles condenses to form a cone-shaped, eccentrically located nucleoid. The translocation of p24 from a peripheral position to a more internal position relative to the viral membrane is also evident during maturation (Gelderbloom *et al.*, 1987). This migration of the core protein (p24) has been shown to be essential for viral infectivity, suggesting that disruption of correct viral core assembly could serve as an anti-HIV therapy (Gottlinger *et al.*, 1989).

Comparison of HIV and EIAV Major Core Proteins

The major EIAV core protein shares significant homology with the major core proteins of other members of the lentivirus family of viruses, most notably a 55% amino acid similarity (30% identity) with the major core antigen of HIV-1 (Figure 22)(Stephens *et al.*, 1986). Indeed, serum from EIAV infected horses has been shown to precipitate the major core antigens of HIV-1 (Montagnier *et al.*, 1984), and FIV (Egberink *et al.*, 1990; Steinman *et al.*, 1990). This is indicative of the presence of conserved inter-species determinants among the lentiviral core proteins, prompting suggestion that the major core proteins of HIV-1, EIAV, FIV and BIV are organized into similar antigenic domains (Chong *et al.*, 1991; Egberink *et al.*, 1990). The major antigenic sites for these proteins appear to be localized within their carboxyl-termini (Windheuser *et al.*, 1989).

Structural similarities between the major core proteins of HIV-1 and EIAV have been suggested as a result of a previous study conducted using a series of p26 fusion proteins. It was demonstrated that a 15 amino acid region of p26 (153-167), 10 of which are identical to the corresponding region in p24 from HIV-1, defines an epitope which binds a particular monoclonal antibody which also bind to the core proteins of HIV-1, HIV-2 and SIV (Nishimura *et al.*, 1984). This sequence forms part of a 20 residue sequence termed the major homology region (MHR)(Figure 22) (Patarca *et al.*, 1985).

Figure 22: Alignment of EIAV (above) and HIV-1 (below) Core Protein Sequences.
The major homology region (MHR)(amino acids 277 to 296) is indicated. (Gap Weight = 3.000; Average Match = 0.540; Length Weight = 0.100; Average Mismatch = -0.396; Quality = 151.1; Length = 239; Ratio = 0.654; Gaps = 4; Percent Similarity = 54.825; Percent Identity = 29.386).
(KEY: A bar (¦) indicates identity; a colon (:) indicates conservation; a period (.) indicates weak conservation. Cysteine (*) and tryptophan (+) residues of EIAV core are indicated).

```
                           +                        *
  1 PIMVDGAGNRNFRPLTPRGYTTWVNTIQTNGLLNEASQNLFGILSVDCTS 50
    ||: : .|..  .|:.||......||..::..::  |. . :|: |. :.|.
  1 PIVQNLQGQMVHQPISPRTLNAWVKVVEEKAFSPEVIP.MFSALAEGATP 49

                                   +
 51 EEMNAFLDVVPGQAGQKQILLDAIDKIADDWDNRHPLPNAPLVAPPQGPI 100
    :::|.:|:.|.|:.:..|:| :.|:. |.:||. ||:..:| :||.| .
 50 QDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGP.IAPGQMRE 98

                                       +
101 PMTARF...IRGLGVPRERQMEPAFDQFRQTYRQWIIEAMSEGIKVMIGK 147
    |... :   ...|. . ::  : :  .. :.|:.||| ::.. :: |.:.
 99 PRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVR.MYSP 147

                 MHR
        ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬
148 PKAQNIRQGAKEPYPEFVDRLLSQIKSEGHPQEISKFLTDTLTIQNANEE 197
    ..  :||||:|||:.::|||:.. ::.|. .|::..::|:|| :||||.:
148 TSILDIRQGPKEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNANPD 197

    *                   *
198 CRNAMRHLRPEDTLEEKMYACRDIGTTKQKMMLLAKAL 235
    |:. :: | |..||||.| ||.::|.. :| .:|
198 CKTILKALGPAATLEEMMTACQGVGGPGHKARVL.... 231
```

This region, in addition to harboring a conserved B-cell epitope, facilitates protein interactions that result in self-assembly of the Gag precursor (Ehrlich *et al.*, 1992; Rose *et al.*, 1992; Strambio-De-Castillia and Hunter, 1992). The importance of the MHR in capsid assembly is highlighted by the observation that it is conserved among all retrovirus core proteins (Patarca *et al.*, 1985; Wills and Craven, 1991).

Close examination of the amino acid sequences of EIAV and HIV core proteins reveals several intriguing homologies. The first is a conserved pair of cysteine residues; specifically Cys198 and Cys218 of EIAV p26 with two cysteine residues in HIV-1 p24 (Figure 22). This raises the possibility of the presence of a conserved disulfide bridge, which, if present, could perform an important structural role in each of these proteins. The conservation of such a structural element would highlight an apparent necessity for the retention of this bond, suggesting it performs an important function in the formation of subviral particles.

A second interesting homology exhibited by these two proteins involves tryptophan residues. p26 from EIAV has three tryptophan residues (Trp23, Trp81 and Trp132), each of which is conserved in the primary sequence of p24 from HIV (Figure 22). The intrinsic ability of tryptophan residues to fluoresce renders them useful for probing the structure of proteins. This conservation of tryptophan residues enables their fluorescence properties to be exploited in comparing the

structure of these two proteins.

In light of the high degree of homology between the major core proteins of members of this family of viruses, we propose that structural information for p26 of EIAV may be applicable to other members of the *Lentiviridae*. It is particularly relevent to the major core proteins of FIV, HIV-1, and BIV, which collectively form a subfamily of the *Lentiviridae* exhibiting serological cross-reactivity of their major core proteins (Egberink *et al.*, 1990; Goudsmit *et al.*, 1986). A working knowledge of structural features, and ultimately the 3-dimensional structure of capsid proteins provides important structural information which helps to develop an understanding of how the structure of these proteins is involved in their functionality. This is of particular interest with this family of proteins as they have been shown to be essential for viral infectivity.

The high degree of amino acid similarity, combined with immunological cross reactivities, suggests that the 3-dimensional structure of the lentiviral core proteins are likely to be related to some degree. At the present time, the structural information for lentiviral core proteins is limited to models assigned to HIV-1 core protein (p24). The models have been generated largely based on data generated from the mapping of antigenic determinants using monoclonal antibodies, in concert with computer-assisted structure predictions (Argos, 1989; Langedijk *et al.*, 1990). The

computer generated model, proposed by Argos, indicates that p24 of HIV-1 is folded in the form of an anti-parallel beta-barrel with a single stretch of alpha-helix. Langedijk and colleagues later assigned linear antibody binding sites to every turn and loop of the beta-barrel, and to the alpha-helix. However, several antibody binding sites were also assigned to beta-sheet regions which one would not predict to be surface exposed, suggesting that the model may not be entirely accurate. These inconsistencies highlight the necessity for structural determinations of lentiviral core proteins which are based on experimental evidence rather than computer generated models.

Diagnosing EIAV Infection

The 'Coggins' test, which is an agar gel immunodiffusion assay, has been the accepted diagnostic for EIAV infection since it was developed almost 25 years ago (Coggins and Norcross, 1970). This assay detects antibodies to p26, the major core protein of EIAV, using protein which is purified from cell culture. The high relative abundance, 40% of total virion protein mass in EIAV, tied to the highly immunogenic and remarkably conserved nature of viral core proteins has made them popular choices in diagnostic assays for detecting viral infection. The use of retroviral envelope proteins for detecting viral infection is hampered by the rapid mutation rate of these proteins. Therefore, core proteins are generally

considered more reliable for capturing circulating antibodies.

In the absence of a source of recombinant full-length protein, peptides or fusion proteins which incorporate major antigenic determinants are popular choices for creating diagnostic assays. Inhibition assays using monoclonal antibodies are also widespread. However, both of these approaches have potential shortfalls. Peptides and fusion proteins often fail to mimic antigenic sites to a sufficient degree, particularly if the sites are discontinuous in nature; while inhibition assays are at the mercy of a single amino acid substitution. Such a mutation is often sufficient to destroy the monoclonal antibody binding site resulting in a failure to detect infected samples. The latter is exemplified by the recent discovery of hepatitis B virus 'escape mutants' which were not detected by a monoclonal antibody based assay directed at the surface protein of the Hepatitis B virus (Fujii *et al.*, 1992).

We propose that a source of recombinant p26 could have useful applications in the design of new diagnostic reagents for EIAV. This could possibly take the form of an ELISA, which would have many potential benefits over the rather time-consuming and insensitive immunodiffusion assay.

SPECIFIC AIMS

1. Clone, express and purify the major core protein (p26) of EIAV using an *E.coli* expression system.

2. Study the structural and physical properties of this protein and in the process evaluate it as a model for understanding lentiviral core proteins, particularly p24 from HIV-1.

3. Develop a rapid and sensitive ELISA for detecting anti-p26 antibodies in serum, and determine it's potential to serve as a marketable diagnostic assay for monitoring EIAV infection.

BACKGROUND

Cloning Strategy

The cloning of the EIAV core gene, such that it does not include extraneous amino acids at the amino-terminus, requires the 'engineering' of the sequence around the initiator methionine residue. A popular choice for overcoming this problem is the incorporation of methionine-containing restriction sites at the 5' end of the gene (e.g. NcoI, NdeI, SphI). For p26, the incorporation of an NcoI restriction site requires the alteration of the 4th nucleotide in the gene from an adenine to a guanine, this changes the second triplet from ATA (Ile) to GTA (Val) (Figure, 23a and b). However, since valine and isoleucine both have aliphatic, hydrophobic side chains, we do not expect this conservative alteration to be of significance in terms of the global structure of the protein.

Processing of the amino terminus of this protein would be expected in *E.coli* since the initiator methionine is followed by valine. The following amino acids; His, Gln, Glu, Phe, Met, Lys, Tyr, Trp and Arg, when located as the penultimate amino acid, appear to supress the removal of the

Figure 23: Cloning Strategy for p26.
Panel A - Nucleotide sequence of the primers used to amplify the p26 gene. The 5′ primer introduces an NcoI restriction site, while the 3′ primer introduces a HindII restriction site and a termination codon. **Panel B** - Alignment of the 5′ ends of the original and cloned p26 sequences. **Panel C** - Alignment of the 3 ′ ends of the original and cloned p26 sequences. Restriction sites are underlined (NcoI - CCATGG; HindIII - TTCGAA), and 'TAA' is a termination codon.

A

5′-GCCCATGGTAGATGGGGCTGGAAACAG -- 5′NcoI Primer

5′-CGAAGCTTAAAGTGCTTTTGCCAATAACAT -- 3′HindIII Primer

B

```
                    P   I   M   I   D   G   A   G
Original ... ... CCA ATC ATG ATA GAT GGG GCT GGA ... ...
Cloned   ... ...     CC ATG GTA GAT GGG GCT GGA ... ...
                    -   -   M   V   D   G   A   G
```

C

```
                    A   K   A   L   Q   T   G   L
Original ... ... GCA AAA GCA CTT CAG ACT GGT CTT
Cloned   ... ... GCA AAA GCA CTT TAA GCT TCG
                    A   K   A   L   *
```

initiator methionine (Hirel *et al.*, 1989). Since the initiator methionine would be expected to be removed on this protein, valine is a good choice as the penultimate amino acid according to the 'N-end rule' proposed by Tobias and coworkers (Tobias *et al.*, 1991). They reported protein half-lives of only 2 minutes when the following amino acids were present at the amino terminus; Arg, Lys, Phe, Leu, Trp, and Tyr.

The Gag precursor of EIAV is synthesized as a polyprotein, meaning that p26, which is flanked by p15 (matrix protein) and p11 (nucleocapsid protein) (Figure 21), does not contain a termination codon. Therefore it was necessary to insert a termination codon at the predicted 3' end of the cloned p26 gene (Stephens *et al.*, 1986). This was achieved by inserting a termination codon immediately upstream of the HindIII cloning site in the reverse PCR primer (Figure 23c).

Predicted Features of p26

The predicted length of the equine infectious anaemia major core protein is 233 amino acids (Figure 22). The core protein in this study is expected to be 231 amino acids long due to the fact that is starts at the first methionine residue, where an NcoI restriction site could be conveniently inserted (Figure 24). Therefore, our clone lacks the first two amino acid, proline and isoleucine (Figure 22). When refering to specific amino acids, numbering is done in accordance with

```
  3  MVDGAGNRNF RPLTPRGYTT WVNTIQTNGL LNEASQNLFG ILSVDCTSEE

 53  MNAFLDVVPG QAGQKQILLD AIDKIADDWD NRHPLPNAPL VAPPQGPIPM

103  TARFIRGLGV PRERQMEPAF DQFRQTYRQW IIEAMSEGIK VMIGKPKAQN

153  IRQGAKEPYP EFVDRLLSQI KSEGHPQEIS KFLTDTLTIQ NANEECRNAM

203  RHLRPEDTLE EKMYACRDIG TTKQKMMLLA KAL*
```

Figure 24: Amino acid Sequence of Cloned EIAV Core Protein.

the predicted full-length protein, the sequence of which is displayed in figure 22. Thus, the initiator methionine of the p26 clone is in fact residue 3.

Successful expression of the gene constructed in this study is predicted to yield a protein with a molecular weight of 26kDa. There are no predicted sites of glycosylation on this protein. It has an isoelectric point of 6.78, which means that the protein would be expected to carry zero net charge at this pH. The extinction coefficient for this protein is calculated to be 22,370 $M^{-1}cm^{-1}$.

MATERIALS AND METHODS

Cloning of the EIAV Core Gene

Isolation of Viral RNA from EIAV Infected Serum

Total RNA was prepared from EIAV infected serum using the procedure described by Simmonds (Simmonds *et al.*, 1990), with minor alterations, including those described by Castillo (Castillo *et al.*, 1992). To 200μl of serum, 500μl of a solution containing 4.2M guanidine thiocyanate, 0.5% sarkosyl, 25mM sodium citrate, pH 7.0, and 5μl of 2-mercaptoethanol were added. The solution was mixed and 50μl 2M sodium acetate (pH4.0), 500μl phenol, and 200μl chloroform were also added, the samples were mixed and incubated for 15 minutes on ice. After centrifugation at 14,000xg for 15 minutes at 4°C, the aqueous phase was twice re-extracted with chloroform. The viral RNA was precipitated with isopropanol at -70°C overnight. Next morning the RNA was pelleted by centrifugation at 14,000xg for 30 minutes at 4°C, washed with 70% ethanol, and dried. The pellet was resuspended in 3μl of TE buffer (10mM tris-Cl, pH 8.0, 1mM EDTA), and used as the template for the reverse transcription reaction.

Reverse Transcription and PCR Amplification

The reverse transcription reaction was prepared according to the manufacturer's directions ('GeneAmp', Perkin, Elmer & Cetus). The total reaction volume was 20μl and random hexamers (2.5μM) were used to prime the reaction. Following a 10 minute incubation at 25°C the mixture was incubated at 42°C for 20 minutes, 99°C for 5 minutes, and finally 4°C for 5 minutes. The contents of this tube were then used as the template for a standard PCR reaction. The following conditions were employed for the PCR; 5 cycles of 94°C, 42°C and 72°C, each for 1 minute; followed by 35 cycles of 94°C, 55°C and 72°C, each for 1 minute. A final 'filling-in' step of 72°C for 5 minutes was incorporated to assure that the ends of the PCR product were 'flush'. The product of the PCR reaction was run on an agarose gel (0.8%), and a 712bp product was extracted using the 'Qiaex' DNA extraction kit (Qiagen).

Restriction Digestion and Cloning

The primers used for the PCR were designed such that the product had an NcoI restriction site at the 5' end, and a stop codon followed by a HindIII restriction site at the 3' end (Figure 23). Two additional nucleotides were incorporated 5' of the restriction sites to facilitate efficient cutting of the PCR product by these enzymes.

The 712bp PCR product was digested with the restriction enzymes NcoI and HindIII and cloned into the vector pUC18N,

which had been cleaved with the same two enzymes. pUC18N is identical to pUC18 except it has an NcoI restriction site introduced at the 5′ end of the multiple cloning site (Zheng *et al.*, 1992). A positive clone was identified by restriction analysis of mini-prep DNA and ultimately by automated DNA sequencing.

To express EIAV core protein the gene was cloned into the expression vector pKK332 (Pharmacia), which uses the 'tac' promotor to drive the expression of cloned genes. The initiator methionine residue is coded for by the ATG triplet within the NcoI restriction site, ensuring that no extraneous amino acids are incorporated into the final gene product.

The NcoI/HindIII fragment bearing the core gene was excised from pUC18N/EIAp26 and ligated into the NcoI/HindIII sites of the expression vector. The host cell strains used for expression were TB1.

DNA Sequencing

The identity of the cloned DNA was confirmed by automated DNA sequencing of pUC18N/EIAp26, performed by the 'MCV-VCU Nucleic Acid Core Facility'.

Site-Directed Mutagenesis of Cysteine-48 to Serine

A two-step PCR reaction was used to mutate cysteine-46 of EIAV core protein to a serine. Two mis-match oligonucleotide primers were designed such that base number 142, a thymidine,

was mutated to an adenine. This changes codon number 48 from 'TGT' which specifies cysteine, to 'AGT' which specifies serine. This single nucleotide mutation also results in the introduction of a new and unique restriction site (ScaI). This is a necessary feature of the mutagenesis procedure (Figure 25).

Two PCR reactions were carried out using pKK322/EIAp26 as the template, and the following pairs of primers:

1. 5' NcoI primer (forward), and the internal mis-match ScaI primer (reverse).
2. 3' HindIII primer (reverse), and the internal mis-match ScaI primer (forward).

The products of the two reactions, an NcoI/ScaI fragment corresponding to the 5' end of the gene, and a ScaI/HindIII fragment corresponding to the 3' end of the gene were recovered and digested with ScaI. The two fragments were ligated to create the mutated gene, and then used as the template for a third PCR reaction. The final PCR reaction, using the original 5' NcoI and 3' HindIII primers, served merely to amplify the mutant gene to acquire sufficient DNA for successful cloning. The final product was digested with NcoI and HindIII, and then ligated into pKK322. The identity of the final clone was confirmed to be that of the Cys48Ser mutant by confirmation of the presence of the ScaI restriction site.

A

5' - AC<u>AGTACT</u>TCTGAAGAAATGAATGC -3'

5' - GA<u>AGTACT</u>GTCTACTGATAATATCCC -3'

B

			V		**D**		**C**		**T**		**S**	**E**
Original	...	...	GTA		GAC		TGT		ACT		TCT	GAA ... ...
Cys48Ser	...	...	GTA		GAC		<u>AGT</u>		<u>ACT</u>		TCT	GAA ... ...
			V		**D**		**S**		**T**		**S**	**E**

Figure 25: Mutation of Cysteine 48 to Serine.
Panel A shows the sequence of the ScaI primers used to PCR amplify the two gene fragments. ScaI restriction sites are underlined. **Panel B** shows the original and mutated gene sequences. Amino acid sequences are shown in bold, and the ScaI restriction site in the Cys48Ser mutant is underlined.

Deletion of the Hydrophobic Carboxyl-Terminus of p26

The hydrophobic carboxyl-terminus of p26 was deleted by designing a PCR primer which terminated the translation of the protein after Lys223 (Figure 26). This primer was used in combination with the original NcoI primer, in a PCR reaction where pUC18N/p26 served as the primer. The PCR product was cloned into the pKK322 expression vector.

Expression and Purification of EIAV Core Protein

Expression of EIAV core protein

E.coli (strain TB1) was transformed with the plasmid pKK/EIAp26 and plated onto an LB plate supplemented with 50μg/ml ampicillin. A single colony was selected and used to inoculate 100ml of TYN media (10g tryptone, 10g yeast extract, 5g NaCl, pH 7.4), supplemented with 1g/L glucose and 50μg/ml ampicillin. This culture was grown for 8 hours at 37°C and then used to inoculate 10L of the same media. This culture was grown for 24 hours, after which time the cells were harvested by centrifugation at 6000 rpm for 5 minutes. Cell pellets were stored at -70°C until needed.

Purification of EIAV Core Protein - Method 1

Cell pellets were resuspended in water (30ml/L of cells) and lysed by a single passage through a French pressure cell at 12,500psi. The cell lysate was clarified by centrifugation at 20,000rpm for 30 minutes, and the supernatant was brought

```
 G   T   T   K   Q   K   M   M   L   L   A   K   A   L   *
GGA ACT ACA AAA CAA AAG ATG ATG TTA TTG GCA AAA GCA CTT TAA
GGA ACT ACA AAA CAA AAG TGA TAA GCT T
 G   T   T   K   Q   K   *   *
```

Figure 26: Deletion of the Carboxyl-Terminal Hydrophobic Region of p26.
The carboxyl-terminal 8 amino acid of p26 were deleted in an attempt to increase the solubility of the protein for crystallization experiments. The sequence of the original p26 clone is shown above the truncated mutant. The HindIII restriction site (AAGCTT) is underlined. Two stop codons were included in another attempt to decrease heterogeneity at the carboxyl-terminus of the protein.

to a 50% ammonium sulfate saturation by the addition of solid ammonium sulfate (313g/L). The precipitate was recovered by centrifugation at 20,000 rpm for 10 minutes and resuspended in 10mM sodium phosphate, pH 6.8, before being dialyzed against the same buffer at 4°C, overnight. The dialyzed protein was applied to a hydroxyapatite column which had previously been equilibrated with 10mM sodium phosphate, pH 6.8. Once the protein had entered the column it was washed extensively with 50mM sodium phosphate, pH 6.8. The eluent was collected until the optical density, measured at 280nm, fell below 0.1. The eluent was subjected to a second 50% ammonium sulfate precipitation. The precipitate was recovered in the manner described previously, resuspended in 10mM tris-Cl, pH 8.0, and dialyzed against the same buffer at 4°C, overnight. The protein was loaded onto a DE-51 column (Whatman), which had been equilibrated with 10mM tris-Cl, pH 8.0. Bound proteins were eluted with a linear gradient of 0-300mM NaCl in the same buffer. Fractions were assayed for core protein, using the inhibition ELISA assay described (see below), pooled, and subjected to a 50% ammonium sulfate precipitation. The resultant precipitate was recovered by centrifugation, resuspended in a minimum volume of 10mM tris-Cl, pH 8.0, and dialyzed against the same buffer overnight. The material was then loaded onto a Biogel p-100 sizing column (2.5 x 100cm) which had been equilibrated in 10mM tris-Cl, pH 8.0. 5ml fractions were collected and assayed for core protein. Active

fractions were pooled and its purity tested by SDS-PAGE.

Purification of EIAV Core Protein - Method 2

The second purification procedure was identical to 'method 1' in it's initial stages. However, following the first 50% ammonium sulfate precipitation, the protein was dialyzed against 50mM sodium phosphate, pH 6.8, and applied to a DE-52 anion exchange column (Whatman) which had previously been equilibrated with the same buffer. Once the protein had entered the column it was washed extensively with 50mM sodium phosphate, pH 6.8. The eluent was collected, and subjected to a second 50% ammonium sulfate precipitation. The precipitate was recovered, resuspended in a minimum volume of 50mM sodium phosphate, pH 6.8, and dialyzed against the same buffer overnight, at 4°C. The solution was clarified by centrifugation (18,000xg, 20 min), and loaded onto a p-100 sizing column (Biogel). Fractions (20ml) were collected and assayed for core protein. Active fractions were pooled and purity tested by SDS-PAGE.

Physical and Chemical Characterization of EIAV Core Protein

SDS-PAGE

Analytical polyacrylamide gel electrophoresis was performed according to the method of O'Farrell (O'Farrell, 1975). 4% acrylamide stacking gel and 15% acrylamide

resolving gel were employed.

Immunoblotting

Immunoblotting was performed according to the method of Burnette (1981). Briefly, membranes were blocked with 4% BSA in PBS, and 1:50 dilutions of horse serum were made in 1% BSA in PBS. The serum was incubated with the membrane at room temperature, overnight. A 1:3,000 dilution of anti-horse IgG peroxidase (Sigma) was made in PBS and incubated with the membrane for 1 hour. The immunoblots were developed with chloronapthol/H_2O_2.

Amino-Terminal Sequence Analysis

5μg of pure p26 protein were subjected to amino terminal sequencing. Five cycles of sequence analysis were performed using an Applied Biosystems model 470A sequencer, equipped with an in-line PTH-amino acid analyzer.

Circular Dichroism Spectroscopy

Circular dichroism spectra were generated on a Jobin Yvon Mark III dichograph using a 0.1cm path-length quartz cell. The protein concentration was 0.1mg/ml. The values of mean M_r residue ellipticities were calculated on the basis of 110 as the average M_r per residue and they are reported in terms of $[\theta]_{m.r.w.}$ (degrees.cm^2.$dmol^{-1}$). The secondary structure estimations were performed by computer fit according to the ellipticity

reference values of Bolotina *et al.* (1980), which are based on proteins of known three-dimensional structure.

Fluorescence Spectroscopy

Fluorescence studies were carried out with a Perkin-Elmer MPF-44E spectrofluorimeter fitted with a 150W xenon arc and a differential correction spectra unit (DCSU-2). The temperature in the cuvette was maintained by a Lauda circulating water bath. Excitation was performed at a wavelength of 275nm, and emission spectra measured over a range of 280-450nm.

Quenching studies were performed by adding aliquots of a stock solution of 2M KI, containing 0.1mM $Na_2S_2O_3$ to prevent I_3^- formation. The data from a control experiment performed with 2M NaCl instead of KI, in order to mimic the ionic strength and protein concentration, was subtracted from the quenching data. The analysis of quenching was performed according to a modified form of the Stern-Volmer relationship:

$$F_o/(F_o-F) = 1/([Q]f_aK_a) + 1/f_a$$

where F_o is the fluorescence in the absence of quencher, F is the fluorescence at a molar quencher concentration [Q], K_a is the Stern-Volmer constant and f_a is the fraction of fluorescence accessible to the quenching agent (Lehrer, 1971). From a plot of $F_o/F_o-F)$ versus $[Q]^{-1}$, the values of f_a and K_a can be calculated. At each quencher concentration, F_o was determined from a p26 solution to which aliquots of 2M NaCl

were added.

Free-Cysteine Labelling

The free cysteine residue was labelled with ^{14}C-iodoacetamide in the presence of 6M guanidine hydrochloride. After removal of the excess ^{14}C-iodoacetamide the protein was reduced and alkylated again, this time with unlabelled iodoacetamide. The reduced and alkylated protein was digested with trypsin. Separation of the tryptic peptides was achieved by fractionation using reverse phase HPLC. An 'Ultrasphere ODS column' was used, along with a linear gradient (0-30%, 80 min). Fractions were 'counted' in a Beckman liquid scintillation counter, and those containing radioactivity were hydrolysed with 6N HCl, at 115°C for 24 hours, and analyzed using a Beckman 6300 amino acid analyzer.

Free Thiol Titration

The free sulfhydryl content of p26 was determined using Ellman's reagent (5,5'-dithiobis-2-nitrobenzoic acid, Sigma), under both native and denaturing conditions (Ellman, 1958; Riddles ***et al.***, 1979). To measure the number of surface accessible sulfhydryl groups, a solution comprised of 40μM Ellman's reagent and 100mM Na_2HPO_4 was mixed 1:1 with pure p26 protein (2mg/ml) in 100mM Na_2HPO_4. The reaction was supplemented with 0.1% SDS for the determination of 'total' free sulfhydryl groups. The reaction was allowed to proceed

at 37°C, for 10 minutes, after which time the absorbance was measured at 412nm. The free sulfhydryl concentration was calculated using the extinction coefficient of 13,600 $CM^{-1}M^{-1}$. A prior knowledge of the exact protein concentration facilitated the determination of the number of surface accessible, as well as total free, sulfhydryl groups present in the protein.

Crystallization

The ability of purified EIAV core protein to crystallize was investigated using a commercially available kit comprised of 50 different buffer/salt/precipitant mixtures ('Crystal Screen', Hampton Research), using the 'hanging drop' technique described in the accompanying directions. One milliliter of purified core protein (16mg/ml) was dialyzed against 1mM sodium phosphate overnight. Briefly, 2μl of each of the 50 crystal screen buffers was mixed with 2μl of protein, applied to a siliconized coverslip, inverted, and sealed with vacuum grease above 0.5ml of the same buffer (undiluted). The 50 samples were set-up in duplicate; one series was incubated at 4°C and the other at room temperature.

IMMUNOASSAYS

Conjugation of Core Protein to Horseradish Peroxidase

A p26-peroxidase conjugate was prepared by reacting 2mg of pure p26 protein with 1mg of activated horseradish

peroxidase in 50mM sodium phosphate buffer, pH 7.0, 50mM sodium cyanoborohydride at 37°C for 1 hour. The conjugate was stable for at least 6 months when stored at 4°C.

ELISA for Detecting EIAV Core Protein

Purified p26 protein was coated to the wells of microtitre plates (Costar) by incubation at room temperature, overnight, at a protein concentration of 10μg/ml (50μl per well) in 50mM sodium bicarbonate buffer, pH 9.2. Next morning the plates were blocked with 1% BSA for 1 hour at room temperature, dried, and stored at -20°C until needed.

EIAV positive horse serum was diluted (1:10) with normal horse serum. 50μl of this was added to each well of a microtitre plate, which had previously been coated with p26 protein, and incubated for 10 minutes. The wells were then washed and 50μl of the sample to be assayed for core protein was added to each well. After 5 minutes, 50μl of p26 peroxidase conjugate (1:1000 dilution) was added to each well. Five minutes later the wells were washed and 50μl of 'TM Blue' substrate (TSI) was added. After 10 minutes the reaction was stopped by the addition of 100μl 1N sulfuric acid and the plates read at a wavelength of 450nm. Typically, wells which had not received core protein gave an absorbance of 0.8 to 1.2 whereas those with core protein were significantly lower at 0.1 to 0.4 absorbance units. Fractions exhibiting inhibition of 50% or more were considered positive for p26 protein. All

steps were carried out at room temperature.

ELISA For Detecting EIAV Infection

To assay for anti-core antibodies, 50µl of serum was added to each well of microtitre plates, which had previously been coated with p26 protein, and incubated at room temperature for 30 minutes. The wells were then washed and 50µl of p26-peroxidase conjugate (diluted 1:1000 in normal horse serum) was added to each well. Following a 15 minute incubation the wells were washed again and 50µl of 'TM Blue' substrate was added. The reaction was allowed to proceed for 10 minutes, after which time it was stopped by the addition of 100µl 1N sulfuric acid and read at a wavelength of 450nm. All steps were performed at room temperature.

Generation of Polyclonal Rabbit Antisera

100µl of p26 protein (1mg/ml) was diluted to 500µl with PBS and mixed 1:1 with Freund's complete adjuvant (Sigma). Following brief sonication the homogenous emulsion was administered to a single New Zealand white rabbit by making multiple injection, subcutaneously, on the back of the rabbit. One month later the rabbit was 'boosted' by injecting antigen in a similar manner, except that the antigen was mixed 1:1 with Freund's incomplete adjuvant. This procedure was repeated after two weeks, and one week later the rabbit was sacrificed.

RESULTS

Cloning, Expression and Purification of EIAV Core Protein

The gene coding for the major core protein (p26) of the lentivirus EIA was successfully reverse transcribed from viral RNA which had been extracted from infected serum, amplified by PCR, and cloned. The gene, whose identity was confirmed by automated DNA sequencing, was then cloned into pKK322, an *E.coli* expression vector. The protein, expressed in a soluble form, was subsequently purified using standard biochemical procedures (Figure 27). The final preparation was determined to be >95% pure by SDS-PAGE analysis (Figure 27, Lane 6). Approximately 50-80mg of p26 protein was routinely purified from 20 liters of culture.

Chemical and Physical Characterization of EIAV Core Protein

Amino-Terminal Sequence Analysis

The identity of the purified protein was confirmed by amino-terminal sequence analysis (Table 3). The first five residues were determined to be Val, Asp, Gly, Ala, Gly, which is consistent with residues 2 through 6 of the published

sequence (Kawakami *et al.*, 1987). The endoproteolytic cleavage of the initiator formyl-methionine (fMet) residue, which is catalyzed by methionyl aminopeptidase, is frequently reported when proteins are expressed and purified from *E.coli*. Since the penultimate amino acid of p26 is valine, removal of the amino-terminal fMet would be expected according to the observations of Hirel *et al.* (1989).

Ultra-Violet Spectroscopy

When analyzed by ultra-violet spectroscopy, purified p26 exhibited absorption spectra characteristic of a soluble protein (Figure 28). It exhibited maximum absorbance at a wavelength of 280nm, suggesting that it does not have an affinity for nucleic acid.

Reducing and Non-Reducing SDS-PAGE Analysis

Pure EIAV core protein was subjected to SDS-PAGE analysis in both the presence and absence of reducing agent (5% 2-mercaptoethanol), in order to determine if it contained any intermolecular disulfide bridges. The results, displayed in figure 27, show that no intermolecular disulfide bridges are present. Under either reducing (lane 6), or non-reducing conditions (lane 7), the protein migrates as a monomer of 26kDa. This observation does not exclude the possibility that the protein has an intramolecular disulfide bond.

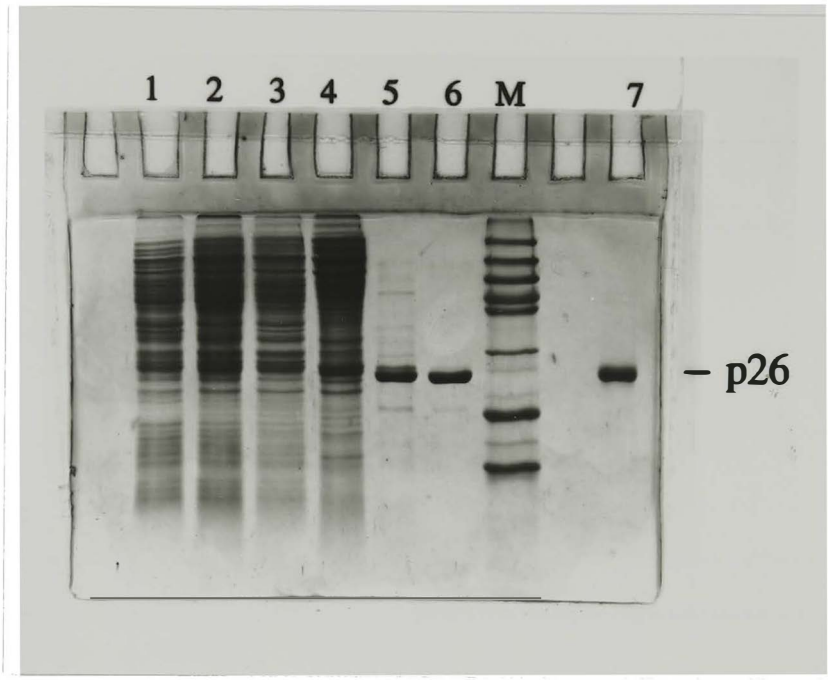


Figure 27: SDS-PAGE to Show the Steps in Purification of EIAV Core Protein. Lane 1: Cell Lysate; Lane 2: 50% ammonium sulfate precipitate; Lane 3: hydroxyapatite eluent; Lane 4: 50% ammonium sulfate pellet; Lane 5: DE-51 fractions; Lane 6: P-100 pooled fractions; Lane 7: P-100 pooled fractions (non-reducing); M = molecular weight markers, 97.4, 66.2, 55.0, 42.7, 40.0, 31.0, 21.5, 14.4 kDa.

Table 3: Amino-Terminal Sequence Data (5 cycles).
The most abundant amino acid at each cycle is indicated with bold type. Units are in picomoles.

CYCLE NUMBER

	1	2	3	4	5
Val	**15.30**	1.88	0.90	0.79	0.00
Asp	0.00	**8.67**	6.11	1.97	0.79
Gly	6.66	3.81	**25.57**	9.66	**23.51**
Ala	8.07	7.09	6.07	**33.56**	11.66

Since p26 contains 3 cysteine residues (Figure 22), it must adopt one of two possible disulfide bonding patterns: either, it must have a single intramolecular disulfide bridge along with a single free cysteine, or it must have 3 free cysteines.

Determination of the Disulfide Bonding Pattern

Using Ellman's reagent it was possible to ascertain both the number of free sulfhydryl groups, in addition to their surface accessibility. The data, shown in table 4, suggest that EIAV core protein contains a single free cysteine residue. This is evident from the fact that one mole of p26 reacts with one mole of Ellman's reagent. Further, this single reactive cysteine must be poorly accessible to solvent since it reacts only poorly when the protein is in its native conformation. However, when the protein is denatured by treatment with SDS the cysteine is fully exposed.

The fact that p26 contains only a single reactive cysteine residue, even when denatured by the presence of SDS, suggests that the other two cysteine residues must form a disulfide bridge. This accounts for their lack of reactivity with Ellman's reagent.

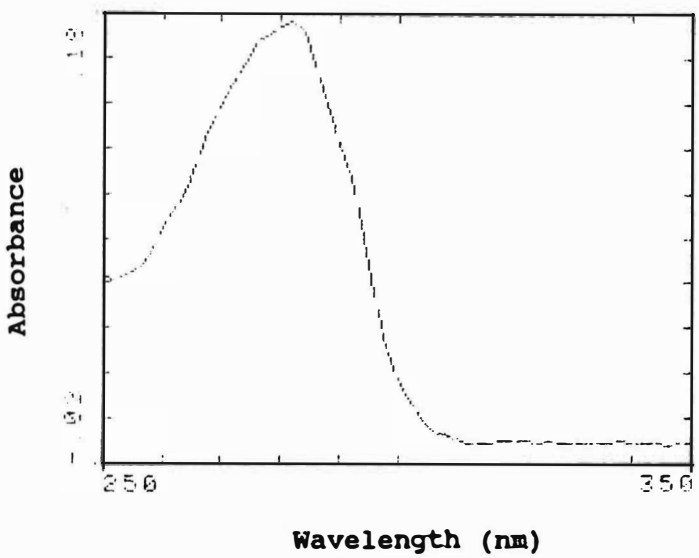


Figure 28: Ultra-Violet Absorption Spectra for p26.

In an attempt to ascertain which of the three cysteine residues was free and which two were involved in forming the disulfide bond, the protein was reacted with iodoacetic acid (^{14}C). Digestion of the labelled protein with trypsin and separation of the peptides by HPLC resulted in the identification of three fractions associated with radioactivity (Figure 29). Each of the three peptides was identified by amino acid analysis, and it was shown that the three reactive peptides contained a cysteine residue. The counts per minute per nanomole (cpm/nmol) of cysteine was determined for each peptide in order to determine the relative amounts of ^{14}C label associated with each of the three cysteine residues. The results are displayed in figure 30. Two-thirds of the radioactivity (14,100 cpm/nmol) was associated with peptide 1, which contained Cys-48, suggesting that it is the free cysteine in p26. The other cysteine containing peptides were also associated with radioactivity, although to a much lesser extent. The Cys-198 containing peptide had 5070 cpm/nmol, and the Cys-218 containing peptide had 2020 cpm/nmol. The reactivity of these residues could result from reduction of the disulfide bond during storage, or the formation of a partial disulfide bond. The latter is unlikely because newly purified protein reacted with Ellman's reagent with a stoichiometry which indicated the presence of exactly one mole of free cysteine residues per mole of protein.

Table 4: Free Thiol Determination of p26
All determinations were done in triplicate and the protein concentration was 34.41μM.

Protein	SDS	A412	[React.Cys]μM
-	-	0.079	0
+	-	0.164	6.25
-	+	0.035	0
+	+	0.517	35.44

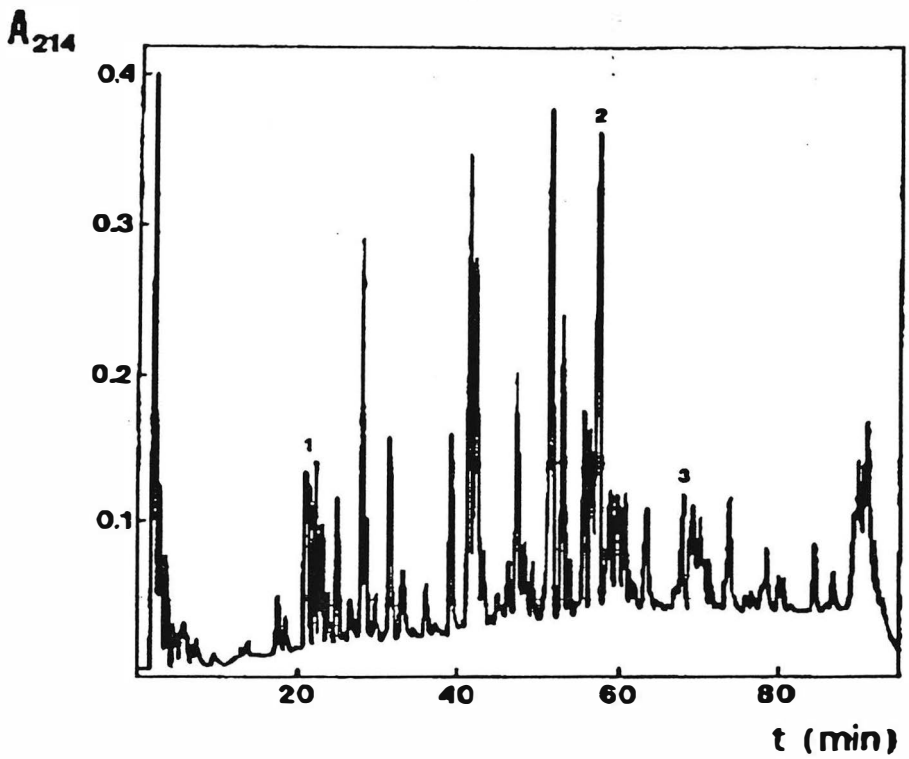


Figure 29: HPLC Elution Profile of Trypsin Digested, Carboxymethylated (^{14}C) and Reduced p26. The three peaks associated with the radioactive label are indicated.

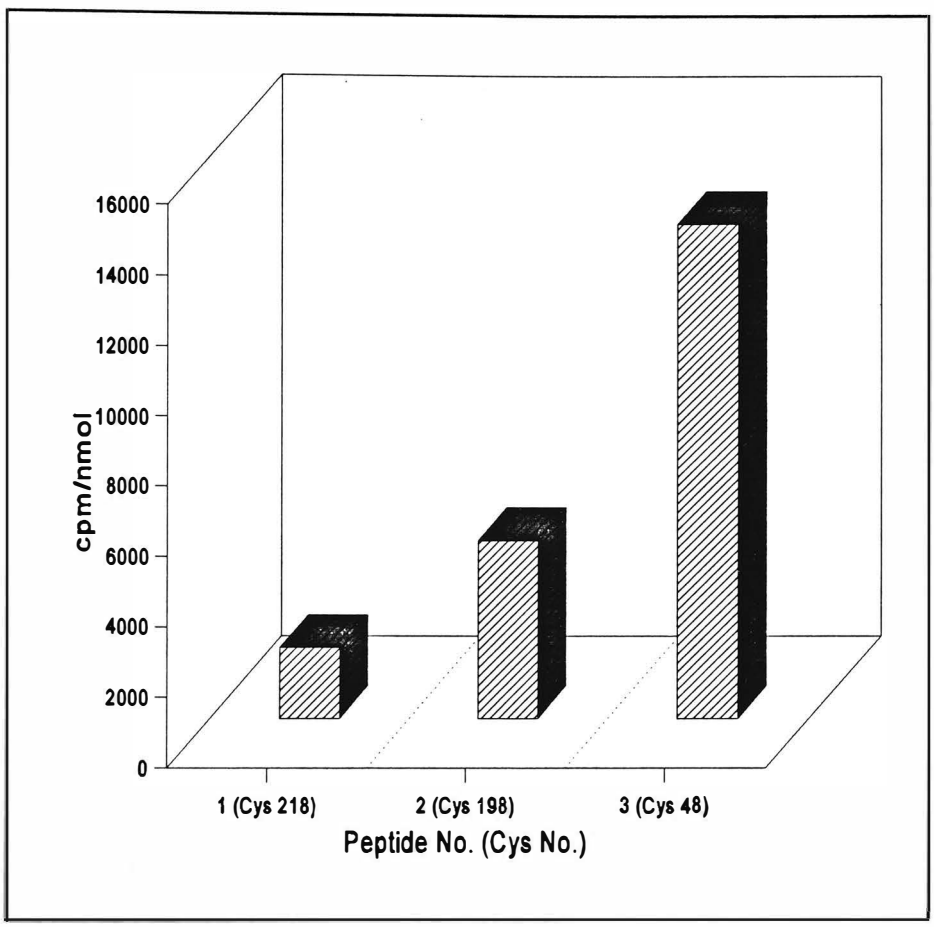


Figure 30 : Quantitation of ^{14}C Labelled Peptides.

To more precisely determine whether Cys-48 is indeed the 'free' cysteine in p26, a mutant was constructed whereby this residue was substituted for a serine. The mutant protein was purified using the same purification procedure described for the wild-type protein. The reactivity of the Cys48Ser mutant, in comparison with wild type p26, with Ellman's reagent was determined and the results are displayed in table 5. It is apparent from these experiments that Cys-48 is indeed the free cysteine residue in p26. The mutant form of p26, which contains a serine at amino acid 48, is no longer reactive toward Ellman's reagent. Further, this indicates that Cys-181 and Cys-213 are connected by an intramolecular disulfide bond.

Circular Dichroism Spectroscopy

Circular dichroism spectroscopy of p26 in the 'far' UV (Figure 31), revealed that the protein exhibits a high content of alpha-helical structure (Table 6). This is in close agreement with the values predicted for this protein using the algorithms of Chou-Fasman (Table 6). Further, reduction and carboxymethylation of p26 causes a significant alteration in the CD spectrum of p26 (Figure 31), which is indicative of significant structural changes in the protein (Table 7).

Table 5: Free Thiol Determination of Wild-Type p26 Versus the Cys48Ser Mutant. (TNB, thionitrobenzene).

Protein	[Protein]μM	A410nm	[TNB]μM	Free cys/mol
Wild Type	34.26	0.47	34.50	1.01
Cys48->Ser	35.45	0.00	0.00	0.00

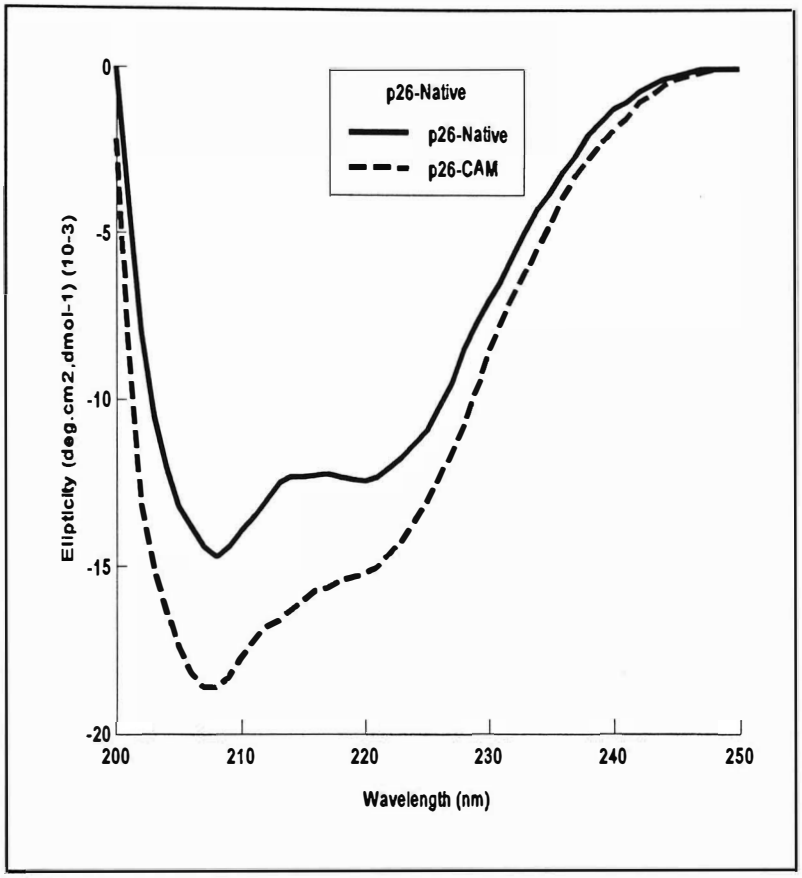


Figure 31: Circular Dichroism Spectra of p26. Far UV spectra were generated over a wavelength of 200-250nm using a protein concentration of 0.1mg/ml.

Table 6: Predicted Versus Actual Secondary Structure Comparison.
Predictions were made using the algorithms of Chou-Fasman.

	Prediction	Actual
Alpha Helix	46%	40%
Beta Sheet	24%	22%
Beta Turn	12%	10%
Random Coil	18%	28%

Table 7: Native versus Reduced/Carboxmethylated Secondary Structure Comparison.

	Native p26	CAM p26
Alpha Helix	40%	42%
Beta Sheet	22%	26%
Beta Turn	10%	0%
Random Coil	28%	32%

Fluorescence Spectroscopy

The emission spectra for native p26 shows a maximum at 337.5nm, indicating the presence of a major population of tryptophan residues, along with a less well-defined peak at 327.5nm (Figure 32). This is indicative that their two different populations of tryptophan residues, and inferences can be made regarding differences in the environment around the tryptophan residues. As a reference, tryptophan which is free in solution has a emission maximum at 348nm, whereas a tryptophan residue which is completely buried in a hydrophobic environment has emission maximum at 310-320nm. Accordingly, one can infer that the major population of tryptophan residues in p26, which emit at 337.5nm, are situated in a less hydrophobic environment than the smaller population of tryptophan residues emiting at 327.5nm. In this form, the technique does not allow for the assignment of individual tryptophan residues, but it is logical to suggest, because p26 contains only 3 tryptophan residues, that 2 of the tryptophans account for the emission maximum (337.5nm), while the remaining one accounts for the smaller maximum at 327.5nm.

In contrast, the reduced and carboxylmethylated protein exhibits a maximum at 325 and a less well-defined maximum at 337.5nm (Figure 32). The shift in fluorescence maxima to a lower wavelength upon reduction and carboxylmethylation of p26 is indicative of a change in the environment of the tryptophan residues. Specifically it suggests that the major

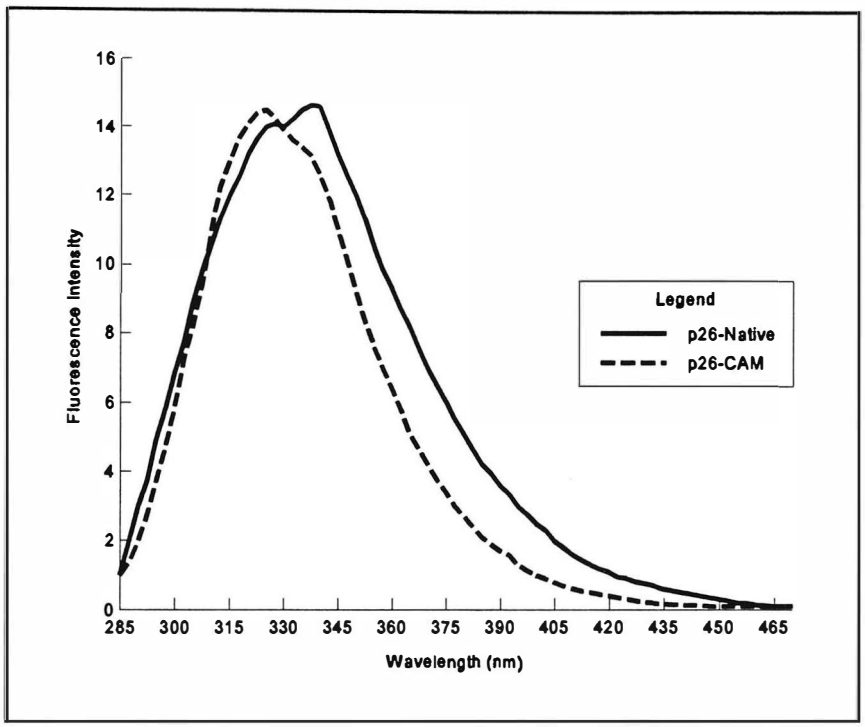


Figure 32: Fluorescence Spectra of Native and Reduced/Carboxymethylated p26.

population of tryptophan residues are in more hydrophobic environment in the modified protein compared with that experienced by the same tryptophan residues in the native protein.

Iodide Quenching of Protein Fluorescence

The modified Stern-Volmer plot for both native and carboxyamidomethylated p26 is shown in figure 33. From this plot, the following results were obtained:

-Native p26 : Fa = 0.86

-Carboxiamidomethylated p26 : Fa = 0.33

(Fa : Fraction of fluorescence accessible to the quenching agent).

The results obtained in this experiment are in concordance with those from the circular dichroism and fluorescence studies, i.e. the protein has a higher degree of ordered structure when the disulfide bridge is reduced, because, in accordance with the Fa values, the accessibility of the fluorophore to iodide is diminished. Since iodide is charged (anionic) and highly hydrated, it will only interact with those flurophores at or near the surface of the protein. Therefore, the carboxyamidomethylated protein appears to have more tryptophan residues buried in a hydrophobic environment. This confirms the observations seen with the fluorescence spectra.

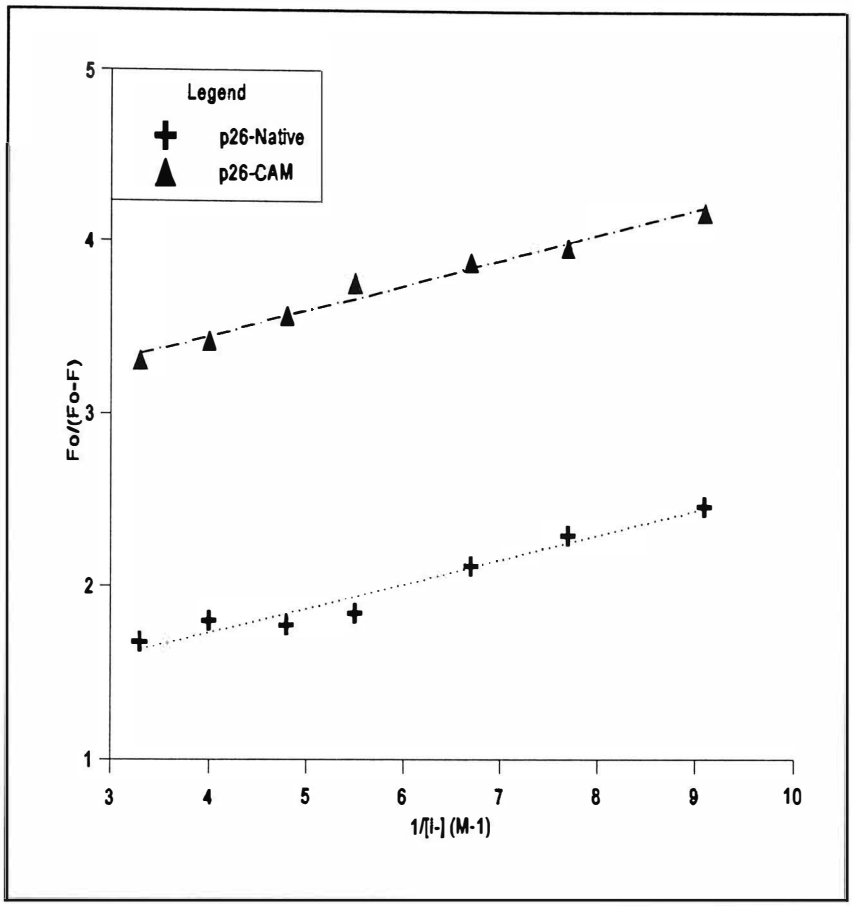


Figure 33: Modified Stern-Volmer Plot to Show the Fluorescence Quenching of Native and Reduced/Carboxymethylated p26.

Crystallization of EIAV Core Protein

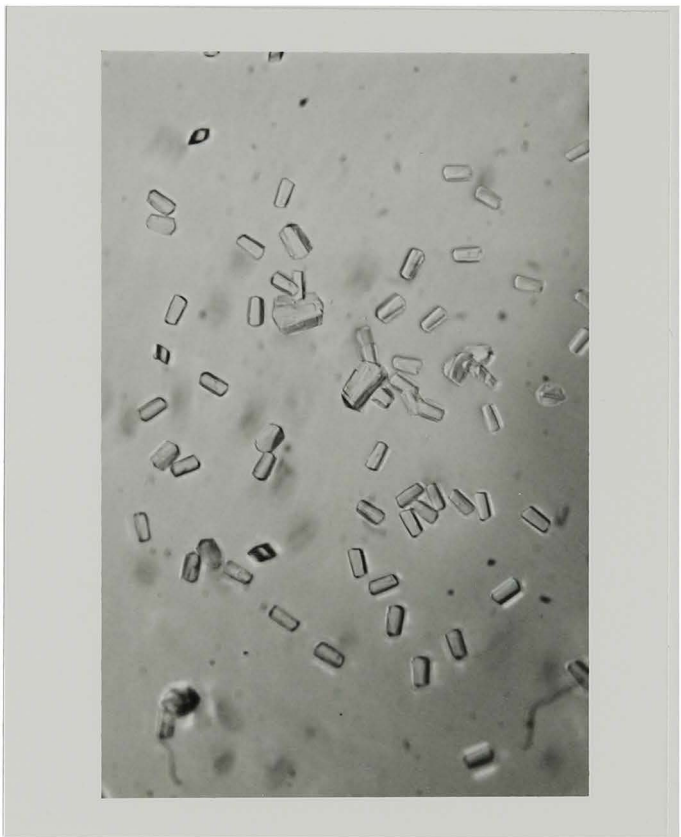
EIAV p26 core protein, purified using method one, was successfully crystallized in only one of the 50 different buffer/precipitant/salt conditions tested. The successful reagent (#7) contains 0.1M sodium cacodylate, pH 6.5, and 1.4M sodium acetate trihydrate. Crystal growth was reproducibly observed when grown at 4°C, but not at room temperature, and they exhibited a parallelopiped appearance (Fig. 34a and b).

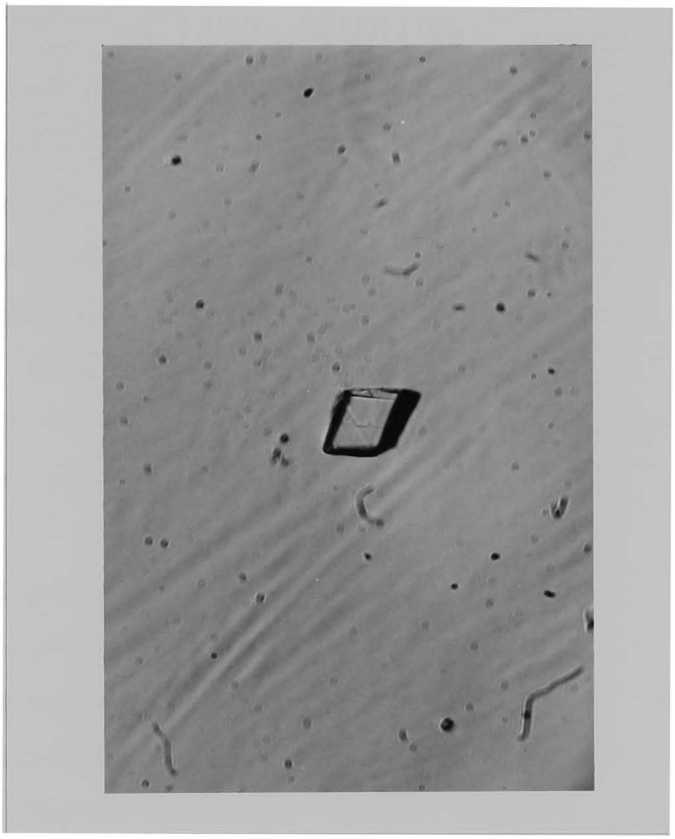
In an attempt to further refine the crystallization conditions the effects of varying the concentration of sodium acetate trihydrate was investigated. It was found that either increasing, or decreasing the acetate concentration in the drop prevented crystal growth. While increasing the acetate concentration from 0.7 to 1.0M caused aggregation of the protein, decreasing it to 0.4M resulted in an absence of crystal formation.

Varying the pH of the crystallization buffer seemed to have a somewhat less dramatic effect on the growth of the crystal. Crystals of a similar form to those observed in the original buffer were successfully grown over a pH range of 6.0-7.5. However, pH values below 6.0 caused precipitation of the protein.

Purification of p26 using method two appeared to yield a form of p26 which was not proteolytically cleaved. However, attempts to crystallize this protein, using the established crystallization conditions for p26, failed. Instead of

Figure 34: Parallelopiped-Like Crystals of EIAV Core Protein. Magnification is x100 (Panel A), and x200 (Panel B).





crystallizing, the protein precipitated, suggesting that it is less soluble than the material purified using purification scheme one.

Immunological Studies of EIAV Core Protein

Immunoblotting with EIAV Infected Serum

Immunoblotting of purified p26 protein with sera which was either positive or negative for EIAV infection, confirmed that p26 binds antibodies which are specific to EIAV infected horses (Figure 35). The four nitrocellulose strips incubated with four different EIAV positive sera show a strong reactivity toward p26, while the two EIAV negative controls show no such reactivity.

ELISA Sensitivity

The relative sensitivity of the assay, as well as relative titres of anti-p26 antibodies, were assessed by assaying 3 randomly selected EIAV positive sera at logarithmic dilutions (Table 8). All three samples were easily identifiable as being EIAV positive at a dilution of 10^{-1}, and two were identifiable at a serum dilution of 10^{-2}.

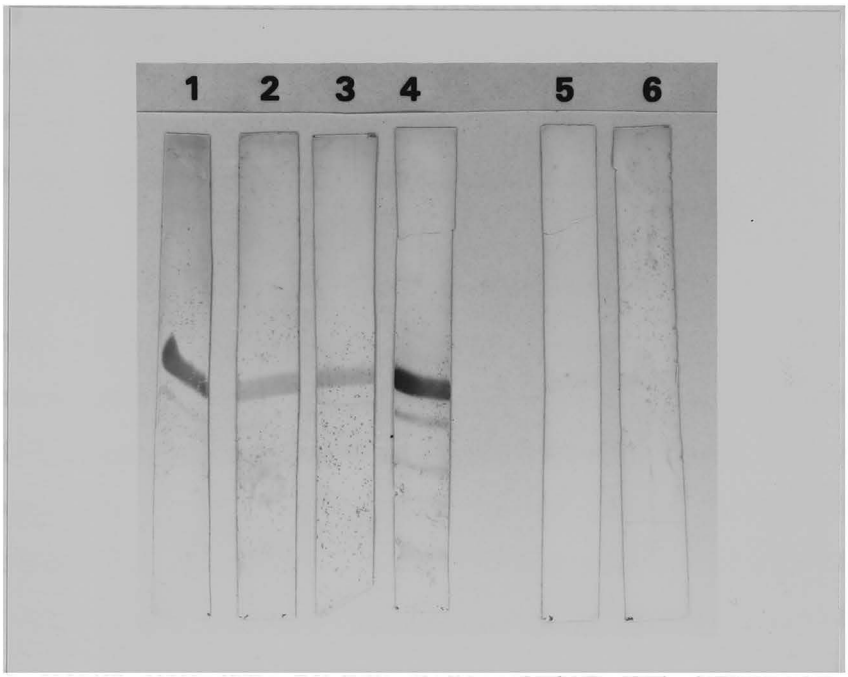


Figure 35: Immunoblotting with EIAV Infected Serum.
Lanes 1-4, EIAV infected serum; Lanes 5 & 6, normal horse serum.

Detection of EIAV Positive Sera by ELISA

The ELISA successfully identified 30 known EIAV positive serum samples which had been selected randomly following field tests conducted in the state of Virginia (Table 9). A total of 23 out of the 30 sera tested (72.6%) exhibited absorbance values of >2.0, while all 30 (100%) of the sera gave absorbance values of >1.0 (Table 9).

ELISA Analysis of a 'Blind Panel' of Horse Sera

The immunoassay was tested for it's ability to differentiate between EIAV positive and EIAV negative sera in a 'blind panel' of 30 sera supplied by the National Veterinary Services Labs (Ames, Iowa). The results, displayed in table 10 and figure 36, show that the ELISA successfully assigned all 30 of the serum samples. All negative sera gave absorbance values which did not exceed 0.05, while all positive sera exhibited absorbance values of >0.5, a value which is greater than 10 times higher than the highest 'negative'. Of the 19 positive sera, 18 (94.7%) exhibited reactivities of >1.0, a value >20 times the highest negative. Only a single serum sample (#30) gave a test result of less than 1.0; however, at greater than 10 times above the most reactive 'negative', it was still easily identifiable using this test. Serum sample #30, was confirmed by the 'National Veterinary Labs.' as having a very low titre of anti-p26 antibody.

Table 8: Results of ELISA to Measure Anti-p26 Antibody Titres in 3 Different EIAV infected horses. (U, undiluted serum).

Dilution	1	2	3
10,000	0.009	0.138	0.012
1,000	0.007	0.170	0.080
100	0.690	0.821	0.057
10	>2.0	>2.0	0.978
U	>2.0	>2.0	>2.0

Table 9: Anti-p26 ELISA Results for 30 EIAV Positive Sera. Samples were collected during random field tests in the state of Virginia.

A450	No.
0.00-0.09	0
0.10-0.49	0
0.50-1.99	8 (27%)
>2.00	22 (73%)

Table 10: Anti-p26 ELISA Results for a Panel of EIAV Positive and EIAV Negative Sera.
The results shown in parentheses are those indicated by National Veterinary Services Labs (Ames, Iowa), the supplier of the panel.
* Confirmed as the weakest positive by National Veterinary Service Labs.

No.	A450nm	Result	No.	A450nm	Result
1	1.190	+ (+)	16	>2.0	+ (+)
2	0.034	- (-)	17	>2.0	+ (+)
3	1.520	+ (+)	18	>2.0	+ (+)
4	>2.0	+ (+)	19	0.000	- (-)
5	>2.0	+ (+)	20	0.000	- (-)
6	0.037	- (-)	21	0.013	- (-)
7	>2.0	+ (+)	22	0.000	- (-)
8	0.018	- (-)	23	>2.0	+ (+)
9	1.667	+ (+)	24	0.000	- (-)
10	>2.0	+ (+)	25	>2.0	+ (+)
11	0.000	- (-)	26	>2.0	+ (+)
12	>2.0	+ (+)	27	0.000	- (-)
13	>2.0	+ (+)	28	>2.0	+ (+)
14	>2.0	+ (+)	29	0.013	- (-)
15	>2.0	+ (+)	30*	0.582	+ (+)

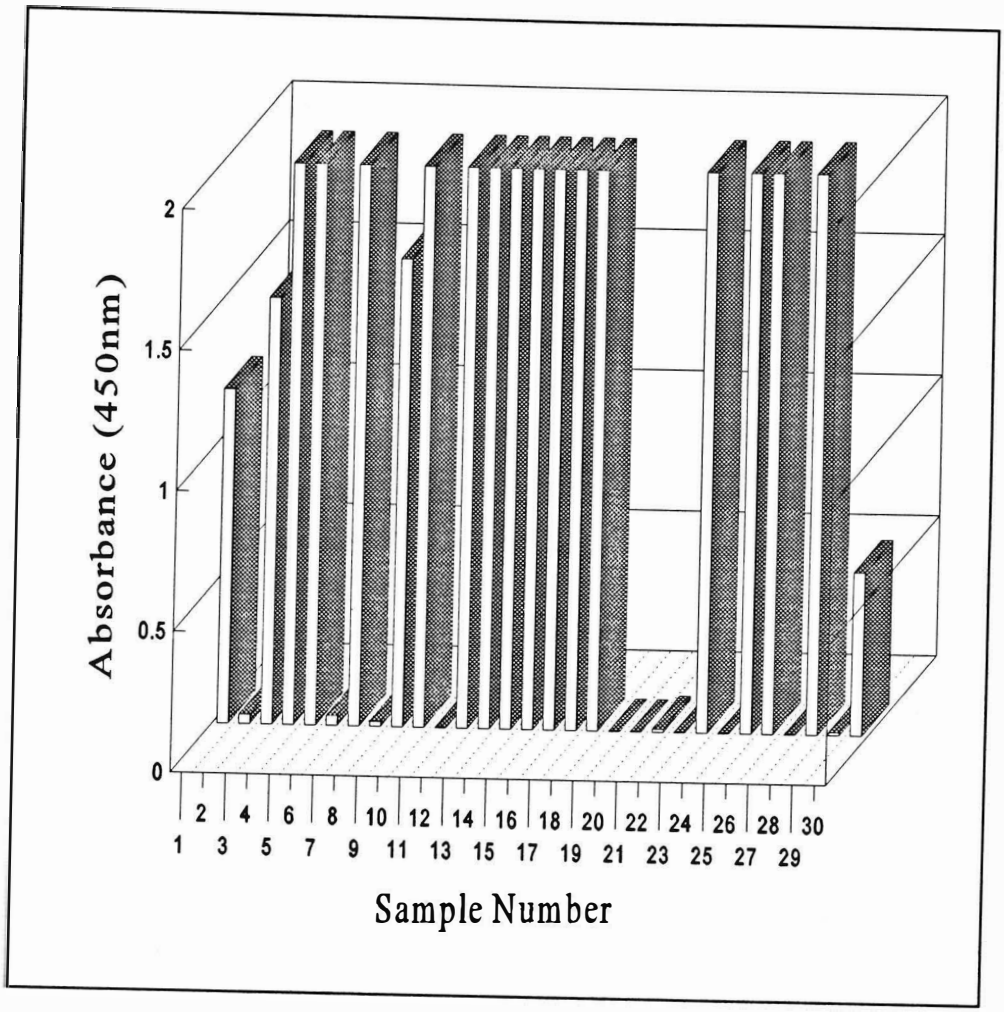


Figure 36: Plot of ELISA Data for EIAV Test Panel.

DISCUSSION

Equine infectious anaemia (EIA) virus represents a valuable animal model for studying the human immunodeficiency virus (HIV). The two viruses share many common features which extend from conserved genomic organization to striking homologies at the amino acid level. This is highlighted by the 55% similarity, and 30% amino acid identity shared by the major core proteins of HIV and EIAV (Stephens *et al.*, 1986).

The high degree of homology exhibited by p26 of EIAV and p24 of HIV-1 at the level of primary sequence would suggest that this similarity may extend to secondary and tertiary structural similarities. Indeed, three dimensional homologies between the two core proteins have been proposed as a result of previous studies conducted using a series of p26 fusion proteins. It was demonstrated that a region of 15 amino acids in p26 of EIAV, 10 of which are identical to p24 from HIV, defines an epitope which binds a particular monoclonal antibody which also binds to the core proteins of HIV-1, HIV-2 and SIV (Chong *et al.*, 1991). The existence of such an 'interspecies determinant' lends further evidence to the usefulness of p26 as a model for understanding lentiviral core

proteins (Goudsmit *et al.*, 1986). Understanding the structure and function of this family of proteins, which have important roles in virus morphology, has applications in anti-viral therapy and vaccine development (Wills and Craven, 1991).

In this report we detail the cloning, high-level expression, and purification of the major core protein (p26) of EIAV from *E.coli*. Our ability to isolate milligram quantities of this protein has enabled us to begin studying its structure and to evaluate it as a model for understanding lentiviral core proteins in general. Further, we have developed a diagnostic ELISA for capturing anti-p26 antibodies and have analyzed its suitability for detecting EIAV infection.

The structure of p26 was studied using a variety of biophysical techniques, including circular dichroism and fluoresence spectroscopy. Circular dichroism data for p26 are not consistent with the models proposed by Argos (1989) for p24 of HIV. We propose that p26 has significantly more alpha helix (40%), than proposed for p24 (24% alpha helix), which has a single, short alpha helical region. Our observation is indirectly supported by the data of Langedijk (1990), who observed that several antibody binding sites mapped to regions proposed to be beta sheet in the Argos model. The hydrophobic nature of beta sheet structure makes this highly improbable. These discrepancies may be due to legitimate structural differences between the two proteins. However, because of

their high degree of amino acid similarity and conserved function, it is highly probable that p26 from EIAV and p24 from HIV-1 are structurally similar (Chong *et al.*, 1991; Grund *et al.*, 1994; Stephens *et al.*, 1986; Goudsmit *et al.*, 1986).

In these studies, cysteine-48 of p26 was identified, using a combination of chemical modification and site-directed mutagenesis, as being a free cysteine residue. Interestingly, this residue is not conserved in the p24 from HIV (Figure 22), suggesting that it may not have an essential role in subviral particle assembly in EIAV. This observation is in contrast to cysteine residues 198 and 218, which appear to form an intramolecular disulfide bond in p26. Examination of the sequence of p26 (Figure 22), indicates that this disulfide bond results in the formation of a loop at the carboxyl-terminus of the protein. While it is not possible to speculate as to the importance of this bond in the formation of subviral core particles, it is of interest to note that both of the cysteine residues involved in forming this bond are conserved in p24 (Figure 22). This conservation would appear to suggest that this bond could be an essential feature of lentiviral major core proteins.

The role of this disulfide bond in the structure of p26 was investigated using both native and reduced and carboxyl-methylated p26 in circular dichroism, fluorescence and fluorescence quenching studies. Overall, it was demonstrated that reduction of the bond caused significant perturbation in

the structure of the protein.

The circular dichroism data for the two forms of the protein suggests that there is significant difference in the amount of beta-turn structure in the two proteins. While the native protein has around 10% beta-turn, the reduced/carboxyl-methylated protein does not have any (Figure 31 and Table 7). One plausible interpretation of these data is that the intramolecular disulfide bond stabilizes a beta-turn or loop structure at the carboxyl-terminus of the protein. The fact that the two linked cysteine residues are separated by just 20 amino acids, representing approximately 10% of the total sequence, suggests that this is indeed a possibility. However, it should be noted that the beta-turn motif contributes minimally to the overall nature of a CD spectrum, and therefore CD is not an accurate technique for beta-turn estimation.

Fluorescence and fluorescence quenching data indicate that the environment of the tryptophan residues of p26 (Trp23, 81, 132) is influenced by the presence of the disulfide bond connecting cysteine residues 198 and 218 (Figures 32 & 33). This is despite the fact that the closest tryptophan is located over 60 amino acids away from either of the two cysteine residues. Therefore, this suggests that the influences of the disulfide upon the conformation of p26 are not limited to the region of the protein in the vicinity of the bond, rather, it would appear to have an influence on the

global structure of p26.

The role of the intramolecular disulfide bond of p26 in the formation of subviral particles will be an interesting area of study. It is tempting to speculate, due to the apparent conservation of this disulfide bond, that it serves an essential role in the assembly of subviral particles.

Since subviral particles have not been successfully assembled *in vitro*, an *in vivo* system would probably be necessary for studying the role of the intramolecular disulfide bond of p26 on subviral particle assembly. A recent report by McGuire and coworkers (1994) described the successful expression of subviral EIAV particles in equine kidney cells infected with vaccinia viruses expressing the EIAV *gag* gene. This would appear to be an excellent system for studying the role of this intramolecular disulfide bond in the in vivo assembly of subviral particles. Their reports showed that the Gag precursor protein (p55), when expressed in the absence of other EIAV proteins, is able to assemble into subviral particles. However, the final processing of the precursor does require the co-expression of the viral protease, a product of the proteolytic processing of the Pol precursor (McGuire *et al.*, 1994). In addition to an *in vivo* system for studying particle formation, a mutated form of p26 would be necessary to undertake such studies. Specifically, the p26 mutant would be required to have one of the cysteine residues involved in the formation of the intramolecular

disulfide bond (Cys198 or Cys218), mutated to a similar amino acid (e.g. serine). This could easily be achieved using site-directed mutagenesis.

Assuming that the recombinant form of p26 accurately mimics the structure of the native protein found in virions, and that the intramolecular disulfide bond is relevant *in vivo*, the intramolecular disulfide bond exhibited by p26 must form at one of two possible times. Firstly, the disulfide could form immediately following translation, when p26 is still part of the 55 kDa precursor, indicating that it could have a role in the function of p55. The second possibility is that the disulfide bond does not form until p55 has been processed to yield p26, in which case it may be involved in subviral particle assembly.

If the bond does form in p55, it may be involved in the maintenance of p55-Gag in a form which can interact with the plasma membrane. If this is the case, and the bond is important in the function of the p55 precursor, then subviral particle formation would be expected to be affected by a Cys 218 (or Cys198) mutation. Alternatively, the bond could be involved in maintaining the p55 precursor in a conformation that can be processed effectively by the viral protease. Either of these scenarios could be effectively tested using the vaccinia virus system of McGuire *et al*. (1994).

If the disulfide bond plays a role in the function of p55, then subviral particle formation would be expected to be

affected by the mutation. However, if it is important for processing of the gag precursor, then particle formation would not be expected to be affected since subviral particle formation has been observed in the absence of processing of the p55 precursor (McGuire *et al.*, 1994). Rather one would predict that the processing of p55 to yield p26 in the subviral particles would be affected. This could easily be tested by examining the subviral particles for evidence of processing by immunoblotting.

Another possibility is that the bond does not have a role in the targeting or processing of p55, but instead forms after the proteolytic processing of p55, an event which occurs after subviral particle assembly (McGuire *et al.*, 1994). This would suggest that the disulfide functions to stabilize the formation of p26 containing subviral particles. Since the bond is intramolecular, the study of it's role in stabilizing p26 monomers does not require particle formation. Therefore, an *in vitro* technique such as differential scanning calorimetry (D.S.C.), could be used to study this.

The recent appreciation of the importance of lentiviral Gag proteins in the assembly of infectious viral particles has led to new approaches to the development of anti-viral therapies and vaccines. Interfering with the Gag-mediated budding of viral particles, an essential feature of the viral lifecycle, has been proposed as a therapy for lentiviral infections. One approach which has shown some degree of

promise as a future therapy involves the mutation of HIV-1 Gag protein such that it can no longer form sub-viral particles (Trono, *et al.*, 1989). The mutant Gag protein, when co-expressed with wild-type Gag, is able to exert a dominant interference on normal particle assembly, resulting in a decrease in the amount of infectious virus produced. Confirmation of an essential role for the intramolecular disulfide bridge in p26, and possibly p24 of HIV-1, could provide another mechanism for interfering with subviral particle assembly *in vivo*. It is envisaged that this type of treatment, via gene therapy, could function by decreasing the amount of infectious virus produced by the host, and so slow the development of the disease. This type of approach to lentiviral therapy highlights the need for a greater understanding of the structure and function of lentiviral Gag proteins.

The structure of Gag proteins also has important implications in lentiviral vaccine development. One of the most promising vaccines for targeting lentiviral infections has been immunization with whole-killed virus. Unfortunately, ethical concerns are certain to prevent this kind of approach to human vaccine development. However, the apparent effectiveness of whole-killed virus preparations, tied to the disappointing results attained with individual proteins, has accelerated the development of a 'pseudovirus' approach to vaccine development. In these systems the viral genome is

mutated to render it non-infectious, and the viral structural genes (*pol, env, and gag*) are expressed from heterologous promoters. The intent of these systems is to create virus particles, which while harmless, contain the full complement of structural antigens. The priming of the host's immune system to antigens which would be presented in an analogous manner to infection is believed to be far superior to immunization with individual viral antigens. An additional feature of this type of vaccine is that the structural antigens can be engineered to enhance their immunogenicity and decrease immunosuppressive properties (Haynes *et al.*, 1991). The fact that Gag proteins contain large regions of sequence which are dispensable for sub-viral particle formation means that additional copies of protective or virus neutralizing epitopes can be engineered into the Gag protein sequence (Wills and Craven, 1991).

Our ability to crystalize p26 represents the first report of this being achieved for a major lentiviral core protein in the absence of other proteins. It should be noted that a 1990 paper reported the co-crystallization of p24, the major core protein of HIV-1, with an Fab antibody fragment (Prongay *et al.*, 1990). However, the crystal structure of p24 has not been published, suggesting that this approach may have not generated crystals which were suitable for X-ray diffraction.

In this study, the maximum diameter of the crystal grown was approximately 0.1mm, which is slightly smaller than

required for X-ray diffraction. One possible explanation for our inability to grow crystals beyond 0.1mm diameter could be heterogeneity in the purity in the protein. Accordingly, we developed a more rapid purification procedure for p26 (method 2), with the rationale that reducing the exposure time of the protein to cellular proteases may increase the homogeneity of the preparation, and in the process afford the growth of larger crystals. Interestingly, the proteolytic cleavage of p26 observed in method 1 apppeared to facilitate crystal growth rather than inhibit it. The protein purified by method two appears to be less soluble than that purified using method one, and precipitates when set up to crystallize. We propose that this observation could be due to proteolytic cleavage of the crystallizable material during purification. Because method two is a more rapid purification scheme, and lacks the hydroxyapatite step, it is plausible that p26 purified in this manner does not get cleaved. Analysis of the sequence of p26 at it's carboxyl-terminus reveals a very hydrophobic stretch of amino acids with the following sequence, MMLLAKAL. Seven of the eight amino acids in this stretch are hydrophobic, suggesting that if they are not cleaved from p26, they may function to decrease the solubility of the protein, which has been observed. Accordingly we have created a truncated form of p26 which ends at Lys227, the amino acid which precedes the hydrophobic tail. This clone directs the expression of a 224 amino acid version of p26, and

should provide a more homogenous source of truncated protein than that produced by the proteolytic cleavage of p26. Investigations are currently underway to assess the ability of this protein to crystallize, and specifically whether it can be used to grow larger crystals than those which have been grown to date.

A knowledge of the 3-dimensional structure of capsid proteins provides important structural information which assists in understanding how the structure of the protein is involved in its functionality, which is accepted as being essential for infectivity. Such structural information could ultimately be exploited in aiding the design of anti-viral agents for blocking capsid assembly/disassembly. This approach has been shown to be an effective therapy in targeting human rhinoviruses. In these studies, a series of hydrophobic organic compounds which bind to a hydrophobic pocket in the capsid were developed and shown to impart stability on the assembled capsids (Smith *et al.*, 1986). This in turn prevents disassembly of the capsids, blocking virus 'uncoating', an essential stage in the viral lifecycle; further, it was also shown that the compounds inhibit receptor binding. A similar approach has been proposed for targeting HIV, however no lentiviral capsid structures have been solved to date.

The *in vitro* study of viral core assembly for HIV-1 has been approached by Ehrlich and co-workers using recombinant p24 purified from *E.coli* (Ehrlich *et al.*, 1990). They have

shown that *in vitro*, recombinant core protein assembled into long rod-like structures (Ehrlich *et al.*, 1992). Although these structures do not mimic the cone-shaped capsid seen *in vivo*, the data does demonstrate an intrinsic ability of p24 to form stable, oligomeric structures. Self-assembly appears to be dependent on protein concentration, pH and ionic strength (Ehrlich *et al.*, 1992). Further, it has been shown using analytical affinity chromatography (AAC) that p24 can reversibly form dimers *in vitro* (Rose S. *et al.*, 1992).

We are currently investigating these observations with p26 to see if it too is also able to self assemble into oligomeric structures, although we have no evidence of this at the present time. Ultimately, our findings should help to determine whether such 'spontaneous oligomerization' is a conserved feature of lentiviral core proteins. If it is, it would provide a valuable '*in vitro*' system for evaluating the ability of drugs to stabilize/destabilize core protein structures; such interactions could be studied using a variety of biophysical approaches. However, recent observations made with EIAV indicated that the p55 Gag precursor, and not processed p26, initially forms subviral particles *in vivo* (McGuire, *et al.*, 1994). This would suggest that *in vitro* studies of core particle assembly may require unprocessed Gag precursors instead of just the major core protein. This observation may help to explain why previous *in vitro* studies of core particle assembly for p24 do not mimic the events

observed *in vivo*.

The inability of lentiviral core proteins to self-assemble *in vitro* into the cone-shaped capsids seen *in vivo* indeed suggests that other factors, including other viral proteins, may be important players in the assembly process. In HIV-1, the viral infectivity factor (Vif) appears to be a candidate for a role in nucleocapsid assembly. It had previously been demonstrated that Vif increases viral infectivity at the time of virus production, most probably by affecting viral protein processing, virus assembly, or virus maturation. (Gabuzda *et al.*, 1992; Blanc *et al.*, 1993; Sakai *et al.*, 1993). Recently, it has been reported that the *vif* gene product may play a role in packaging of the core of HIV-1 (Hoglund *et al.*, 1994). In the absence of Vif, the majority of virus particles did not exhibit the normal dense, cone-shaped core structure. Instead they appeared as heterogeneous, irregular structures. The mechanism by which Vif exerts its effect on nucleocapsid assembly is not fully understood; however, the recent observation that Vif is predominantly a cytoplasmic protein associated with the cytoplasmic side of the plasma membrane has lead some to suggest that Vif may play a role during the final stages of virion assembly at the plasma membrane (Goncalves *et al.*, 1994).

Another interesting observation regarding the *vif* gene product was recently reported by Ma and coworkers (1994). They identified a pair of cysteine residues in the *vif* gene of

HIV-1, which are conserved among the animal lentiviruses, which appear to serve a critical role in the function of the virus. When the cysteine residues in the full-length HIV-1 genome were mutated to leucines, it was demonstrated that the virions adopted a *vif*- phenotype. The *vif*- phenotype was subsequently abolished by complementation with a Vif expression vector (Ma *et al.*, 1994). These observations suggested that the pair of cysteine residues, which could function by forming a disulfide bond, were critical for viral function and infectivity.

Like the analogous protein in HIV (p24), p26 is a highly immunogenic protein which is very highly conserved among different viral isolates. For this reason, antibody responses to it form the basis of diagnostic assay for determining infection. The standard assay for determining EIAV infection in horses is the 'Coggins Test' which was introduced in the early 1970's (Coggins and Lombard, 1970). This assay takes the form of an agar gel immunodiffusion reaction, or oucterlony, and detects antibodies to p26 which is isolated from cell cultures infected with the EIA virus. Despite the obvious shortfalls of an immunodiffusion reaction for diagnostic purposes, it remains the most widely used diagnostic tool and is regarded as the most reliable assay currently available.

Using recombinant p26 from *E.coli*, we have developed a rapid, sensitive and highly accurate sandwich ELISA for

detecting anti-p26 antibodies in sera. The assay takes just 1 hour to complete and has successfully identified all EIAV positive samples tested to date. Most notably it passed a 'blind' test from the National Veterinary Service Labs (Ames, Iowa); successfully identifying 19 positive sera, of varying antibody titre, and 11 negative sera.

An ELISA such as the one described here has many potential advantages over an immunodiffusion reaction. The first is the speed with which the reaction can be conducted (1 hour), in comparison to the 'Coggins' Test' which takes 6-12 hours. A second advantage is the ease with which the result of the test, a colorometric reaction, can be interpreted. Further, the assay detects antibodies to p26, a highly conserved and immunogenic protein, which has formed the basis of the longstanding 'Coggins' test.

One potential drawback of an ELISA is the necessity for diluting serum samples prior to testing, a step which is essential when using anti-antibody conjugates for detecting antibody. However, in our ELISA this problem was circumvented by making a p26-peroxidase conjugate for detecting captured antibody. It is inferred that only one of the Fab portions of the antibody is utilized in binding the antibody to the p26 immobilized on the plate, leaving a second Fab binding site exposed. It is this site which binds the p26-peroxidase conjugate, allowing detection of bound antibody. Because the p26-peroxidase does not have an affinity for the constant

regions of immunoglobulin molecules, unlike anti-antibody conjugates, background reactivity is not a concern, even when using undiluted serum. One of the few potential problem with such a 'sandwich ELISA' is the possible loss in sensitivity due to the binding of both of the Fab sites on an antibody to the immobilized p26. This does not appear to be a major concern since all positive sera have been correctly identified so far. However, it may be possible to increase the sensitivity of the assay still further by elevating the relative excess of antibody over immobilized antigen, and therefore reduce the possibility of both Fab sites on the antibody interacting with the immobilized p26. This could be approached by reducing the amount of p26 coated to the microtitre wells.

It is envisaged that ultimately the assay could be modified to adopt the form of a 'dip-stick' reaction. This type of assay, which forms the basis of current pregnancy tests, can be conducted in minutes, without the need for any laboratory equipment or expertise. Because this virus is a worldwide concern, the market for such a simple and rapid diagnostic test would be very significant.

A recombinant source of p26 protein will also permit investigation of the immunological properties of this protein. The sequence homology of the MHR with core proteins from other retroviral core proteins has led to speculation that this sequence may harbor an inter-species cross-reactive

determinant (Grund *et al.*, 1994; Wills and Craven, 1991). Recent attempts to precisely identify the site of cross-reactivity using a panel of overlapping synthetic peptides were unsuccessful, although the cross-reactive antibodies were shown to be specific for the carboxyl-terminal 83 residues of p26 (153-235) (Grund *et al.*, 1994). The inability to precisely localize this epitope was presumed to indicate that the cross-reactive epitope of p26 is conformation-dependent, and therefore not mimicked by short synthetic peptides. Mutational analysis of p26, by substitution of amino acids within the MHR, would certainly enable accurate mapping of the cross-reactive epitope.

In this report we emphasize the tremendous potential of the animal lentiviruses to serve as models for understanding this important family of viruses, which includes the human immunodeficiency virus. Clearly a recombinant source of p26 from EIAV has important applications in a variety of structural, functional and immunological studies. In light of the conserved nature of the lentiviral major core proteins, it is envisaged that other major core proteins from other members of the *Lentiviridae*, particularly the antigenically related proteins from FIV and BIV, could be cloned, expressed, purified, and studied in the manner described here for p26. A recombinant source of BIV and FIV core protein would also permit the development of a sandwich ELISA, analogous to the one described here for EIAV, for commercial use as a

diagnostic assay.

CONCLUSIONS

The suitability of p26, the major core protein from equine infectious anaemia virus (EIAV), to serve as a model for understanding the structure and function of lentiviral core proteins was studied. The protein was expressed, using an *E.coli* expression system, and purified in sufficient quantities to allow structural studies to be made.

Circular dichroism spectroscopy revealed that p26 has the following assignment of secondary structure; 40% alpha helix, 22% beta sheet, 10% beta turn and 28% random coil. The high degree of alpha helical content of p26 was in agreement with the predicted values acquired using the method of Chou and Fasmann (46% alpha helix). However, these results contrasted to the proposed structural features of another lentiviral core protein, p24 from HIV, which is predicted to be comprised predominantly of beta structure. This is despite the fact that these two core proteins share 55% amino acid similarity.

It was determined, using a combination of site-directed mutagenesis and chemical modification, that p26 contains an intramolecular disulfide bridge connecting Cys198 and Cys218, and a single free cysteine (Cys48). Further, it was demonstrated by fluorescence studies that the integrity of the

disulfide bond affects the global structure of the protein. The presence of this pair of cysteine residues in p24 of HIV-1 suggests that this disulfide bond may have a conserved role in the function of lentiviral core proteins, and may be essential in the lifecycle of these viruses.

A sandwich ELISA was developed for measuring anti-p26 antibodies in sera, and was evaluated as a diagnostic tool for detecting EIAV infection. The assay appears to be extremely reliable for detecting infected sera, and exhibits extremely low background reactivities with EIAV negative sera. Accordingly, the interpretation of the data generated using this test would appear to be simple and unambiguous.

The expression system and purification procedure developed for EIAV core protein may be useful for producing milligram-gram quantities of other lentiviral core proteins for structure, function and immunological studies. It is envisaged that this could be extended to include the development of marketable diagnostic ELISA's for other animal lentiviruses such as feline immunodeficiency virus (FIV).

LITERATURE CITED

LITERATURE CITED

Aach, R.D., Szumuness, W., Mosley, J.W., Hollinger, F.B., Kahn, R.A., Stevens, C.E., Edwards, V.M., and Werch, J. (1981) "Serum Alanine Aminotransferase of Donors in Relation to the Risk of Non-A, Non-B Hepatitis in Recipients", *N. Engl. J. Med.*, **304,** 989-994.

Agawai, K.L., Buchi, H., Caruthers, M.H., Gupta, N., Khorana, H.G., Kleppe, K., Kumar, A., Ohtsuka, E., Raj Bhandary, U.L., van de Sande, J.H., Sgaramella, V., Weber, H., and Yamada, T. (1970) "Total Synthesis of the Gene for an Alanine Transfer Ribonucleic Acid from Yeast", *Nature* **227,** 27-34.

Akatsuka, T., Donets, M., Scaglione, L., Ching, W.-M., Shih, J.W.-K., Di Bisceglie, A., and Feinstone, S.M. (1993) "B-cell Epitopes on the Hepatitis C Virus Nucleocapsid Protein Determined by Human Monospecific Antibodies", *Hepatology* **18(3),** 503-510.

Alter, J.H., and Hoofnagle, J.H. (1984) "Non-A, Non-B. Observations on the First Decade." In: Vyas, G.N., Dienstag, J.H., and Hoofnagle, J.H. (Eds): Viral Hepatitis and Liver Disease. Orlando, Grune & Stratton, 345-355.

Argos, P. (1989) "A Possible Homology Between Immunodeficiency Virus p24 Core Protein and Picornaviral VP2 Coat Protein: Prediction of HIV p24 Antigenic Sites", *EMBO Journal* **8,** 779-785.

Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and Jacobsen, H. (1993) "Nonstructural Protein 3 of the Hepatitis C Virus Encodes a Serine-Type Proteinase Required for Cleavage at the NS3/4 and NS4/5 Junctions", *Journal of Virology* **67(7),** 3835-3844.

Blanc, D., Patience, C., Schultz, T.F., Weiss, R., and Spire, B. (1993) "Transcomplementation of Vif-HIV-1 Mutants in CEM Cells Suggests that Vif Affects Late Steps of the Viral Life Cycle", *Virology* **193,** 186-192.

Botarelli, P., Brunetto, M.R., Minutello, M.A., Calvo, P., Unutmaz, D., Weiner, A.J., Choo, Q.-L., Shuster, J.R., Kuo, G., Bonini, F., Houghton, M., and Abrignani, S. (1993) "T-Lymphocyte Response to Hepatitis C Virus in Different Clinical Courses of Infection", *Gastroenterology* **104,** 580-587.

Bryant, M. and Ratner, L. (1990) "Myristoylation-Dependent Replication and Assembly of Human Immunodeficiency Virus 1", *Proc. Natl. Acad. Sci. USA* **87,** 523-527.

Burnette, W.N. (1981) "Westerm Blotting: Electrophoretic Transfer of Proteins from Sodium Dodecyl Sulphate-Polyacrylamide Gels to Unmodified Nitrocellulose and Radiographic Detection with Antibody and Radioiodinated Protein A", *Anal. Biochem.* **112,** 195-203.

Castillo, I., Bartolome, J., Quiroga, J.A., and Carreno, V. (1992) "Comparison of Several Different Procedures for Detection of Serum HCV-PCR Using Different Regions of the HCV Genome", *J. Virol. Meth.* **38,** 71-80.

Chambers, T.J., Hahn, C.S., Galler, R., and Rice, C.M. (1990) "Flavivirus Genome Organization, Expression, and Replication", *Annu. Rev. Microbiol.* **44,** 649-688.

Charman, H.P., Bladen, S., Gilden, R.V., and Coggins, L. (1976) "Equine Infectious Anemia Virus: Evidence Favoring Classification as a Retrovirus", *J.Virol.* **19,** 1073-1079.

Chemello, L., Cavalletto, D., Pontisso, P., Bortolotti, F., Donada, C., Donadon, V., Frezza, M., Casarin, P., and Alberti, A. (1993) "Patterns of Antibodies to Hepatitis C Virus in Patients with Chronic Non-A, Non-B Hepatitis and Their Relationship to Viral Replication and Liver Disease", *Hepatology* **17(2),** 179-182.

Chen, P.-J., Wang, J.-T., Hwang, L.-H., Yang, Y.-H., Hsieh, C-L., Kao, J.-H., Sheu, J.-C., Lai, M.-Y., Wang, T.-H., and Chen, D.-S. (1992) "Transient Immunoglobulin M Antibody Response to Hepatitis C Virus Capsid Antigen in Posttransfusion Hepatitis C: Putative Serological Marker for Acute Viral Infection", *Proc. Natl. Acad. Sci. USA* **89,** 5971-5975.

Chiba, J., Ohba, H., Matsuura, Y., Watanabr, Y., Katayama, T., Kikuchi, S., Saito, I., and Miyamura, T. (1991) "Serodiagnosis of Hepatitis C Virus (HCV) Infection with an HCV Core Protein Molecularly Expressed by a Recombinant Baculovirus", *Proc. Natl. Acad. Sci. USA* **88,** 4641-4645.

Chiu, I-M., Yaniv, A., Dahlberd, J.E., Gazit, A., Skuntz, S.F., Tronick, S.R., and Aaronson, S.A. (1985) "Nucleotide Sequence Evidence for Relationship of AIDS Retrovirus to Lentiviruses", *Nature* **317,** 366-368.

Chong, Y-H., Payne, S.L., Issel, C.J., Montelaro, R.C., and Rushlow, K. (1991) "Characterization of the Antigenic Domains of the Major Core Protein (p26) of Equine Infectious Anaemia Virus", *Journal of Virology* **65,** 1007-1012.

Choo, Q.-L., Kuo, G., Weiner, A.J., Overby, L.R. Bradley, D.W., and Houghton, M. (1989) "Isolation of a cDNA Clone Derived from a Blood-Borne Non-A, Non-B Viral Hepatitis Genome", *Science* **244,** 362-364.

Choo, Q.-L., Kuo,G., Ralston, R., Weiner, A., Chien, D., Van Nest, G., Han, J., Berger, K., Thudium, K., and Kuo, C. (1994) "Vaccination of Chimpanzees Against Infection by the Hepatitis C Virus", *Proc. Natl. Acad. Sci. USA* **91(4),** 1294-1298.

Choo, Q.-L., Weiner, A.J., Overby, L.R., Kuo, G., Houghton, M., and Bradley, D.W. (1990) "Hepatitis C Virus: The Major Causative Agent of Viral Non-A, Non-B Hepatitis", *Br. Med. Bull.* **46,** 423-441.

Chou, P.Y. and Fasman, G.D. (1974) "Conformational Parameters for Amino Acids in Helical, β-sheet and Random Coil Regions Calculated from Proteins", *Biochemistry* **13,** 211-222.

Coates, A.R., Cookson, J., Barton, G.H., Zvelebil, M.J., and Sternberg, M.J.E. (1987) "AIDS vaccine predictions", *Nature* **326,** 549-550.

Cleland, J.L., Builder, S.E., Swartz, J.R., Winkler, M., Chang, J.Y., and Wang, D.I. (1992) "Polyethylene Glycol Enhanced Protein Refolding", *BioTechnology* **10,** 1013-1019.

Coggins, L. and Norcross, N.L. (1970) "Immunodiffusion Reaction in Equine Infectious Anaemia", *Cornell Vet.* **60,** 330-335.

Colombo, M., Kuo, G., Choo, Q.-L., Donato, M.F., Del Ninno, E., Tonnasini, M.A., Dioguardi, N., and Houghton, M. (1989) "Prevalence of Antibodies to Hepatitis C Virus in Italian Patients with Hepatocellular Carcinoma", *Lancet ii,* 1006-1008.

Davis, G.L., Balart, L. A., Schiff, E. R., Lindsay, K., Bodenheimer, H.C., Perrillo, R.P., Carey, W., Jacobson, I.M., Payne, J., Dienstag, J.L., VanThiel, D.H, Tamburro, C., Lefkowitch, J., Albrecht, J., Meschievitz, C., Ortego, T.L., Gibas, A., and the Hepatitis Interventional Therapy Group (1989) "Treatment of Chronic Hepatitis C with Recombinant Interferon Alfa. A Muliticenter Randomized, Controlled Trial", *N. Engl. J. Med.* **321,** 1501-1506.

DiBisceglie, A.M., Martin, P., Kassianides, C., Lisker-Melman, M., Murray, L., Wagganer, J., Goodman, Z., Banks, S.M., and Hoofnagle, J.H. (1989) "Recombinant Interferon Alfa Therapy for Chronic Hepatitis C. A Randomized, Double-Blind, Placebo-Controlled Trial", *N. Engl. J. Med.* **321,** 1506-1510.

Dupraz, P., Oertle, S., Meric, C., Danay, P., and Spahr, P.F. (1990) "Point Mutation in the Proximal Cys-His Box of Rous Sarcoma Virus Nucleocapsid Protein", *J. Virol.* **64,** 4978-4987.

Egberink, H.F., Ederveen, J., Montelaro, R.C., Pederson, N.C., Horzinek, M.C., and Koolen, M.J.M. (1990) "Intracellular Proteins of Feline Immunodeficiency Virus and their Antigenic Relationship with Equine Infectious Anaemia Virus Proteins", *J. Gen. Virol.* **71,** 739-743.

Ehrlich, L.S., Agresta, B.E., and Carter, C.A. (1992) "Assembly of Recombinant Human Immunodeficiency Virus Type 1 Capsid Protein in vitro", *J. Virol.* **66,** 4874-4883.

Forster, S.M., Osbourne, L.M., Chiengsong-Popov, R., Kenny, C., Burnell, R., Jeffries, D.J., Pinching, A.J., Harris, T.R., and Weber, T.N. (1987) "Decline of Anti-p24 Antibody Precedes Antigenaemia as Correlate of Prognosis of HIV-1 Infection", *AIDS* **1,** 235-240.

Fujii, H., Moriyama, K., Sakamoto, N., Kondo, T., Yasuda, K., Hiraizuma, Y., Yamazaki, M., Sakaki, Y., Okochi, K., and Nakajima, E. (1992) "Gly145 to Arg Substitution in HBs Antigen of Immune Escape Mutant of Hepatitis B virus", *Biochem. Biophy. Res. Comm.* **184,** 1152-1157.

Gabuzda, D.H., Lawrence, K., Langhoff, E., Terwilliger, E., Dorfman, T., Haseltine, W.A., and Sodroski, J. (1992) "Role of Vif in Replication of Human Immunodeficiency Virus Type 1 in CD4+ T-Lymphocytes", *J. Virol.* **66,** 6489-6495.

Goncalves, J., Jallepalli, P., and Gabuza, D. (1994) "Subcellular Localization of the Vif Protein of Human Immunodeficiency Virus Type 1", *J. Virol.* **68,** 704-712.

Gorelick, R.J., Nigida, S.M., Bess, J.W., Artur, L.O., Henderson, L.E., and Rein, A. (1990) "Noninfectious Human Immunodeficiency Virus Type 1 Mutants Deficient in Genomic RNA", *J. Virol.* **64,** 3207-3211

Gottlinger, H.G., Sodroski, J.G., and Haseltine, W.A. (1989) "Role of Capsid Precursor Processing and Myristoylation in Morphogenesis and Infectivity of Human Immunodeficiency Virus Type 1", *Proc. Natl. Acad. Sci. USA* **86,** 5781-5785.

Goudsmit, J., Houwers, D.J., Smit, L., and Nauta, I.M. (1986) "LAV/HTLVIII *gag* Gene Product p24 Shares Antigenic Determinants with Equine Anaemia Virus but not with Visna Virus or Caprine Arthritis Encephalitis Virus", *Intervirology* **26,** 169-173.

Grakoui, A., Wychowski, C., Lin, C., Feinstone, S.M., and Rice, C.M. (1993) "Expression and Identification of Hepatitis C Virus Polyprotein Cleavage Products", *J.Virol.* **67,** 1385-1395.

Grund, C.H., Lechman, E.R., Issel, C.J., Montelaro, C., and Rushlow, K.E. (1994) "Lentivirus Cross-Reactive Determinants Present in the Capsid Protein of Equine Infectious Anaemia Virus", *J. Gen. Virol.* **75,** 657-662.

Haist, S., Stuber, G., Matz, J., Rohrhoffer, A., Wolf, H., and Modrow, S. (1993) "Immunological Reactivity Against Gag Proteins in Patients" *Annual Meeting for the German Society for Virology*, Hamburg, March 17-20.

Harada, S., Watanabe, Y., Takeuchi, K., Suzuki, T., Katayama, T., Takebe, Y., Saito, I., and Miyamura, T. (1991) "Expression of Processed Core Protein of Hepatitis C Virus in Mammalian Cells", *J.Virol.* **65,** 3015-3021.

Haynes, J.R., Cao, S.X., Rovinski, B., James, O., and Klein, M.H. (1991) "Production of Immunogenic HIV-1 Virus-like Particles in Stably Engineered Monkey Cell Lines", *AIDS Res. Hum. Retroviruses* **7,** 17-27.

Henderson, L.E., Sower, R.C., Smythers, G.W., and Oroszlan, S. (1987) "Chemical and Immunological Characterization of Equine Infectious Anemia Virus Gag-Encoded Proteins", *J. Virol.* **61,** 1116-1124.

Hijikata, M., Kato, N., Ootuyama, Y., Nakagawa, M., and Shimotohno, K. (1991) "Gene Mapping of the Putative Structural Region of the Hepatitis C Virus Genome by *in vitro* Processing Analysis", *Proc. Natl. Acad. Sci. USA* **88,** 5547-5551.

Hirel, P.H., Schmitter, M.J., Dessen, P. Fayat, G., and Blanquet, S. (1989) "Extent of N-Terminal Methionine Excision from *Escherichia coli* Proteins is Governed by the Side Chain Length of the Penultimate Amino Acid" *Proc. Natl. Acad. Sci. USA* **86,** 8247-8251.

Hochuli, E. (1989) "Purification of Recombinant Proteins with Metal Chelate Adsorbent", **12,** 87-90, In *"Genetic Engineering, Principle and Methods"* (J.K.Setlow, ed.), Plenum Press, New York.

Hoglund, S., Ohagen, A., Lawrence, K., and Gabuzda, D. (1994) "Role of Vif During Packing of the Core of HIV-1", *Virology* **201,** 349-355.

Hoofnagle, J.H. and Di Bisceglie, A.M. (1991) "Serologic Diagnosis of Acute and Chronic Viral Hepatitis", *Seminars in Liver Disease* **11(2),** 73-83.

Hosein, B., Fang, C.T., Popovsky, M.A., Ye, J., Zhang, M., and Wang, C.Y. (1991) "Improved Serodiagnosis of Hepatitis C Infection with Synthetic Peptide Antigen from Capsid Protein", *Proc. Natl. Acad. Sci. USA* **88,** 3647-3652.

Hsu, H.H., Donets, M., Greenberg, H.B., and Feinstone, S.M. (1993) "Characterization of Hepatitis C Virus Structural Proteins with a Recombinant Baculovirus Expression System", *Hepatology* **17(5),** 763-771.

Hussain, K.H., Issel, C.J., Rwambo, P.M., Arnizaunt, A.B., Ball, J.M., Schnorr, K.L., and Montelaro, R.C. (1988) "Identification of Gag Precursor of Equine Infectious Anaemia Virus with Monoclonal Antibodies to the Major Viral Core Protein, p26", *Journal of General Virology* **69,** 1719-1724.

Imberti, L., Cariani, E., Bettinardi, A., Zonoro, A., Albertini, A., and Prima, D. (1991) "An Immunoassay for Specific Amplified HCV Sequences", *J. Virol. Methods* **34,** 233-243.

Issel, C.J. and Coggins, L. (1979) "Equine Infectious Anaemia: Current knowledge", *Journal of the American Veterinary Association* **174,** 727-733.

Kato, N., Ootsuyama, Y., Ohkoshi, S., Nakazawa, T., Sekiya, H., Hijikata, M., and Shimotohno, K. (1992) "Characterization of Hypervariable Regions of the Putative Envelope Protein of Hepatitis C Virus", *Biochem. Biophys. Res. Comm.* **189(1),** 119-127.

Kawakami, T., Sherman, L., Dahlbert, J., Gazit, A., Yanov, A., Tronic, S.R., and Aaronson, S.A. (1977) "Nucleotide Sequence of Equine Infectious Anaemia Proviral DNA", *Virology* **158,** 300-312.

Khorana, H.G. (1979) "Total Synthesis of a Gene", *Science* **203,** 614-625.

Khudyakov, Y.E., Fields, H.A., Favorov, M.O., Khudyakova, N.S., Bonafonte, M.-T., and Holloway, B. (1993) "Synthetic Gene for the Hepatitis C Virus Nucleocapsid Protein", *Nuc. Acids. Res.* **21,** 2747-2754.

Kumar, U., Cheng, D., Thomas, H., and Monjardino, J. (1992) "Cloning and Sequencing of the Structural Region and Expression of Putative Core Gene of Hepatitis C Virus from a British Case of Chronic Sporadic Hepatitis", *J. Gen. Virol.* **73,** 1521-1525.

Kuo, G., Choo, Q.-L., Alter, H.J., Gitnick, G.L., Redeker, A.G., Purcell, R.H., Miyamura, T., Dienstag, J.L., Alter, M.J., Stevens, C.E., Tegtmeier, G.E., Bonino, F., Colombo, F., Lee, W.S., Kuo, C., Berger, K., Shuster, J.R., Overby, L.R., Bradley, D.W., and Houghton, M. (1989) "An Assay for Circulating Antibodies to a Major Etiologic Virus of Human Non-A, Non-B Hepatitis", *Science* **244,** 362-364.

Lehrer, S.S. (1971) "Solute Perturbation of Protein Fluorescence. The Quenching of the Tryptophyl Fluorescence of Model Compounds and of Lysozyme by Iodide Ion", *Biochemistry* **10,** 3254-3263.

Letvin, N.L. (1990) "HIV and the Immune System: Animal Models for AIDS", *Immunology Today* **11,** 322-326.

Levy, J.A. (1993) "Pathogenesis of Human Immunodeficiency Virus Infection", *Microbiological Reviews 57,* 183-289.

Lo, S.-Y., Selby, M., Tong, M., and Ou, J.-H. (1994) "Comparative Studies of the Core Gene Products of Two Different Hepatitis C Viral Isolates: Two Alternative Forms Determined by a Single Amino Acid Substitution", *Virology* **199,** 124-131.

Ma, X.Y., Sova, P., Chao, W., and Volsky, D.J. (1994) "Cysteine Residues in the Vif Protein of Human Immunodeficiency Virus Type 1 are Essential for Viral Infectivity", *J. Virol.* **68,** 1714-1720.

Machida, A., Ohnuma, H., Tsuda, F., Munekata, E., Tanaka, T., Akahone, Y., Okamoto, H., and Mishiro, S. (1992) "Two Distinct Subtypes of Hepatitis C Virus Defined by Antibodies Directed to the Putative Core Protein", *Hepatology* **16(4),** 886-891.

Martarano, L., Stephens, R., Rice, N., and Derse, D. (1994) "Equine Infectious Anemia Virus *trans*-Regulatory Protein Rev Controls Viral mRNA Stability, Accumulation, and Alternative Splicing", *J. Virol.* **68,** 3102-3111.

Matsuura, Y., Harada, S., Suzuki, R., Watanabe, Y., Inoue, Y., Saito, I., and Miyamura, T. (1992) "Expression of Processed Envelope Protein of Hepatitis C Virus in Mammalian and Insect Cells", *Journal of Virology* **66(3),** 1425-1431.

McFarlane, I.G., Smith, H.M., Johnson, P.J., Bray, G.P., Vergani, D., and Williams, R. (1990) "Hepatitis C Virus Antibodies in Chronic Active Hepatitis: Pathogenic Factor or False Positive Result?" , *Lancet* **335,** 754-757.

McGuire, T.C., Crawford, T.B., and Henson, J.B. (1971) "Immunofluorescent Localization of Equine Infectious Anemia Virus in Tissue", *Am. J. Pathol.* **62,** 283-294.

McGuire, T.C., O'Rourke, K.I., Baszler, T.V., Leib, S.R., Brassfield, A.L., and Davis, W.C. (1994), "Expression of Functional Protease and Subviral Particles by Vaccinia Virus Containing Equine Infectious Anaemia Virus *gag* and 5' *pol* Genes", *J. Gen. Virol.* **75,** 895-900.

Meric, C., Gouilloud, E., and Spahr, P.-F. (1988) "Mutations in Rous Sarcoma Virus Nucleocapsid Protein p12 (NC): Deletions of Cys-His Boxes", *J. Virol.* **62,** 3328-3333.

Meric, C. and Spahr, P.-F. (1986) "Rous Sarcoma Virus Nucleic Acid-Binding Protein p12 is necessary for Viral 70s RNA Dimer Formation and Packaging", *J. Virol.* **60,** 450-459.

Milich, D.R., McLachlan, A., Thornton, G.B., and Hughes, J.L. (1987) Antibody Production to the Nucleocapsid and Envelope of the Hepatitis B Virus Primed by a Single Synthetic T Cell Site", *Nature* **329,** 547-549.

Miller, R.H. and Purcell, R.H. (1990) Hepatitis C Virus Shares Amino Acid Sequence Similarity with Pestiviruses and Flaviviruses as well as Members of Two Plant Virus Supergroups", *Proc. Natl. Acad. Sci. USA* **87,** 2057-2961.

Montagnier, L., Dauguet, C., Axler, S., Chambaret S., Gruest, J., Nugeyre, M.T., Rey, F., Barre-Sinoussi, F., and Chermann, J.C. (1984) "A New Type of Retrovirus Isolated from Patients Presenting with Lymphadenopathy and Aquired Immunodeficieny Syndrome: Structural and Antigenic Relatedness with Equine Infectious Anaemia Virus", *Annals Virol. (Institute Pasteur)* **135E,** 119-134.

Nasoff, M.S., Zebedee, S.L., Inchauspe, G., and Prince, A.M. (1991) "Identification of an Immunodominant Epitope Within the Capsid Protin of Hepatitis C Virus", *Proc. Natl. Acad. Sci. USA* **88,** 5462-5466.

Nishimura, M. and Nakajima, H. (1984) "Structural Proteins of Equine Infectious Anaemia Virus and their Antigenic Activity", *American Journal of Veterinary Research* **45,** 5-9.

Nishioka, K., Watanabe, J., Furuta, S., Tanaka. E.,Iino, M., Suzuki, H., Tsuji, T., Yano, M., Kuo, G., Choo, Q.-L., Houghton, M., and Oda, T. (1991) "A High Prevalence of Antibody to the Hepatitis C Virus in Patients with Hepatocellular Carcinoma in Japan", *Cancer* **67,** 429-433.

O'Farrell, P.H. (1975) "High Resolution Two-Dimensional Electrophoresis of Proteins", *J. Biol. Chem.* **250,** 4007-4021.

O'Rourke, K., Perryman, L.E., and McGuire, T.C. (1988) "Antiviral, Anti-Glycoprotein and Neutralizing Antibodies in Foal with Equine Infectious Anaemia", *Journal of General Virology* **69,** 667-674.

Ohkoshi, S., Kojima, H., Tawaraya, H., Miyajima, T., Kamimura, T., Asakura, H., Satoh, A., Hirose, S., Hijikata, M., Kato, N., and Shimotohno, K. (1990) "Prevalence of Antibody Against Non-A, Non-B Hepatitis Virus in Japanese Patients with Hepatocelluluar Carcinoma", *Jpn. J. Cancer Res.* **81,** 550-553.

Okada, S., Akahane, Y., Suzuki, H., Okamoto, H., and Mishiro, S. (1992) "The Degree of Variability in the Amino Terminal Region of the E2/NS1 Protein of Hepatitis C Virus Correlates with Responsiveness to Interferon Therapy in Viremic Patients", *Hepatology* **16(3),** 619-624.

Osbourne, S., Cecconato, E., Griva, S., Garetto, F., Calogero, R., Rosa, C., and Bonelli, F. (1993) "Expression in *E.coil* and Purification of a Chimeric p22-NS3 Recombinant Antigen of Hepatitis C Virus (HCV)", *FEBS* **324(3),** 253-257.

Patarca, R. and Haseltine, W.A. (1985) "A Major Retroviral Core Protein Related to EPA and TIMP", *Nature* **318,** 390.

Penny, D. (1988) "Origin of the AIDS Virus", *Nature (London)* **333,** 494-495.

Prongay, A., Smith, T.J., Rossmann, M.G., Ehrlich, L.S., Carter, C.A., and McClure, J. (1990) "Preparation and Crystallization of a Human Immunodeficiency Virus p24-Fab Complex" *Proc. Natl. Acad. Sci. USA* **87,** 9980-9984.

Puoti, M., Zonara, A., Ravaggi, A., Grazia Marin, M., Castelnuovo, F., and Carini, E. (1992) "Hepatitis C Virus RNA and Antibody Response in the Clinical Course of Acute Hepatitis C Infection", *Hepatology* **16(4),** 877-881.

Resh, M.D. (1989) "Specific and Saturable Binding of pp60v-src to Plasma Membrane: Evidence for a Myristyl-src Receptor", *Cell* **58,** 281-286.

Rice, N.R., Lequarre, A.-S., Casey, J.W., Lahn, S., Stephens, R.M., and Edwards, J. (1989) "Viral DNA in Horses Infected with Equine Infectious Anemia Virus", *J. Virol.* **63,** 5194-5200.

Rose, S., Hensley, P., O'Shannessy, D.J., Culp, J., Debouck, C., and Chaiken, I. (1992) "Characterization of HIV-1 p24 Self-Association Using Analytical Affinity Chromatography", *Proteins* **13,** 112-119.

Rossi, J.J., Kierzek, R., Huang, T., Walker, P.A., and Itakura K. (1982) "An Alternate Method for Synthesis of Double-Stranded DNA Segments", *J.Biol.Chem.* **257,** 9226-9229.

Rossmann, M.G. (1988) "Antiviral Agents Targeted to Interact with Viral Capsid Proteins and a Possible Application to Human Immunodeficiency Virus", *Proc. Natl. Acad. Sci. USA* **85,** 4625-4627.

Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, Y., Koi, S., Onj, M., Ohta, Y., Choo, Q.-L., Houghton, M., and Kuo, G. (1990) "Hepatitis C Virus Infection is Associated with the Development of Hepatocellular Carcinoma", *Proc. Natl. Acad. Sci. USA* **87,** 6547-6549.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) "Molecular Cloning: A laboratory manual, 2nd Edition", *Cold Spring Harbor Laboratory Press*.

Santolini, E., Migliaccio, G., and La Monica, N. (1994) "Biosynthesis and Biochemical Properties of the Hepatitis C Virus Core Protein", *J. Virol.* **68,** 3631-3641.

Scarpulla, R.C., Narang, S.A., and Wu, R. "Use of a New Retrieving Adaptor in the Cloning of a Synthetic Human Insulin A-Chain Gene", *Anal. Biochem.* **121,** 356-365.

Sellon, D.C., Perry, S.T., Coggins, L., and Fuller, F.J. (1992) "Wild-Type Equine Infectious Anemia Virus Replicates in vivo Predominantly in Tissue Macrophages, Not in Peripheral Blood Monocytes", *J. Virol.* **66,** 5906-5913.

Simmonds, P., Zhang, L.Q., Watson, H.G., Rebus, S., Ferguson, E.D., Balfe, P., Leadbetter, G.H., Yap, P.L., Peutherer, J.F., and Ludlam, C. (1990) "Hepatitis C Quantification and Sequencing in Blood Products, Haemophiliacs and Drug Users", *Lancet* **336,** 1469-1472.

Smith, T.J., Kremer, M.J., Luo, M., Vriend, G., Arnold, E., Kramer, G., Rossmann, M.G., McGinlay, M.A., Diana, G.D., and Otto, M.J. (1986) "The Site of Attachment in Human Rhinovirus 14 for Antiviral Agents that Inhibit Uncoating", *Science* **233,** 1286-1293.

Spearman, P., Wang, J.-J., Van Der Heyden, N., and Ratner, L. (1994) "Identification of Human Immunodeficiency Virus Type 1 Gag Protein Domains Essential to Membrane Binding and Particle Assembly", *J. Virol.* **68,** 3232-3242.

Stahl, S., Mackay, P., Magazin, M., Bruce, B., and Murray, K. (1982) "Hepatitis B Viral Core Antigen : Synthesis in *E.coli* and Application in Diagnostics", *Proc. Natl. Acad. Sci. USA* **79,** 1606-1610.

Steinman, R., Dombrowski, J., O'Conner, T., Montelaro, R.C., Tonelli, Q., Lawrence, K., Seymor, C., Goodness, J., Pedeson, N., and Anderson, P.R. (1990) "Biochemical and Immunological Characterization of the Major Structural Proteins of Feline Immunodeficiency Virus", *J. Gen. Virol.* **71,** 701-706.

Stephens, R.M., Casey, J.W., and Rice, N.R. (1986) "Equine Infectious Anaemia Virus Gag and Pol Genes: Relatedness to Visna and AIDS virus", *Science* **231,** 589-594.

Strambio-De-Castilla, C. and Hunter, E. (1992) "Mutational Analysis of the Major Homology Region of Mason-Pfizer Monkey Virus by use of Saturation Mutagenesis", *J.Virol.* **66,** 7021-7032.

Takahashi, K., Kishimoto, S., Yoshizawa, H., Okamoto, H., Yoshikawa, A., and Mishiro, S. (1992) "p26 Protein and 33-nm Particle Associated with Nucleocapsid of Hepatitis C Virus Recovered from the Circulation of Infected Hosts" *Virology* **191,** 431-434.

Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I., and Okayama, H. (1991) "Structure and Organization of the Hepatitis C Virus Genome Isolated from Human Carriers", *J.Virol.* **65,** 1105-1113.

Takeuchi, K., Kubo, Y., Boonmar, S., Watanabe, T., Katayama, T., Choo, Q.-L., Kuo, G., Houghton, M., Saito, I., and Miyamura, T. (1990) "The Putative Nucleocapsid and Envelope Genes of Hepatitis C Virus Determined by Comparison of the Nucleotide Sequence of Two Isolates Derived from Isolates Derived from an Experimentally Infected Chimpanzee and Healthy Human Carriers", *J. Gen. Virol.* **71,** 3027-3033.

Tobias, J.W., Schrader, T.E., Rocap, G., and Varshavsky, A. (1991) "The N-End Rule in Bacteria", *Science* **254,** 1374-1377.

Tomei, L., Failla C., Santolini, E., De Francesco, R., and La Monica, N. (1993) "NS3 Is a Serine Protease Required for Processing of Hepatitis C Polyprotein", *Journal of Virology* **67(7),** 4017-4026.

Trono, D., Feinberg, M.B., and Baltimore, D. (1989) "HIV-1 Gag Mutants can Dominantly Interfere with the Replication of the Wild-Type Virus", *Cell* **59,** 113-120.

Tsubota, A., Chayama, K., Arase, Y., Koida, I., Saitoh, S., Ikeda, K., Iwasaki, S., Matsumota, T., Kobayashi, M., and Kumada, H. (1993) "Factors Useful in Predicting the Response to Interferon Therapy in Chronic Hepatitis C", *J. Gastroenterol. Hepatol.* **8(6),** 535-539.

Van der Poel, C.L., Cuypers, H.T.M., Reesink, H.W., Weiner, A.J., Quan, S., di Nello, R., van Boven, J.J.P., Winkel, I., Mulder-Folkerts, D., Exel-Oehlers, P.J., Schaasberg, W., Leentvaar-Kuypers, A., Polito, A., Houghton, M., and Lelie, P.N. (1991) "Confirmation of Hepatitis C Virus Infection by New Four-Antigen Recombinant Immunoblot Assay", *Lancet* **337,** 317-319.

Wills, J.W. (1989) "Retro-Secretion of Recombinant Proteins", *Nature (London)* **340,** 323-324.

Wills, J.W. and Craven, R.C. (1991) "Form, Function and Use of Retroviral Gag Proteins", *AIDS* **5,** 639-654.

Zheng, J., Schodel, F., and Peterson, D.L. (1992) "Structure of Hepadnaviral Core Antigens", *J. Biol. Chem.* **267,** 9422-9429.

VITA