Characterisation of the Hepatitis C Virus NS3 Serine Protease

by

Russell John Thompson

A thesis presented for the degree of Doctor of Philosophy

in

The Faculty of Science, University of Glasgow

Division of Virology

Church Street

Glasgow

G11 5JR

June 1998

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Acknowledgements

I would like to thank Prof. J. H. Subak-Sharpe for letting me undertake my PhD here in Glasgow and to GlaxoWellcome for funding my studentship.

A great deal of thanks and gratitude must go to Prof. Richard Elliott for taking me on as a PhD student, helping and advising me with my experimental work as a PhD student, steering me in the correct direction as a PhD student, for his input as I was writing this thesis as a PhD student and now and then getting me drunk, as a PhD student.

Thanks to Berwyn Clarke and everyone else at GlaxoWellcome who gave me help and input throughout my work here and to Dr. Rick Randall for the monoclonal antibody and epitope tag.

A whole heap of merci beaucoups must go to all the busy workers who were the G18 posse, but a special mention must go to Marion (for her maternal influence) and Gill (for being under the influence - only kidding!). An extra special "TA" to Ewan "Sempai" Dunn for help with my Marathon problems, saving me from my Doom and his unfailing belief that "science is cool!".

Throughout my three years I have had two great friends in Ricky and Ross who I owe a great debt of gratitude, Ricky for being a good, steadying influence (!) and Ross for letting me know that no matter how bad it gets it could always be worse. I must also thank Ross for being a quite, conscientious flatmate (except for during Rangers matches on TV), a great friend and potentially a disaster as a best man (whenever I have the need of one).

For obvious reasons I must thank Dad, Mum and Mandy for supplying me with love and a stable home life for 21 years, ensuring that I would be the well adjusted scientist I am today.

Last, but no means least, I owe a tremendous amount to Katie for her love and support over the last two years, for her adoration of Barry Manilow (!) and for basically making me happy.

Summary

Hepatitis C Virus (HCV) is the major cause of post-transfusion non-A. non-B hepatitis. HCV is a member of the *Flaviviridae*, a family of viruses characterised by a single-stranded, positive sense RNA genome. The genome of HCV is approximately 9.6kb and contains a large open reading frame (ORF) that encodes the viral polyprotein.

The viral polyprotein contains the structural proteins (Core, E1 and E2) at the N-terminus and the non-structural proteins (NS2-NS5B) within the remainder of the polyprotein. The HCV polyprotein is processed by cellular and viral-encoded proteases. A serine protease, contained within the NS3 protein, is responsible for the majority of cleavage events within the non-structural region of the polyprotein.

A cell-based assay to study the cleavage of an HCV substrate *in trans* by the NS3 serine protease was developed, and using this assay characterisation of the substrate specificity of the NS3 serine protease was carried out. HCV proteins were expressed from plasmid DNAs in vTF7.3 infected mammalian cells and the cells lysed and immuneprecipitated before analysis by SDS-PAGE. A truncated NS5 substrate was constructed with an epitope tag at both termini to enable immuneprecipitation of the substrate and cleavage products.

The consensus amino acid motif for cleavage by NS3 had been previously determined (D/E X X X X C/T S/A) and mutagenesis within an HCV NS5 substrate confirmed the importance of these residues (P6, P1, P1'). The importance of the P6 residue appeared to be more important for cleavage at the NS5A/NS5B cleavage site than at other cleavage sites within the HCV polyprotein.

The amino acid sequence of the HCV Gla polyprotein was searched for the consensus NS3 cleavage site motif. In addition to the sites cleaved by NS3 an additional three potential sites were found. Analysis of one of these sites demonstrated that an acidic residue at P5 was important for efficient cleavage in trans by the NS3 serine protease. In addition to this the presence of an additional serine residue in close proximity to the P1 and P1' residues was observed to have a negative effect on the efficiency of processing.

The protease requirement for processing of an HCV NS5 substrate was investigated. The NS5A/NS5B cleavage site was processed when the NS5 substrates were coexpressed with NS34A. Coexpression with NS3 and NS3P did not result in the generation of cleavage products. When specific amino acids were substituted with residues found within other HCV isolates the NS3 alone was able to process *in trans*. These investigations illustrated that additional residues to the catalytic triad within the NS3 protease are important for cleavage *in trans*, at least for the NS3 protein alone.

An NS3 serine protease from a different HCV subtype was shown to be able to process the HCV Gla NS5 substrate.

Contents

Amino Acids and their Symbols

Amino Acid	Three letter code	Single letter code
Alanine	Ala	A
Cysteine	Cys	C
Aspartic Acid	Asp	D
Glutamic Acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

Abbreviations

Alanine Transaminase	ALT
Bovine Viral Diarrhea Virus	BVDV
Calf Intestinal Phosphatase	CIP
Complementary DNA	cDNA
Cytotoxic T Lymphocyte	CTL
Deionised Water	dH_2O
Deoxyribonucleic Acid	DNA
Dulbecco's Modified Eagles Medium	DMEM
Endoplasmic Reticulum	ER
Enzyme Substrate Complex	ES
Foetal Calf Serum	FCS
Glycoprotein	gp
Hepatitis A Virus	HAV
Hepatitis B Virus	HBV
Hepatitis C Virus	HCV
Hepatitis D Virus	HDV
Hepatitis E Virus	HEV
Hepatitis G Virus	HGV or GBV-C
Hepatocellular Carcinoma	HCC
Human Immunodeficiency Virus	HIV

Hydrochloric Acid	HCl
Hypervariable Region	HVR
Immunoprecipitation	IP
Interferon	IFN
Interferon Sensitivity Determining Region	ISDR
Internal Ribosome Entry Site	IRES
Kilobases	kb
Long Stable Hairpin	LSH
Long Terminal Repeat	LTR
Lymphotoxin Beta Receptor	LT beta R
Mega Units	MU
Messenger RNA	mRNA
Microcurie	µCi
Microgram	µg
Microlitre	µl
Milligram	mg
Millilitre	ml
Millimolar	mM
Molar	M
Nanogram	ng
Open Reading Frame	ORF
Phosphate Buffered Saline	PBS
Picogram	pg

Polyacrylamide Gel Electrophoresis	PAGE
Polyethylene Glycol	PEG
Polyethylene Imine	PEI
Polymerase Chain Reaction	PCR
Polypyrimidine Tract Binding Protein	PTB
Replicative Intermediate	RI
Replicative Form	RF
Reverse Transcription	RT
Ribonucleic Acid	RNA
RNA Dependent RNA Polymerase	RdRp
Signal Receptor Protein	SRP
Simian Virus 5	SV5
Simian Virus 40	SV40
Sodium Dodecyl Sulphate	SDS
Transfer RNA	tRNA
Terminal Nucleotide Transferase	TNTase
Untranslated Region	UTR

Tables

Figures

Chapter 1 - Introduction

1. Viral Hepatitis.

Viral hepatitis is a major cause of liver disease worldwide. In addition to hepatitis C virus (HCV) there are a number of different and varied viruses that cause hepatitis.

Hepatitis A virus (HAV) is transmitted primarily by the faecal-oral route and is acid and heat stable. It was originally classified as Enterovirus 72 but has been classified as an Hepativirus. Unlike other picornaviruses it has a significantly different nucleic acid and amino acid sequence, it adapts with difficulty to growth in cell culture and has only one serotype. It is a non-enveloped particle of approximately 27-32nm in diameter. The HAV capsid contains a linear, single-stranded, positive sense RNA genome (~7.5kb) and multiple copies of 3 or 4 viral proteins. The RNA genome is uncapped and contains a 5′ non-coding region (NCR) constituting 10% of the genome that is covalently linked to the viral protein VPg. There is a single open reading frame encoding all the viral proteins followed by a short 3′ NCR. Both the 5′ and 3′NCRs are believed to play roles in viral replication (Hollinger and Ticehurst, 1996).

Hepatitis B virus (HBV) is transmitted by sexual contact, blood to blood contact and from mother to child (parenteral transmission). The primary infection may be asymptomatic or can exhibit varying degrees of acute hepatitis. Five percent of these cases lead to chronic infection with varying symptoms. The HBV virion is a complex, double shelled particle known as the Dane particle. The outer shell is a lipid containing envelope of approximately 7nm containing the viral surface antigen (HBsAg). The electron-dense inner core (capsid) is 27nm in diameter and contains the viral DNA and polymerase. The HBV genome is a relaxed, circular, partially duplex DNA of 3.2kb. The circularity of the genome is maintained by 5′ cohesive ends. The two DNA strands are not perfectly symmetrical. The negative strand is of unit length and has a protein covalently attached to the 5′ end. The positive strand is less than unit length and is capped by an oligoribonucleotide at the 5′ terminus. There are 4 ORFs contained within the DNA, encoding the viral proteins (Ganem, 1996; Hollinger, 1996).

Hepatitis D virus (HDV) is a subviral agent and requires HBV to provide envelope proteins. The HDV particle is approximately 28-39nm in diameter and contains a spherical nucleocapsid of approximately 19nm in diameter. The outer surface of the capsid is composed of lipids and the three HBV surface antigens. The HDV nucleocapsid contains a 1.7kb single-stranded RNA genome and multiple copies of the only HDV encoded protein, the D antigen. The linear genome has the potential to fold on itself

(~70% can base pair) to form an unbranched rod-like structure and can also act as a ribozyme to carry out self-cleavage. Superinfection with HDV produces a varied but more severe hepatitis than the other hepatitis viruses and fulminant hepatitis is a frequent occurrence. Sixty to seventy percent of cases result in liver cirrhosis and a chronic infection is usually preceded by acute infection. Co-infection with HBV usually results in an acute, self-limiting disease (Taylor, 1996).

Hepatitis E virus (HEV) is defined as epidemic or enteric non-A, non-B hepatitis. Similar to HAV it is a water-borne virus that has been likened to caliciviruses. The virion is a non-enveloped 27-30nm particle of indefinite surface substructure containing a 7.5kb single-stranded, positive-sense RNA genome. A short 5'NCR (~27 nucleotides) is followed by three overlapping ORFs encoding the structural and non-structural proteins. Little is known about HEV replication. Not all HEV infections are apparent. Approximately a third of HEV infections result in jaundice, anorexia and hepatomegaly with a half also exhibiting abdominal pains, nausea and vomiting (Purcell, 1996).

Hepatitis G Virus (HGV or GBV-C) is a member of the *Flaviviridae* family (as is HCV). HGV contains a positive-sense RNA genome containing a single ORF and its organisation is similar to HCV although it only has a 25% homology to HCV at the nucleotide level. HGV is transmissable through blood transfusion and by exposure to blood products or intravenous drug abuse. Infection may result in acute or chronic hepatitis and co-infection with HBV and HCV is common (Thomas *et al.*, 1997).

1.1. History of HCV

The major cause of viral hepatitis associated with blood transfusions was designated non-A, non-B hepatitis (NANBH) until 1989 when Choo *et al.* isolated a cDNA clone and designated it hepatitis C virus (HCV). A cDNA library derived from infectious material was constructed in the bacteriophage λgt11. This vector, which allows efficient expression of cDNA-encoded polypeptides, was used to produce a library that was screened for the presence of viral antigens with serum from a chronic NANBH patient. The library was derived from chimpanzee plasma containing a relatively high infectious titre to increase the likelihood of interaction with viral antibodies from the NANBH-infected serum.

After ultracentrifugation nucleic acid obtained from the pellet believed to contain the virus was denatured and cDNAs produced from both the RNA and DNA present using random primers. Two clones were shown not to be derived from the host genome and could hybridize to total RNA extracted from infectious chimpanzee liver. The initial clone (5-1-1) was significantly smaller than the second (clone 81), which was isolated from the same library and was determined to be larger and overlapped clone 5-1-1. These clones were demonstrated to be derived from an exogenous RNA molecule associated with NANBH infection. As only one strand of the

cDNA clone was able to hybridize to the HCV RNA extracted the genome was believed to be single-stranded. The RNA extracted from the infected chimpanzee was between 5,000 and 10,000 nucleotides in length. This RNA bound to oligo (dT)-cellulose indicating that there may be an A-rich stretch somewhere within the molecule.

The nucleotide sequence of one clone (5-1-1) demonstrated that the cDNA strand that produced the immunoreactive polypeptide contained a single, continuous open reading frame (ORF) and the polyprotein encoded was shown to be closely associated with NANBH infection. The serum used to initially identify the clone also reacted specifically to the polypeptide encoded by this clone. The genome itself was shown to have a positive polarity with respect to the translation of the viral polyprotein. This analysis led to the assumption that HCV was a member of the *Flaviviridae* as the virus contained a positive-stranded RNA molecule of approximately 10,000 nucleotides (Choo *et al.*, 1989).

The first full-length HCV sequence was derived from a chimpanzee infected with a contaminated Factor VIII concentrate (Choo *et al.*, 1991) and this was determined as the HCV prototype HCV-1. The similarity of HCV to both flaviviruses and pestiviruses was seen in the size of the polyproteins encoded by the genome (3010, 3400 and ~4000 respectively). There were also small regions of amino acid homology with the NS3 and NS5 proteins of flaviviruses (Miller and Purcell, 1990; Choo *et al.*, 1991). The virus particle has not been characterised because of poor *in vitro* replication systems. It has been shown that it has a chloroform sensitive outer envelope (Bradley *et al.*, 1985) which is believed to be lipid based (Tremolada *et al.*, 1992). Icosahedral particles have been isolated using anti-core monoclonal antibody-coated grids showing that native core is present as part of the virus particle (Takahashi *et al.*, 1992a). Yuasa *et al.* (1991) determined that HCV particles were between 30-38nm in diameter.

HCV infection can be detected by anti-HCV antibodies, antigens to immune response elements and the detection of HCV RNA by reverse transcriptase (RT) PCR (Nakatsuji *et al.*, 1992). HCV infection is often asymptomatic but leads to a chronic infection in greater than 50% of cases (Kuo *et al.*, 1989). The development of chronicity varies but is seen to occur at a higher level than for HBV. Chronic infection has been linked with liver cirrhosis and hepatocellular carcinoma (Alter *et al.*, 1989; Saito *et al.*, 1990). It is estimated that approximately 500 million people are chronic carriers of HCV (Dhillon and Dusheiko, 1995) illustrating that it is a serious human pathogen requiring study.

1.2. Exposure to HCV.

HCV is a blood-borne virus and is detected in low titres only. A significant degree of HCV infection due to transfusions with contaminated blood was observed, especially before an adequate detection system was established. Patients suffering from renal failure, who rely on dialysis and

eventually a kidney transplant, were also under a high risk of HCV infection. Effective screening for HCV contamination of blood has been invaluable in the reduction of the number of infected individuals. Patients with coagulation disorders, treated with contaminated plasma products, were also at high risk before HCV screening. Intravenous drug users are currently the highest risk group for HCV infection (Watson *et al.*, 1992; Degos, 1994). The use of sterile needles would reduce the spread of disease within this particular group. Healthcare workers are also at risk from needlestick injuries.

The majority (60-70%) of HCV cases are believed to result from parenteral transmission (Bastie *et al.*, 1995), i.e. through blood transfusions with contaminated blood or intravenous drug abuse, although other routes exist. Improved analysis should help to clarify non-parenteral routes of infection. Infection with HCV occurs through interaction between adults (i.e. through exchange of infected blood, needles, etc.). Transmission from mother to unborn child has been documented in a number of cases (Uehara *et al.*, 1993; Resti *et al.*, 1995). The titre of the HCV RNA found in the mother was directly proportional to the risk of transmission (Ohto *et al.*, 1994), and a titre of approximately 10^6/ml or higher was shown to be required for transmission of HCV to the unborn child (Moriya *et al.*, 1995; Ohto *et al.*, 1994). It has been shown that mothers testing seropositive for HCV antibodies, but negative for HCV RNA do not infect the unborn child (Zuccotti *et al.*, 1995). There is some evidence that coinfection with HIV increases the risk of vertical transmission (Zanetti *et al.*, 1995). Infection of children from breast milk is not considered a risk (Moriya *et al.*, 1995; Kage *et al.*, 1997). Although sexual transmission of HCV is possible the risk of infection is extremely low (Healey *et al.*, 1995; Tong *et al.*, 1995; Scotto *et al.*, 1996). Prevention of exposure to HCV can be accomplished through the use of clean blood supplies, improvement of general hygiene (i.e. through the use of sterile needles, extensive hand washing and proper disinfection and sterilisation) and reduction of high HCV-RNA titres in infected mothers.

1.3. HCV infection.

There are three distinct phases throughout the clinical course of HCV infection:-

(i) The acute phase - This phase can last 2-3 years after disease onset and is usually asymptomatic.

(ii) The silent phase - Throughout this period there is a low level of serum transaminases observed.

(iii) The reactivated phase - The serum transaminase/transferase level increases and thereafter fluctuates in a moderately high range until hepatocellular carcinoma (HCC) develops. Consequently all sufferers of HCC have abnormal serum transferase levels.

The acute phase of HCV is asymptomatic. There is an observable difference between HCV cases leading to chronicity. Chronicity is defined by the

persisitence of an abnormal serum amino acid transferase level for more than one year (Kiyosawa *et al.*, 1994). In cases where infection originated from a blood transfusion, 74% of patients developed a chronic infection. In cases of sporadic infection (i.e. cases of infection not resulting from infected blood) 55% of patients degenerated to chronicity. The likelihood of chronic infection may depend on the size of the infectious inoculum at the initial infection.

In chronic carriers there is always a detectable amount of HCV RNA. A "first generation antibody test" for HCV, anti-C100, can be used as a diagnostic tool but the protein it identifies has been observed to decline despite persistence of HCV RNA. This antibody reacts with an epitope located in the NS4 region (Watson *et al.*, 1992). A second generation antibody test, anti-HCV2, detects a protein that is always present (Kiyosawa *et al.*, 1994) and reaction with this indicates an HCV infection. The relatively low titre of HCV in the large percentage of patients does not permit efficient detection of HCV antigens (Takahashi *et al.*, 1992b) and a more conclusive detection method is required. The detection of HCV infection by identifying the HCV RNA genome is a more efficient system. Targetting the highly conserved untranslated region at the 5' end of the RNA genome (5'UTR), primers conserved between different HCV genotypes against this region are used in a combined reverse transcription and polymerase chain reaction (RT-PCR) to get a highly sensitive HCV detection assay (Weiner et al., 1990).

The progression to chronicity is unclear, probably due to the complexity of the factors involved. The likelihood of developing chronic HCV infection may depend on the size of the infectious inoculum at the initial infection, the route of infection or the HCV genotype (Yoshioka *et al.*, 1992). There are at least six genotypes of HCV and theoretically each could evolve within an individual, different genes evolving at different rates (Kiyosawa *et al.*, 1994). Cases have been documented where a patient has been infected with two genotypes and infection with a higher number of genotypes is therefore theoretically possible.

1.4. Treatment.

Treatment of HCV infection is with interferon-α, which acts as an antiviral and immunomodulating agent. In the case of HCV it is observed to reduce levels of alanine transaminase (ALT) in some chronic HCV sufferers. After 6 months of thrice-weekly treatment of 3-6 mega-units (MU) of interferon-α there is a normalisation of ALT levels in approximately 50% of those treated. Unfortunately only 25% of these cases retain normal ALT levels once therapy has been stopped. The use of HCV RT-PCR corroborates these results. It should be noted that the absence of HCV RNA does not necessarily mean that there is no HCV infection. Virus may persist in the liver, extrahepatic sites or in plasma in levels undetectable by RT-PCR (reviewed by van der Poel *et al.*, 1994).

Although the use of interferon-α does have side effects these are resolved after therapy cessation and are partly due to the dose level. Interferon treatment in the early/acute phase has been postulated to reduce the likelihood of chronic infection.

Ribavirin, a guanosine analogue administered orally, is observed to normalise ALT levels while not eliminating HCV RNA. Treatment termination results in a return to abnormal levels of ALT (Di Bisceglie *et al.*, 1992). Combination therapy of ribavirin and interferon-α induced a sustained response in patients who initially responded to interferon treatment but relapsed after cessation of treatment. Normal levels of ALT were observed in 75% of these patients after combination therapy, with no HCV RNA detection after 6 months (Brillanti *et al.*, 1994, 1995). Combination therapy of interferon-α and ribavirin increases the sustained response two-fold in comparison with interferon-α treatment alone (Brillanti *et al.*, 1995; Bellobuono *et al.*, 1997).

Because of the problems with vaccine development for HCV (1.5.) the most effective and realistic treatment for HCV is antivirals. In addition to interferon-α and ribavirin treatment other therapies have been investigated. A combination therapy of inteferon with ofloxacin (Takada *et al.*, 1996) or thymosin alpha (Rasi *et al.*, 1996), as well as amantidine treatment (Smith, 1997), have been tested although these investigations have shown only limited success.

Specific antivirals against proteins believed to be essential for virus replication (such as the protease or the polymerase) are most likely to be of use in treatment of HCV and prevention of disease. A detailed knowledge of these proteins is vital for development of an efficient antiviral against HCV.

1.5. Prevention.

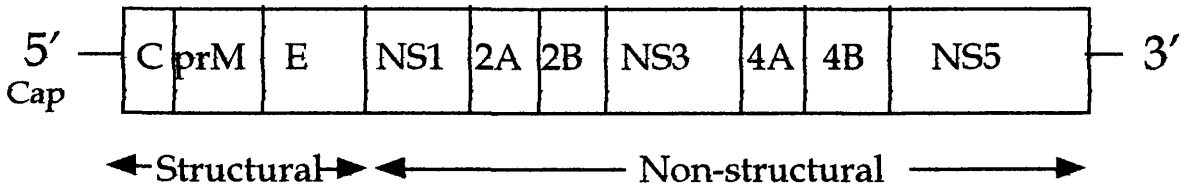
Without any effective therapy against HCV infection a vaccine against HCV would be invaluable. Development of a vaccine is complicated by the fact that re-infection with the same or different isolates of HCV results in a second incidence of disease (Farci *et al.*, 1991). The presence of different genotypes of HCV and the evolution of quasispecies within an infected individual has made the development of a "universal" vaccine problematic. Original vaccines took the form of recombinant envelope vaccines that were only able to combat low titre challenge with HCV in chimpanzees. It has been speculated that use of the highly conserved core protein in a vaccine may be useful in increasing efficacy (Hitomi *et al.*, 1995). Tokushige *et al.* (1996) immunised BALB/c mice with vectors expressing the highly conserved core gene that resulted in a low-level anti-HCV core humoral immune response. As well as this experiment a myeloma cell line constitutively expressing HCV core protein was inoculated into BALB/c mice, resulting in plasmacytomas. A strong CTL

response was generated in comparison to mice injected with vector DNA. These experiments show that the HCV core protein can be used to generate strong CTL activity and is a promising candidate as an antiviral agent. The use of conserved viral epitopes in a vaccine may result in a less aggressive disease with milder symptoms that is more susceptible to interferon treatment (Koshy and Inchauspe, 1996).

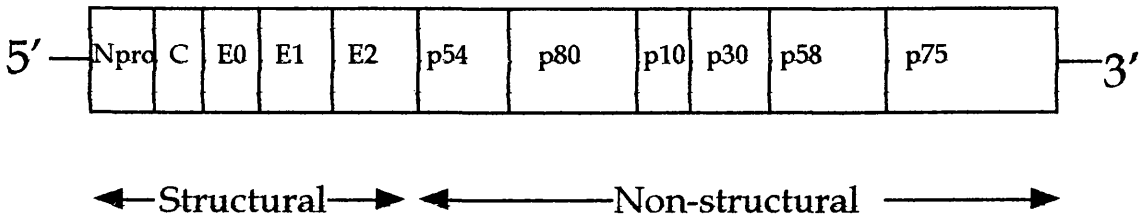
1.6. Genome structure of the *Flaviviridae.*

All members of the *Flaviviridae* i.e. flaviviruses, pestiviruses and HCV, have a positive sense, single-stranded RNA genome. The genomes encode a single open reading frame (ORF) flanked by non-coding regions at the 5′ and 3′ termini. The open reading frame encodes a single polyprotein containing the structural proteins at the N-terminus followed by the non-structural proteins. The basic genomic organisation of the *Flaviviridae* is illustrated in Fig. 1.1. The polyprotein is processed by host and virally encoded proteases both co- and post-translationally to produce the mature viral proteins. Genomes of flaviviruses are approximately 11kb in size, encoding an open reading frame (ORF) of approximately 10kb preceded by a short, capped 5′ untranslated region (UTR) of approximately 95-132 bases in length. The 5′ UTR has a type I cap (m7GpppAmp) followed by the conserved dinucleotide sequence AG and lacks cap-associated and internal base methylated adenine residues (Chambers *et al.*, 1990). The 3′ UTR is significantly longer than the 5′ end and is extensively heterogeneous in size and sequence between different flaviviruses. The bulk of this heterogeneity is found at the proximal 5′ end (Proutski *et al.*, 1997; Wallner *et al.*, 1995) and it is sometimes followed by a polyA stretch. The 3′ end contains approximately 350 nucleotides and exhibits a higher degree of conservation. This section can be divided into 3 regions of distinct secondary structure, the last of which consists of a long, stable hairpin (LSH) structure (Proutski *et al.*, 1997). The 3′ LSH is believed to be important for regulation of replication and is known to bind host cellular proteins thought to be part of the replication complex (Blackwell and Brinton, 1995).

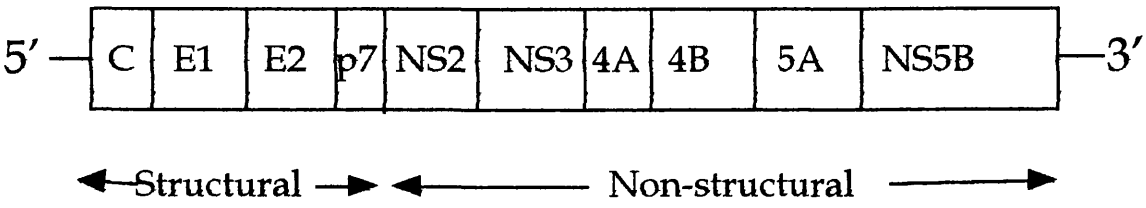
The pestivirus genome is significantly longer than that of the flaviviruses (approximately 12.5kb), with an ORF producing a polyprotein of approximately 4000 amino acids and a 5′ UTR of approximately 385 nucleotides. The 5′ UTR contains an internal ribosome entry site (IRES) directing cap-independent translation (Poole *et al.*, 1995; Rijnbrand *et al.*, 1997). The 3′ UTRs of pestiviruses are between 188 and 228 nucleotides in length and can be divided into two distinct regions - the variable (3′V) and the conserved (3′C) regions (Deng and Brock, 1993). The variable region can be as long as 127 nucleotides and is preceded by the stop codon of the ORF. This region exhibits a high degree of sequence heterogeneity as well as different pestiviruses containing deletions. The 3′V region is not believed to be important in replication. The 3′C region is 102 nucleotides in length, is highly conserved and contains an AU rich stretch and imperfect repeat



Flaviviruses



Pestiviruses



HCV

Fig. 1.1. Genomic organisation of the *Flaviviridae*

sequences with considerable conservation. This region may contain some important signals for viral RNA replication. Between different viruses the position of the repeat stretches and the nucleotide sequence is not as highly conserved. Between different isolates, however, the conservation is very high (Deng and Brock, 1993). The 3'UTR as a whole has considerable secondary structure consisting of four predicted stem loops. Stem loop I is contained within the conserved region and the sequence is conserved between all 5 pestiviruses analysed. Stem loop II uses sequence from both 3'C and 3'V and is relatively well conserved structurally. The two remaining stem loops (III and IV) are both in 3'V and are less conserved in terms of length and structural detail (Deng and Brock., 1993).

The HCV genome is approximately 9.6kb and contains a highly conserved 5' UTR (between different genotypes) of approximately 341 nucleotides in length. Due to the highly conserved nature of the 5' UTR there is very little difference in the nucleotide length within this region. Translation of the HCV polyprotein is directed by an IRES element. Direction of protein translation by an IRES element is also seen in pestiviruses with one significant difference, in that twenty-one nucleotides of the HCV ORF are required for IRES activity (Lu and Wimmer, 1996; Reynolds *et al.*, 1995). Almost the entire 5' UTR constitutes the IRES element with the first 29 nucleotides at the extreme 5' terminus not required (Wang et al., 1995). The remaining region forms a highly conserved secondary structure of stem-loops and a pseudoknot that are essential (Rijnbrand *et al.*, 1995, 1996; Wang *et al.*, 1995). The single exception is the first hairpin structure (Rijnbrand *et al.*, 1996) which is not required. The AUG codon used in translation is located in a stem loop whose sequence is conserved throughout HCV genotypes (Honda *et al.*, 1996). This stem loop, although believed to be unstable, is probably responsible for the direct ribosome entry observed close to the authentic AUG (Reynolds *et al.*, 1996). The 5'UTR is believed to bind polypyrimidine tract binding protein (PTB) utilising 3 distinct regions (Ali and Siddiqui, 1995). A 25 kDa protein has also been demonstrated to bind to the 5' UTR (Fukushi *et al.*, 1997). This protein binds to the α-branch of stem loop II. The primary sequence of the α-branch may act as a *cis*-element for translation initiation. There is no direct evidence of stem-protein interaction and this α-branch alone may not be able to bind cellular factors.

The role of PTB in the HCV IRES is unclear. It may be that other proteins associate with PTB and that this interaction is important. It has been reported that recombinant PTB does not bind to the HCV 5' UTR and that this protein did not stimulate HCV IRES function (Reynolds *et al.*, 1995). Although highly conserved, the 5'UTR does contain some small differences between different genotypes. These differences are thought to explain the observation that IRES elements from different genotypes have varying efficiencies at promoting translation initiation (Buratti *et al.*, 1997; Collier *et al.*, unpublished data). The secondary structure of the 5'UTR of HCV is illustrated in Fig. 1.2.

The ORFs of HCV genomes are between 9027 and 9111 nucleotides in length and hence encode 3009 to 3037 codons (Adams *et al.*, 1997). The open

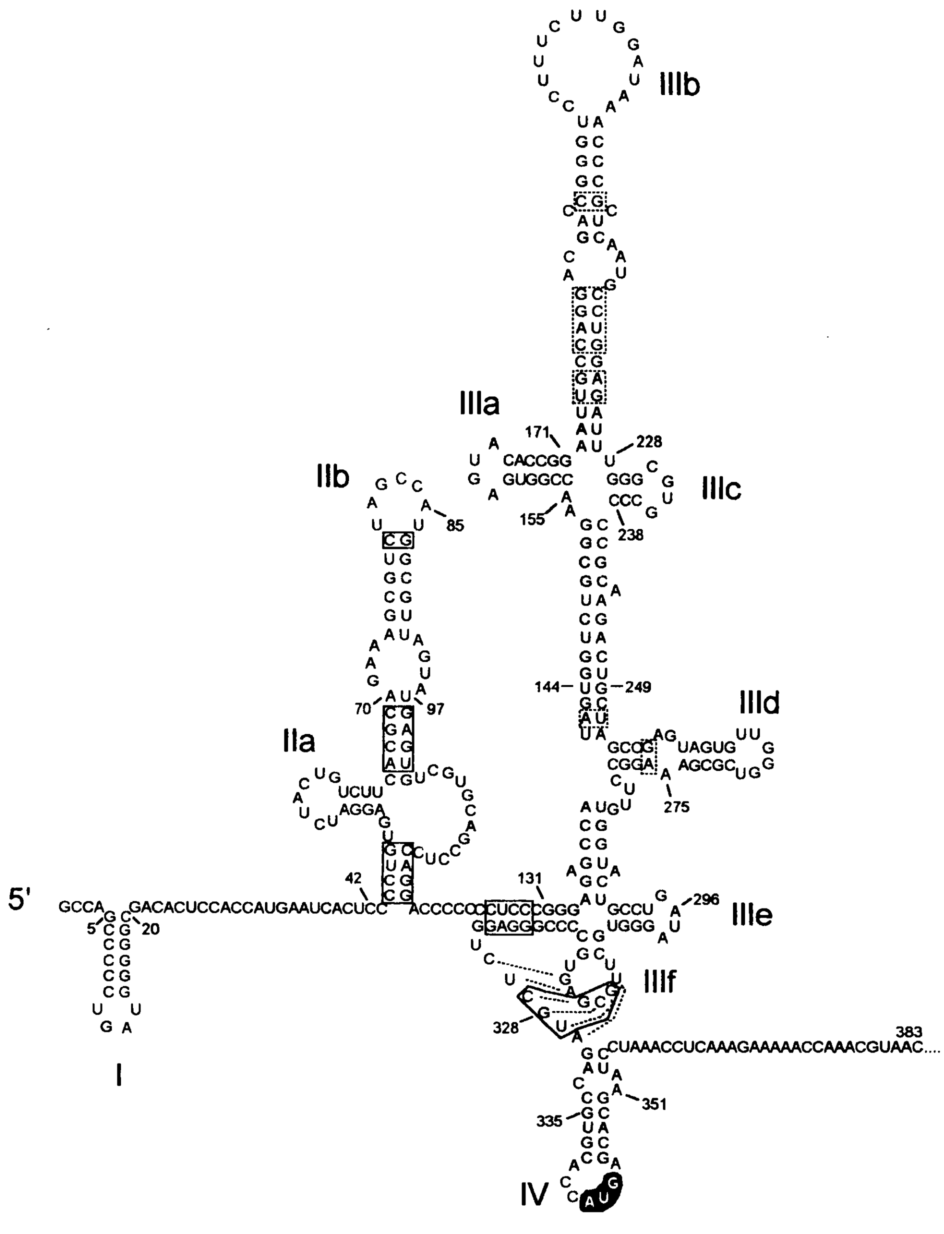


Fig. 1.2. Secondary and tertiary structure within the 5′NCR of HCV (from Lemon and Honda, 1997). Major domains are labeled I-IV from the end, with individual stem-loops assigned letter codes within each domain. Translation initiates at the shaded AUG triplet.

reading frames are all followed by a 3′ UTR consisting of a short variable region, a variable polypyrimidine tract of between 73 and 98 nucleotides in length (Yamada *et al.*, 1996) and a 98 nucleotide sequence denoted the X-tail (Tanaka *et al.*, 1995). The X-tail is highly conserved between different HCV genotypes (Tanaka *et al.*, 1996; Kolykhalov *et al.*, 1996) and is believed to have a high degree of secondary structure (Kolykhalov *et al.*, 1996). The 3′ terminal 46 bases form a stable stem loop structure (stem loop I) with a 6 nucleotide single stranded loop. The remaining 52 nucleotides are believed to fold into more than one secondary structure. The X-tail is probably involved in multiple RNA-protein interactions with stem loop I acting as a recognition site for viral and/or cellular proteins (Blight and Rice, 1997). It has been shown that the X-tail binds the 58-kDa protein PTB. The sequence requirement for this binding is very specific and sensitive to mutation (Ito and Lai, 1997) despite differing from the consensus PTB-binding consensus sequence. The predicted optimal secondary structure of the 3′ UTR is required for PTB binding and involves primary loop sequences and stem structures of SL2 and SL3 (Ito and Lai, 1997). It is speculated that the X-tail may be important in replication of the HCV genome, modulation of translation or RNA stability. The predicted secondary structure of the 3′UTR is shown in Fig. 1.3.

1.7. Heterogeneity of the HCV genome.

All members of the *Flaviviridae* replicate using a virus-encoded RNA polymerase and so exhibit a high degree of sequence variation. Different isolates of HCV show significant differences in sequence in specific regions of the virus genome. Within a single HCV isolate there is a high degree of sequence variation and these variants are called quasispecies. The existence of quasispecies is not unique to HCV. RNA polymerases encoded by RNA viruses lack a proof-reading function. The absence of a proof-reading function leads to a high frequency of mutation after replication. It has been observed in HIV-1 that biological properties vary from clone to clone during the course of infection (Duarte *et al.*, 1994a). The existence of quasispecies has also been observed in influenza A and the human hantaviruses as reviewed by Duarte *et al.* (1994b).

Some regions within HCV, for example the envelope region, are particularly variable leading to different HCV groups having different antigenic and biological properties. The NS5 region also contains a highly variable, type-specific region that reduces the effectiveness of antibody assays against this particular protein, in HCV genotype determination. In contrast other regions are very highly conserved (e.g. the 5′UTR and the X-tail).

Genotype determination has been carried out using regions encoding the core, NS3 and NS5 proteins. Despite some variation in sequence the genotype determined corroborated with that obtained by sequencing the entire genome. These sequencing studies, particularly those in the NS5

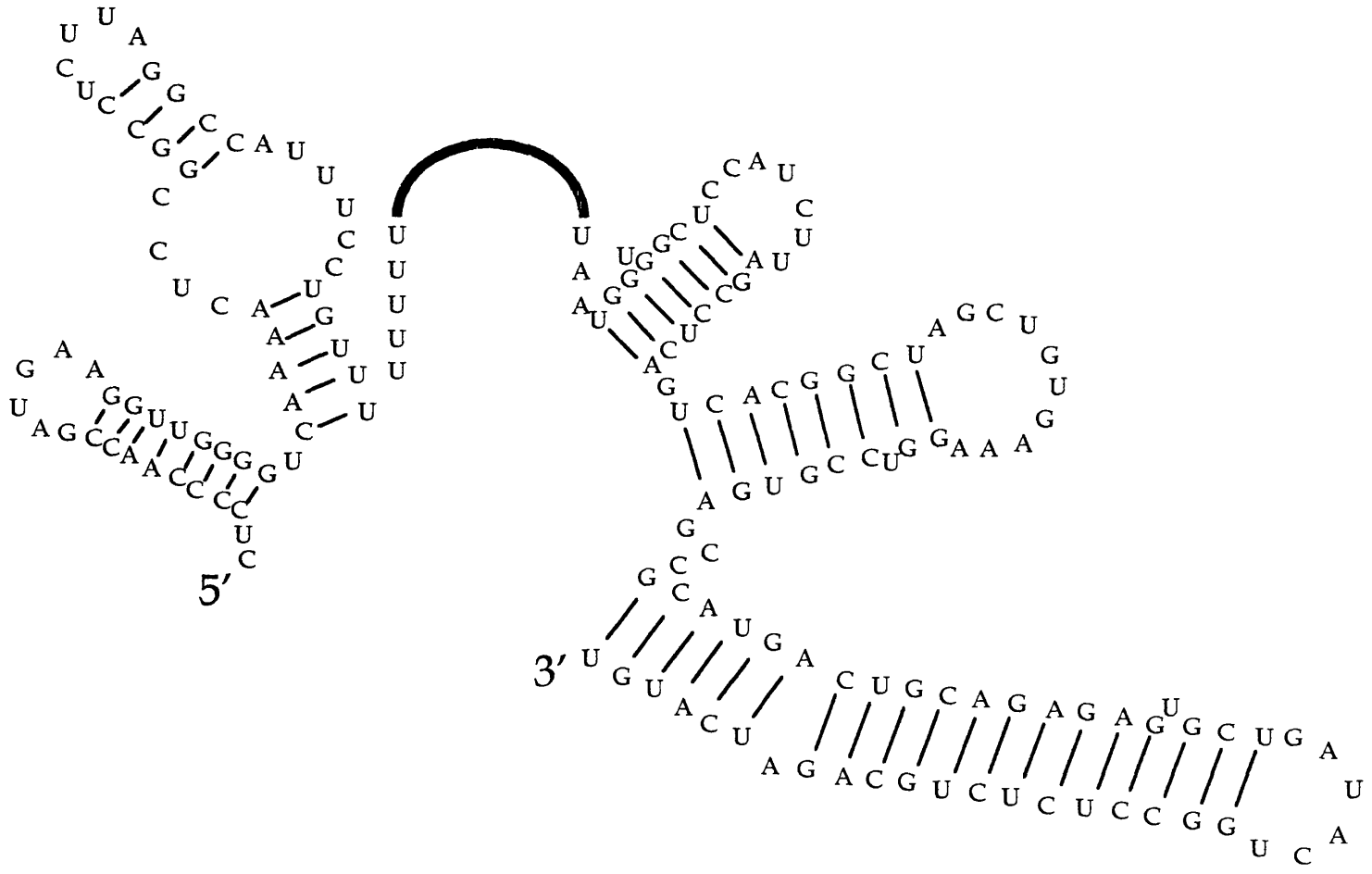


Fig. 1.3. Secondary Structure prediction of the HCV-H 3′ NTR RNA derived from Kolykhalov *et al.*, 1996. The poly (U) tract is represented by the curved black line.

region, have led to the classification of different HCV isolates into a distinctive hierarchy.

The broadest classification for HCV is that of genotypes. There are six major genotypes numbered 1-6 (Simmonds *et al.*, 1993) which are between 66 and 72% similar at the nucleotide level. If an HCV sequence has less than 72% similarity to any of the existing sequences it will be classified as a new HCV genotype. The sequence similarity between different HCV genotypes is illustrated in Table 1.1. (taken from Adams *et al.*, 1997).

In each genotype there are usually two to three clusters of more closely related variants. These groups are called subtypes and are between 75 and 86% similar. For genotype 1 there are three subtypes (1a, 1b, 1c).

A phylogenetic tree can be constructed for different HCV genotypes and is illustrated in Fig. 1.4. The phylogenetic tree can be produced by comparison between the 5'UTR, core, NS3 and NS5 regions of different HCV isolates and classifies them in to types and subtypes.

Individual HCV sequences are classified as isolates. Analysis of HCV genotypes has been carried out using the NS5, core and 5'UTR regions (although the 5'UTR shows very little sequence disimilarity).

A further feature of the heterogeneity of the HCV genome regards the existence of quasispecies within an individual isolate. This feature was observed in individuals infected with one isolate of HCV, yet still showing heterogeneity between the genomes present. The rate of mutation in HCV genomic sequence has been estimated at being between 10^{-3} and 10^{-4} base substitutions per genome per year (Martell *et al.*, 1992). The presence of quasispecies may give HCV an adaptive advantage during infection and may play a role in vaccination failure, establishment of persistent infection, resistance to antiviral agents and changes in virulence.

Fifty-five percent of the infecting HCV genomic sequences have been shown to differ from the consensus sequence of the infecting isolate. Although many of these base changes will have no effect on the encoded amino acid sequence, some will introduce stop codons resulting in defective genomes while others result in amino acid changes that may have an effect (Martell *et al.*, 1992).

1.8. HCV encoded proteins.

HCV, like all members of the *Flaviviridae,* produces a single polyprotein that is subsequently processed into individual proteins. These proteins can be classified into two distinct groups, the structural proteins and the non-structural proteins. HCV produces at least three structural proteins (Core, E1 and E2) and six non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B).

Genotype	n*	Percentage amino acid divergence											
1a	4	—	15.2	14.5	28.1	28.4	27.5	25.7	26.1	25.6	23.0	24.7	24.8
1b	13	21.6	—	15.8	27.6	28.6	27.9	24.9	26.0	25.0	22.8	24.1	24.5
1c	1	20.1	22.1	—	27.6	28.0	27.6	25.2	25.5	24.7	23.0	24.3	24.3
2a	1	31.4	31.0	31.1	—	15.9	13.2	29.2	30.3	28.4	29.3	28.9	29.0
2b	1	32.2	32.2	32.0	22.9	—	13.6	29.4	30.7	29.0	29.4	29.2	29.4
2c	1	30.9	31.5	31.4	19.8	22.4	—	29.9	31.0	29.7	29.4	29.0	29.3
3a	1	31.1	30.9	30.8	32.4	33.4	33.6	—	14.7	17.8	26.7	26.3	27.3
3b	1	31.3	31.2	31.0	33.7	34.0	33.7	21.8	—	18.7	27.8	27.0	27.5
JK049	1	31.1	31.0	30.8	32.9	32.9	33.1	24.8	25.0	—	26.1	25.9	26.5
4a	1	28.4	28.9	28.7	31.9	32.1	32.5	31.5	31.2	31.5	—	25.9	26.1
JK046	1	29.6	29.6	29.8	32.2	32.8	31.8	21.0	32.1	31.3	30.7	—	20.7
EUHK2	1	30.3	30.1	30.4	32.8	32.8	32.9	32.1	32.4	31.9	31.1	27.7	—
		1a	1b	1c	2a	2b	2c	3a	3b	JK049	4a	JK046	EUHK2
		Percentage nucleotide divergence											

* n = number of sequences of same genotype used for comparison. Accession numbers for all sequences compared are the same as those given in ref 15.

Table 1.1. Pairwise comparison using the complete genomic sequence from different genotypes of HCV (from Adams *et al.*, 1997).

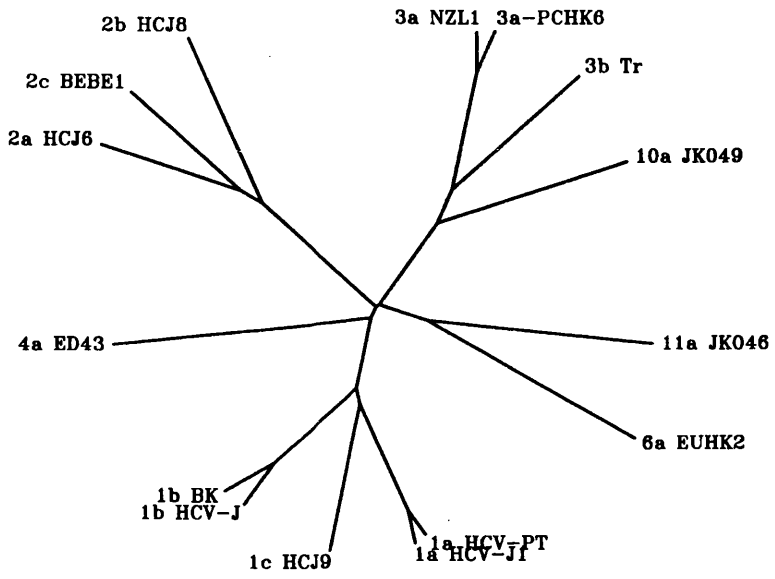


Fig. 1.4. Phylogenetic analysis of the genomic sequences of all major genotypes of HCV (from Adams *et al.*, 1997). The genomic sequences between the 5'NCR and the end of the coding sequence were compared. Sequences were compared using DNADIST and NEIGHBOR in the PHYLIP package and shown as an unrooted tree

1.8.1. Core protein.

The HCV core protein is believed to be composed of the first 191 amino acids of the viral polyprotein (Hijikata *et al.*, 1991b; Ralston *et al.*, 1993). Core is particularly rich in arginine and lysine residues at its N-terminus and is an unglycosylated protein of approximately 21kDa (Hijikata *et al.*, 1991b; Lo *et al.*, 1994; Harada *et al.*, 1991). A second major species of 19kDa (P19) has also been identified (Santolini *et al.*, 1994). Processing to produce P19 occurs at either amino acid 179 or 182 by a eucaryotic signal peptidase. This cleavage event does not require previous processing at the N-terminus of full-length core (Hussy *et al.*, 1996). Besides these two major products a third, minor protein of 151 amino acids is also observed with a molecular mass of 16kDa (P16) (Lo *et al.*, 1995).

The full length core protein is released from the polyprotein through the action of a signal peptidase of the endoplasmic reticulum. Cleavage is mediated by the removal of the putative signal sequence 18 amino acids located at the C-terminus of core. Core protein is normally observed in the cytoplasm of cells, but there is a sequence at the N-terminus of the core protein that targets the protein to the nucleus. The target sequence is only used when the hydrophobic stretch at the carboxy-terminus is removed (Ravaggi *et al.*, 1994; Chang *et al.*, 1994). The amino acid sequence believed to play an important role in nuclear localisation is PRRGPR (Suzuki *et al.*, 1995). The two major species of core (P21 and P19) are associated with the endoplasmic reticulum whereas P16, which lacks the C-terminus of core, is localised in the nucleus (Lo *et al.*, 1995). The different subcellular localisation of these core species suggests that they may have different biological functions.

Multimerisation of core takes place using the amino-terminal hydrophilic region (115 amino acids) with interaction occuring between amino acids 36 and 91 (Matsumoto *et al.*, 1996). This feature of core protein is observed for both membrane-bound and free protein and so must take place at an early stage in viral assembly. Core has been observed to bind membranes and has also been shown to co-localise with the E2 glycoprotein and a cellular ER membrane marker. Besides the ability of core to bind RNA it has also been shown to bind the large ribosomal subunit, utilising the same N-terminal region used for RNA binding. Binding to the ribosome may be mediated by RNA binding (Santolini *et al.*, 1994). The core protein is also phosphorylated by protein kinase A and protein kinase C although phosphorylation is not required for entry into the nucleus (Shih *et al.*, 1995).

Core protein has been shown to activate the Rous sarcoma virus long terminal repeat (LTR), human c-myc and simian virus 40 (SV40) early promoters. Core has also been shown to suppress the c-fos promoter and the human immunodeficiency virus type 1 (HIV-1) LTR (Srinvas *et al.*, 1996). The effect of transcriptional regulation of cellular proto-oncogenes suggests a possible role in deregulation of normal hepatocyte growth and

hepatocarcinogenesis (Ray *et al.*, 1995). The effect of core on the development of carcinogenesis has been further developed. It has been shown that the core gene cotransfected with H-ras can induce rapid proliferation, anchor-independent growth and tumour formation in athymic nude mice (Ray *et al.*, 1996). Core has also been shown to have the ability to inhibit apoptotic cell death (Ray *et al.*, 1996) and suppress the transcriptional activity of the p53 promoter (Ray *et al.*, 1997). Core has been shown to bind the cytoplasmic tail of membrane-bound lymphotoxin-beta receptor (LT beta R) using the same amino acids required for core multimerisation (Matsumoto *et al.*, 1997). LT beta R is a member of the tumour necrosis factor receptor family and is involved in germinal centre formation, developmental regulation of peripheral lymphoid organs, lymph node development and apoptotic signalling. The ability of core to bind LT beta R may have an immunomodulating function possibly important in viral persistence.

1.8.2. The envelope glycoproteins.

E1 (gp35) is a glycosylated protein of approximately 35 kDa. It is has 5 or 6 potential N-linked glycosylation sites (Matsuura and Miyamura, 1993) and after deglycosylation this 192 amino acid protein has a molecular mass of 22kDa. E1 is known to interact through its carboxy-terminus with the HCV core protein (Lo *et al.*, 1996).

After glycosylation at a possible 11 sites the 363 amino acid E2 protein has a molecular mass of 70kDa (Miyamura and Matsuura, 1993) and is known as gp70. E2 is equivalent to the major envelope protein of pestivruses (E2) and NS1 of flaviviruses. Both these proteins confer protective immunity in immunised animals against the respective viral infections. E2 is the most highly variable HCV protein (Kato *et al.*, 1992; Weiner *et al.*, 1991) with hypervariable regions in the N-terminus. Hypervariable region 1 (HVR1), representing the N-terminal 34 amino acids contains B cell epitopes specific to each isolate (Kato *et al.*, 1994). Mutation in this region is believed to be important in the evolution of escape mutants of HCV. Both E1 and E2 are released from the viral polyprotein by host-cell signal peptidases. E2 contains an additional cleavage site 60 amino acids from its C-terminus. Cleavage at this site releases a 7kDa protein called p7 (Lin *et al.*, 1994a; Mizushima *et al.*, 1994; Selby *et al.*, 1994; Lanford *et al.*, 1993; Grakoui *et al.*, 1993c). Processing to give p7 appears to occur post-translationally at a novel, microsomal membrane dependent cleavage site. This cleavage event means that there are two different sized E2 proteins produced by HCV, the different functions of which, if any, are not known. E2 has been observed to interact with NS2 (Matsuura *et al.*, 1994; Selby *et al.*, 1994).

E1 and E2 interact with each other (Dubuisson *et al.*, 1994; Grakoui *et al.*, 1993c; Lanford *et al.*, 1993) usually through non-covalent bonding (Ralston *et al.*, 1993). Disulphide bonding has also been demonstrated

(Grakoui *et al.*, 1993c). It has been shown that the non-covalent complexes form slowly to give a stable heterodimer in stark contrast to the complex formed containing disulphide bonds that are the result of misfolding (Dubuisson *et al.*, 1994). Folding of the individual proteins through the formation of disulphide bonds is required before heterodimer formation can take place. The rate-limiting step for heterodimer formation is thought to be the folding of E1 (Dubuisson *et al.*, 1996). The N-terminus of E1 is important for interaction with E2. Amino acids 415-500 of the polyprotein fall within the N-terminus of E2 and are important for interaction with E1. Within this region several discontinuous sites seem to contribute to strong binding, particularly the amino acid sequence WHY located at residues 489-491. Glycosylation is not required for heterodimer formation (Yuki *et al.*, 1997).

The high degree of variation within the envelope region of HCV presents a problem in developing a vaccine using this region. Although a vaccine against an individual isolate is theoretically possible the hypervariability of E2 coupled with the presence of so many viral strains means that the development of a universal envelope-based vaccine against HCV is unlikely.

1.8.3. NS2 protein.

The non-structural region of the polyprotein is processed by two viral encoded proteases. The first of these proteases cleaves between the C-terminus of NS2 and the N-terminus of NS3 and is encoded by amino acids from both these non-structural proteins. This protease, designated the NS2/3 protease, or C-pro-1, is encoded between amino acids 827 and 1207 of the virus polyprotein (Reed *et al.*, 1995). This protease encompasses the C-terminus of NS2 and all the NS3 serine protease domain (181 amino acids). This 380 amino acid region can process itself when expressed as a single construct (Reed *et al.*, 1995). Two amino acids essential for protease activity are located in the NS2 region (His-952 and Cys-993). Mutations that inactivate the NS3 serine protease have no effect on NS2/3 protease action (Grakoui *et al.*, 1993a) illustrating that the NS2/3 protease cleavage event is independent of the serine protease activity.

Cleavage occurs between the leucine and alanine residues at the following amino acid sequence - KGWRLL/APIT. Although the sequence APIT is present at other positions in the HCV polyprotein (five in HCV-H), the sequence RLLA is not found, and processing is not observed at these sites. The NS2/3 cleavage site amino acid sequence is highly conserved for different strains of HCV from positions P5 to P6'. Cleavage takes place between P1 and P1' with the 5 residues upstream (P1-P5) and 6 residues downstream (P1'-P6') being highly conserved. Mutations at P1 and P1' are well tolerated except for the introduction of acidic residues and proline. The amino acid preferences at P1 and P1' are similar to those of neutral metallo-proteases that cleave next to hydrophobic amino acids. As NS2/3 is an

autocatalytic protease, amino acid substitutions and deletions that inhibit processing may disrupt proper folding of the autoprotease, the cleavage site, or both. It has also been demonstrated that NS2/3 cleavage can take place *in trans*. An NS2/3 protease, inactivated by mutation of His-952 or Cys 993, could be cleaved in mammalian cells when coexpressed with a functional NS2, providing the substrate contained the N-terminal portion of NS3. Similarly an NS2/3 substrate with a C-terminal deletion in NS3 could be processed when cotransfected with an NS3 fragment containing the N-terminal 181 amino acids (Reed *et al.*, 1995). This observation suggests that the two regions interact to from a functional autoprotease. NS2/3 contains a zinc ion which has been demonstrated to have a structural rather than a catalytic role. The structural requirement is explained by its ability to be replaced by a cadmium or cobalt ion. Ubiquitous soluble factors may also be required for efficient proteolysis (Pieroni *et al.*, 1997). The presence of microsomal membranes has been shown to stimulate the activity of the NS2/3 protease and are believed to act by assisting the proper folding of the precursor (Santolini *et al.*, 1995). The NS2/3 protease of different HCV strains is stimulated to different degrees by the presence of membranes. The NS2/3 protease of HCV-J is unaffected by the presence of membranes whereas that of HCV-BK is entirely dependent on membranes. The targetting of the NS2/3 precursor to membranes is dependent on the (Signal Receptor Particle) SRP-SRP receptor machinery.

After the NS2 protein has been processed from the polyprotein it is observed to be a transmembrane protein. Analysis has shown that the C-terminal region of NS2 is translocated in the ER lumen. The way in which this protein interacts with the targetting machinery or how it is inserted into the ER shows some strange features. NS2 contains no long stretches of hydrophobic amino acids that resemble a signal sequence and/or a transmembrane domain. It is speculated that the signal sequence is located downstream of the cleavage site, within NS3. Targetting to and translocation across the ER membrane are distinct events with the NS2 being translocated entirely posttranslationally (Santolini *et al.*, 1995). In an *in vitro* assay non polar detergents have also been shown to stimulate protease activity and are also thought to promote proper folding of the precursor. The action of detergents is believed to bypass the SRP-SRP machinery and so removes the necessity for microsomal membranes for *in vitro* processing for some HCV strains (Pieroni *et al.*, 1997). The function of NS2, once released from the polyprotein is unknown. The analagous cleavage takes place inefficiently in the cytopathic pestiviruses (e.g. BVDV) but is not observed in the non-cytopathic strains (Collett, 1992). The NS2B/NS3 site in flaviviruses is processed efficiently(Chambers *et al.*, 1991) but the NS2B protein is equivelant to the NS4A protein of HCV (i.e. it acts as the NS3 serine protease cofactor).

1.8.4. The NS3 protein.

The NS3 protein (70kDa) of HCV can be split into two distinct domains. The N-terminus encodes a serine protease (Hijikata *et al.*, 1993a; Bartenschlager *et al.*, 1994; Lin *et al.*, 1994b; Tanji *et al.*, 1994b; Failla *et al.*, 1995) as illustrated in Fig. 1.5. The C-terminal domain encodes an NTPase/RNA helicase. This feature is also seen for the flavivirus NS3 protein (Wengler and Wengler, 1991; Warrener *et al.*, 1993) and the pestivirus p80 protein (Tamura *et al.*, 1993).

The serine protease.

Serine proteases are a group of enzymes that hydrolyse peptide bonds in proteins. Many proteases are present as domains in large multifunctional proteins whereas others are smaller polypeptide chains. Proteases belong to one of four families: serine, cysteine, aspartic, or metallo proteases. Classification of a protein into one of these families depends on the nature of the most prominent functional group in the active site. Enzyme catalysis begins with the bonding of enzyme and substrate to form a reversible enzyme-substrate complex (ES). Once formed, ES can either be fully catalysed to give enzyme and products or revert to the reactants. This reaction can be summarised by the equation below:

$$E + S \underset{}{\overset{K_m}{\rightleftharpoons}} ES \overset{k_{cat}}{\longrightarrow} E + P$$

The dissociation constant of the ES complex back to E + S is called K_m and describes the affinity of the enzyme for the substrate. The second step of the reaction from ES to E + P occurs with a rate constant called k_{cat}, or the turnover number. The turnover number is defined as the maximum number of substrate molecules converted to product per active site of the enzyme per unit time. For simple reactions, k_{cat} is the rate constant for the chemical conversion of the ES complex to free enzyme and products. These definitions are only valid when the concentration of the enzyme is very small in comparison to the substrate i.e. the rate of formation of the first few percent of the product before the substrate has been depleted and enough products have been produced that they can interfere with the catalytic reaction. The maximum rate of reaction is reached when the enzyme is saturated with substrate and is defined as V_{max}. The substrate concentration when the half maximal rate ($V_{max}/2$) is achieved is also known as K_m. The rate constant that refers to the properties and reactions of the free enzyme

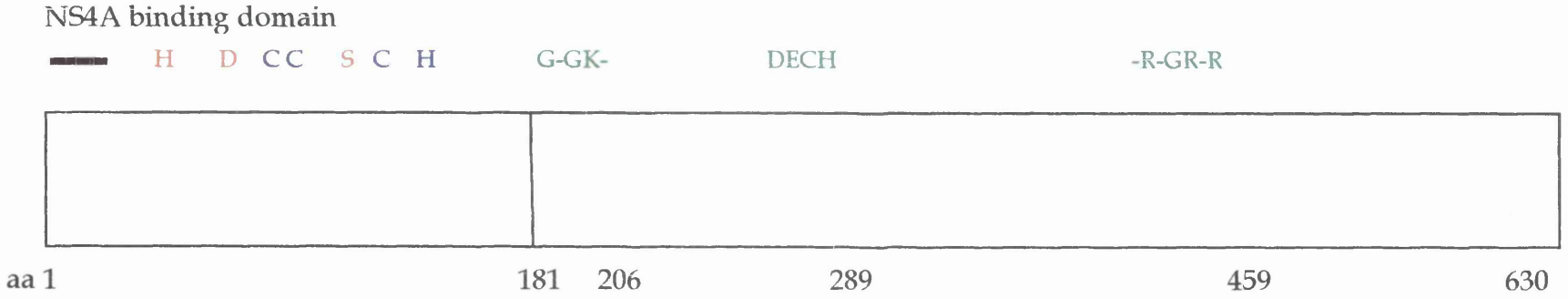


Fig. 1.5. Positional diagram of amino acids essential for NS3 activity. Amino acids essential for serine protease activity are shown in red (illustrating the catalytic triad) and blue (zinc ion coordinating residues) and amino acids required for NTPase/RNA helicase activity are shown in green. The NS4A binding domain is illustrated by the black line.

and free substrate is called the specificity constant (k_{cat}/K_m). This describes the specificity of an enzyme for competing substrates.

The ES complex undergoes several rearrangements to one or several transition states before the products are obtained. These rearrangements require energy. The energy required to bring free enzyme and substrate to the highest transition state is called the activation energy of the reaction. In the absence of enzyme the substrate also proceeds through transition states to give products but in the absence of enzyme the activation energy is significantly higher. Enzymes can reduce the activation by several different methods. They can provide catalytically competent groups for a specific reaction mechanism, bind several substrates in an orientation appropriate to the reaction being catalysed and use the differential binding energy of the substrate in its transition state compared with its normal state. The activation energy for the conversion of ES to E + P is lower if the enzyme binds more tightly to the transition state of S than to its normal structure, i.e. it is catalytically advantageous for the enzyme to be complementary to the transition state of the substrate than to the normal structure of the substrate.

Catalysis by serine proteases occurs in two steps. The first step produces a covalent bond between C_1 of the substrate and the hydroxyl group of the reactive serine residue of the enzyme. To produce this acyl-enzyme intermediate the reaction must first go through a negatively charged, transition state intermediate. In this intermediate the bonds of C_1 have a tetrahedral geometry as opposed to the planar geometry in the peptide group. During this step the peptide bond is cleaved, one peptide product diffuses away while the remaining peptide is attached to the enzyme in the acyl-enzyme intermediate (Fig. 1.6.).

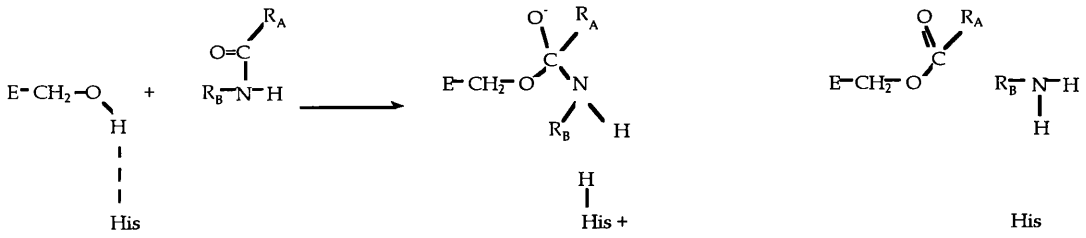


Fig. 1.6. The acylation step of catalysis by serine proteases. The oxygen atom of the catalytic serine residue binds to the carbon atom of the peptide bond to form the negatively charged tetrahedral intermediate while the histidine residue accepts the hydrogen atom. The nitrogen atom of the peptide bond subsequently accepts the hydrogen atom from the histidine residue to form a stable cleavage product that is released from the enzyme. Hydrogen bonding is illustrated by broken lines.

In the second step of the reaction the acyl-enzyme intermediate is hydrolysed by a water molecule to release the second peptide product. This product has a complete carboxy-terminus and the serine residue of the enzyme is restored. This step, known as deacylation, also occurs through a negatively charged tetrahedral transition state intermediate (Fig. 1.7.).

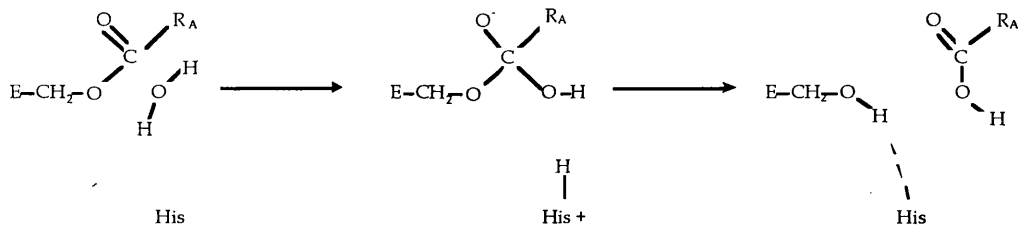


Fig. 1.7. The deacylation step of catalysis by serine proteases. The oxygen atom of the water molecule binds to the enzyme-substrate complex at the carbon atom to form the negatively charged tetrahedral intermediate. The histidine residue of the catalytic triad accepts the hydrogen atom from the water molecule. The second cleavage product is released before the serine residue accepts the hydrogen atom from the positively charged histidine to reform the stable catalytic triad conformation. Hydrogen bonding is illustrated by broken lines.

Serine proteases have four essential structural features that facilitate the catalysis mechanism. The most important feature is that of the catalytic triad, consisting of three side chains from apspartate, histidine and serine residues. The histidine residue is a general base that can accept the proton from the hydroxyl group of the reactive serine residue allowing the formation of the covalent tetrahedral transition state. The aspartic acid stabilises the correct tautomer of histidine, orienting it so that it is able to accept the proton during catalysis (Fig. 1.8.).

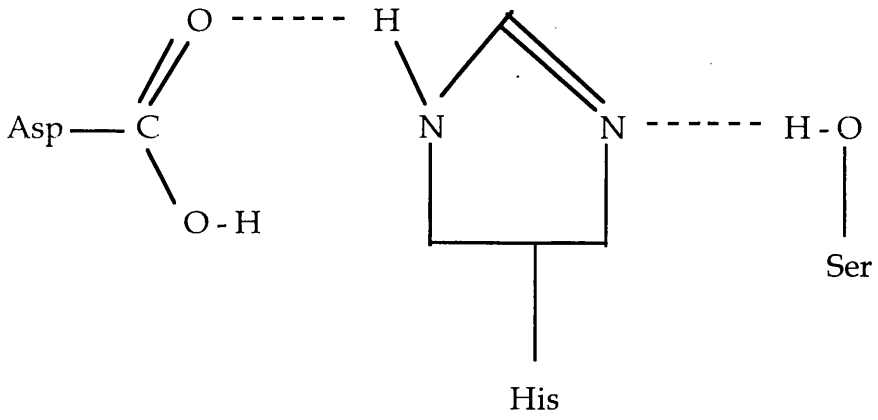


Fig. 1.8. The catalytic triad of the serine protease. The aspartic acid residue forms a hydrogen bond with the histidine residue, orienting the histidine so that it can hydrogen bond with the catalytic serine residue.

The negatively charged transition state intermediate is tightly bound and stabilised by groups that can form hydrogen bonds with the negatively charged oxygen atom at C_1. These groups are located in a region known as the oxyanion hole. The positive charge that develops on the histidine after it has accepted a proton also stabilises the negatively charge transition state. A loop region of the enzyme forms a short antiparallel β sheet, mediated by hydrogen bonding, to main chain atoms of the substrate to give nonspecific binding between enzyme and substrate. Most serine proteases have no strict substrate specificity but many show a preference for a particular side chain before the scissile bond as seen from the N-terminus of the substrate. The preferred side chain is oriented to fit into a pocket of the enzyme called the specificity pocket. Serine protease catalysis was discussed by Branden and Tooze in An Introduction to Protein Structure (1991), Chapter 15.

The NS3 protein of HCV was found to contain the catalytic motif of a trypsin-like serine protease (Miller and Purcell, 1990). The order and spacing of the catalytic triad are conserved between different HCV sequences: His-57, Asp-81 and Ser-139 (amino acid numbering from the start of the NS3 protein). The important amino acid regions of serine proteases can be divided into four separate "boxes" or regions. The first three boxes each contain an amino acid that contributes to the catalytic triad, present in a characteristic spatial pattern. Mutation of amino acids in the catalytic triad results in inactivation of the serine protease (Hijikata *et al.*, 1993a; Manabe *et al.*, 1994). Box 4 contributes to substrate binding. These features are observed for HCV NS3 (Bartenschlager *et al.*, 1993).

The shape of the HCV serine protease specificity pocket is determined by main-chain atoms and by a phenylalanine ring (Pizzi *et al.*, 1994). This indicates that P1 in the substrate has to have a short side chain (Cys or Ser) to fit into the very hydrophobic pocket. The sulfhydryl group of cysteine has been shown to interact favourably with the aromatic ring system of phenylalanine (Bode *et al.*, 1984). The amino acid at P1 in the

substrate is restricted to a short, polar or aliphatic side chain to permit cleavage (Leinbach *et al.*, 1994). Within the substrate P6 is observed to be a conserved acidic amino acid with serine or alanine at P1′ (Grakoui *et al.*, 1993b; Pizzi *et al.*, 1994; Leinbach *et al.*, 1994; Komoda et al., 1994).

Processing by the NS3 protease involves a number of features. NS3 associates with the HCV NS4A protein and is anchored by it to the endoplasmic reticulum (Satoh *et al.*, 1995). Without NS4A, NS3 is usually found in the cytoplasm. The NS4A protein also acts as a cofactor in NS3-mediated processing events (Failla *et al.*, 1994; Bartenschlager *et al.*, 1994). Association of NS3 to the ER membrane through NS4A results in maximal protease activity. ER membrane association of the NS3/NS4A complex may play a role in replicase function (Lin and Rice, 1995). The first 20-30 amino acids of NS3 are required for interaction with NS4A (Failla *et al.*, 1995; Han *et al.*, 1995; Satoh *et al.*, 1995). Association with NS4A stabilises the NS3 protein (Satoh *et al.*, 1995) although association may have to take place at an early stage of expression to create an active protease (Overton *et al.*, 1995). The degree of interaction with NS4A seems to vary between different strains of HCV (Muramatsu *et al.*, 1997). In some strains of HCV, NS3 is found localised, to a minor degree, in the nucleus. This nuclear localisation has been observed to be stimulated by the presence of p53, either through interaction with p53 to form a complex that is transported to the nucleus or by p53 augmenting the expression of a nuclear transporter protein that binds NS3 (Muramatsu *et al.*, 1997).

The Zn^{2+} ion has shown to have an indirect role in NS3 protease function in that it stabilises the correct conformation of the enzyme (Hahm *et al.*, 1995; Han *et al.*, 1995). The Zn^{2+} binding site has a structural composition that utilises the sulfhydryl groups of cysteine residues and the imidazolyl group of a histidine residue. There are four residues involved: Cys-97, Cys-99, Cys-145 and His-149 (Stempniak *et al.*, 1997). The histidine residue is less important than the three cysteine residues in Zn^{2+} binding. This binding motif CXC.....CXXXH is similar to the proposed Zn^{2+} binding motif in enterovirus and rhinovirus proteins. The structural role of the zinc atom is to stabilise the NS3 domain but also to enable recognition of the NS3 region as a substrate for NS2/3 activity (Stempniak *et al.*, 1997). The addition of Zn^{2+} during protease folding stimulates the basal level of polyprotein processing by NS3 (Stempniak *et al.*, 1997). Inhibition of protease activity by Cu^{2+} seems to occur near to or at the active site of the protease (Han *et al.*, 1995). The C-terminal region of NS3 that encodes the NTPase/RNA helicase can be deleted with no effect on protease activity (Failla *et al.*, 1995; Hahm *et al.*, 1995).

Crystallization of the protease domain of HCV has been shown to be difficult due to its poor solubility and tendency to aggregate. The three-dimensional structure of the 189 amino acid protease domain was determined for a type 2b isolate (HCV BK) by Love *et al.* (1996). The protease domain folded into two six-stranded β barrels, similar to those observed for trypsin-like serine proteases despite having virtually no sequence similarity with proteases containing a trypsin-like fold. A 30 amino acid stretch at the N-terminus precedes the first β barrel. This is

significantly larger than in most proteases in their active states. The secondary structure formed by this region is reminiscent of the short β strand found in many cellular proteases. The spatial arrangement of strands within the β barrels is similar to that observed for other trypsin-like proteases with the exception that the loops connecting the strands are relatively short. The two β barrel motifs comprise 140 amino acids of the NS3P domain. After the β barrel structure there is one turn of α helix (4 amino acids) that resembles the first turn of the C-terminal helix found in cellular proteases. This region is followed by a disordered polypeptide stretch. The crystal structure is illustrated in Fig. 1.9. The crystal structure also confirmed the involvement Cys-97, Cys-99, Cys-145 and His-149 in the binding of Zn^{2+} within the protease domain. The three cysteine residues contribute a partial tetrahedral geometry around the zinc ion. The fourth member of the tetrahedral coordination is a water molecule located within hydrogen bond distance to the histidine residue. The zinc ion serves to anchor the turn at βD2-βE2 to the interbarrel loop implying a structural role for the metal. The catalytic triad, the oxyanion hole and strand βE2 (forming one side of the specificity pocket) have the same relative spatial positions as seen in other trypsin-like proteases. A classic catalytic triad configuration would exist during substrate cleavage by the protease domain. The amino acid sequence around the catalytic serine residue is part of the conserved GXSGG motif seen in trypsin-like serine proteases. The oxyanion hole is nearly identical to that observed in other proteases. The βC2-βD2 loop (containing one turn of α helix) situated before the oxyanion hole may add conformational stability at the binding site utilising the region of α helical structure.

A shallow, nonpolar specificity pocket is formed primarily by the side chains of Phe-154, Ala-157 and Leu-135. From the crystal structure it has been proposed that the side chain of P1 in the substrate (OH or SH) will interact with the electron-rich π clouds on the aromatic ring of Phe-154. Amino acids P2-P6 of the cleavage site will form β sheet antiparallel hydrogen bonds with strand βE2. This extensive backbone interaction would allow the acidic side chain at P6 on the substrate to interact with Arg-161 and Lys-165. These interactions also compensate for the apparent lack of P2-P5 side chain to enzyme interactions due to the short βB2-βC2 and βE1-βF1 loops of the protease domain. Phe-154 was shown to be the main determinant of substrate specificity for the S1 pocket by Koch and Bartenschlager (1997). The position of Phe-154 at the bottom of the pocket would determine the specificity of the P1 residue by restricting the length of the P1 amino acid residue that can fit in the pocket and defines the interaction of the residue with the aromatic ring of Phe-154. Ala-157 is located underneath Phe-154 and plays a very minor role in determining the S1 pocket whereas Leu-135 plays a structural role (Koch and Bartenschlager, 1997). The S1 binding pocket is illustrated in Fig. 1.10.

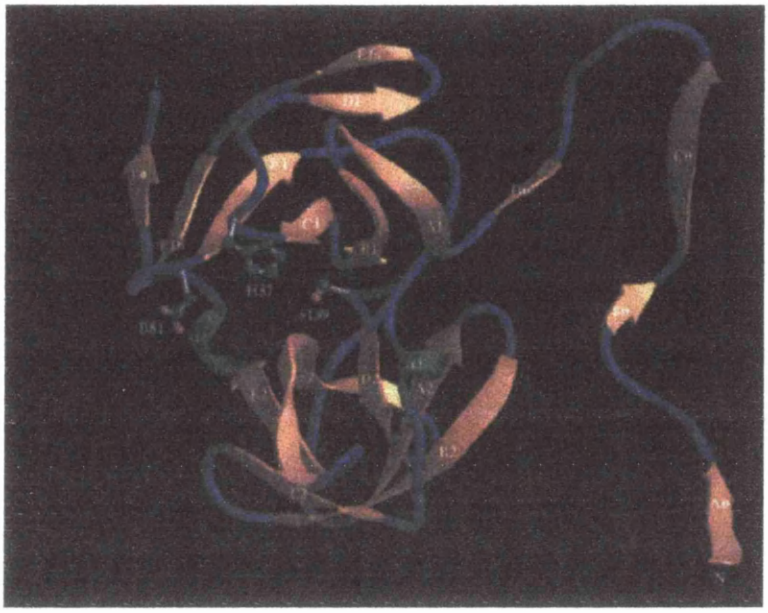


Fig. 1.9. Overall folding of HCV NS3P (taken from Love *et al.*, 1996). View into the active site in the asymmetric unit with catalytic triad residues labeled by the single amino acid code. Secondary structural elements are colour-coded yellow for β strand, green for α helices, and blue for coil. β strands in the trypsin-like barrels (A1-F2) are labelled in the standard convention. Catalytic residues are shown in sphere and cylinder representation, with spheres colour-coded green for carbon, blue for nitrogen , and red for oxygen.

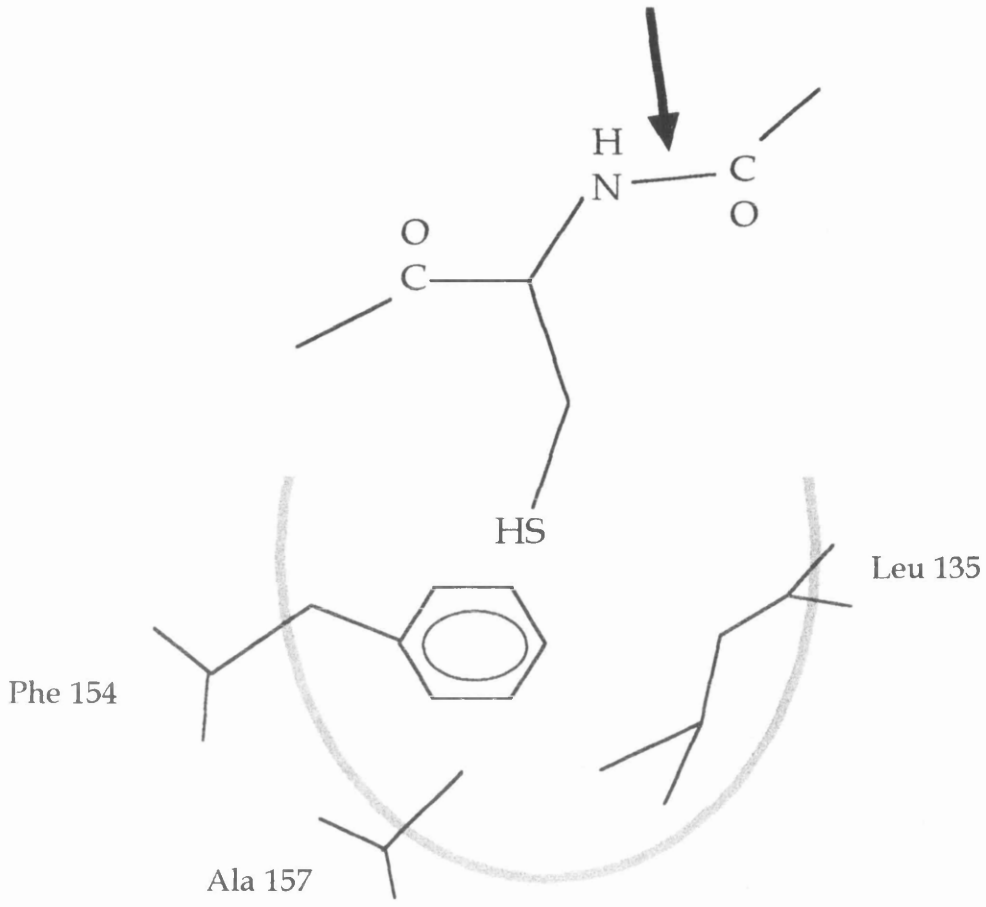


Fig. 1.10. Hypothetical model of the S1 specificity binding pocket of the NS3 serine proteinase. The S1 pocket is defined by the dotted line and containd the P1 side chain of the wild type cysteine residue. Model adapted from Koch and Bartenschlager (1997).

The crystal structure of the protease domain was also determined when bound to the NS4A cofactor (Kim *et al.*, 1996). In this case there were a number of slight differences to that observed for the NS3 protease domain alone. The complex adopted a chymotrypsin-like fold consisting of two domains, each containing a twisted β sheet incorporating a "Greek key" motif. The C-terminal domain (amino acids 1120-1206) consists of a six-stranded β barrel followed by a structurally conserved α helix. The core of this barrel is packed with hydrophobic residues emanating from all six strands of the β barrel. The N-terminal domain (amino acids 1027-1119) contains an eight stranded β barrel with one strand donated from the NS4A cofactor. The C-terminal domain contains the tetrahedrally coordinated zinc ion and the catalytic serine residue of the catalytic triad whereas the histidine and aspartate are located in the N-terminal domain. The NS4A peptide (amino acids 21-39) used in this modelling formed hydrogen bonds with the N-terminal region or one of the strands from its main-chain carbonyl and amide groups from Val-23 to Leu-31. From this analysis it has been shown that NS4A should be considered as an integral structural component of the complex, supporting the observation that NS4A stabilises NS3 (Satoh *et al.*, 1995). Within NS4A several residues found buried in the core of NS3P are critical for complex formation with NS3 (Val-23, Ile-25, Ile-29 and Leu-31). This corroborates with what has been observed in previous NS4A analyses (1.8.6.). Activation of NS3 by NS4A may be a result of incorporation of NS4A into the N-terminal domain β sheet. When incorporated NS4A stabilises the NS3 structure between residues 1030 and 1034 through main-chain hydrogen bonds. The crystal structure of the NS3P/NS4A complex is illustrated in Fig. 1.11.

The NTPase/RNA helicase domain.

The RNA helicase family can be divided into a number of subdivisions. One of these groups called the DEAD box family contains three subfamilies: DEAD, DEAH and DEXH (Koonin, 1991). The HCV NS3 protein has an Asp-Glu-Cys-His sequence and is thought to be a member of the DEXH protein family.

The basal NTPase activities of HCV, flavivirus and pestivirus NS3 proteins are all unaffected by pH, $MgCl_2$ and monovalent cation concentrations when expressed as protein fragments lacking the N-terminal protease domain. In the presence of a polynucleotide stimulator a clear pH optimum of 6.5 was observed. In addition to this there is an optimum $MgCl_2$ concentration of 2.5mM in the presence of 0.35mM ATP (Suzich *et al.*, 1993). All the *Flaviviridae* NTPases tested were capable of hydrolysing NTP and dNTP substrates. Both HCV and pestivirus (BVDV) NTPases were stimulated by poly(U) to a degree higher than that for other polyribonucleotides. HCV was alone however in its equal stimulation by both poly(U) and poly(dU). The HCV enzyme required a higher

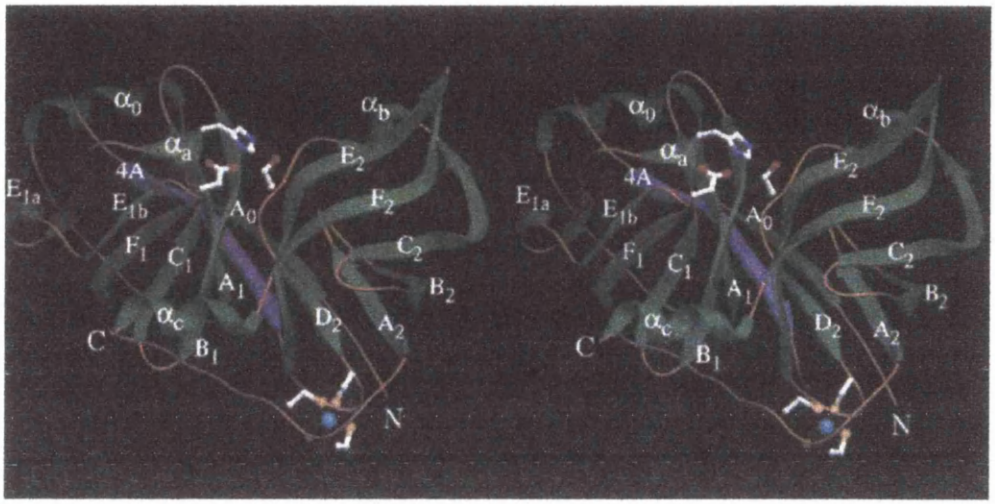


Fig. 1.11. Stereo ribbon diagram of the NS3P: NS4A complex (taken from Kim *et al.*, 1996). View into the active site cleft of the enzyme. The N-terminal domain of the complex is on the left and the C-terminal domain on the right. Secondary structural elements are labelled according to convention used for chymotrypsin. Side-chains of active site residues His-1083, Asp-1107, and Ser-1165, along with Zn^{2+} ligands Cys-1123, Cys-1125, and Cys-1171 are displayed in ball-and-stick representation. Zn^{2+} is coloured cyan and its H_2O ligand red. The β strand formed by NS4A is shown in magenta.

concentration of poly(U) to achieve the half-maximal level of stimulation in comparison to both pestivirus and flavivirus NTPases (Suzich *et al.*, 1993). It has been shown that the NS3 protein interacts with the NS4A protein to form a stable complex with observed effects on protease activity (see **The serine protease**). The NS34A complex has an optimal NTPase pH between 7.5 and 8.5 in contrast to an optimal pH 6.5 for the helicase domain alone (Morgenstern *et al.*, 1997). The catalytic efficiency of the NTPase function was stimulated to a similar extent by saturation with poly(U) for both the NS3 NTPase alone and the NS34A complex. This is not true for saturation with poly(C) and poly(A) that stimulated the NS34A complex NTPase activity significantly more than for the NTPase alone. The NS3-4A complex was also significantly more sensitive to polynucleotide stimulation than the NTPase alone (Morgenstern *et al.*, 1997).

RNA helicases are involved in important cellular functions, such as RNA splicing, ribosome assembly and translational initiation as well as viral RNA helicases functions in viral replication, transcription and protein translation. The C-terminal domain of HCV NS3 has an RNA helicase activity that uses ATP as an energy source (Kim *et al.*, 1995; Tai *et al.*, 1996).

As for the NTPase experiments, a C-terminal fragment only was used to assay RNA helicase activity. A full length NS3 protein may have a higher level of enzymatic activity. The functional domain required for NTPase/RNA helicase activity is 400 amino acids in length between amino acids 1209 and 1608. Of these amino acids two regions (amino acids 1562-1608 and 1193-1224) are believed to be important for correct conformation of the protein (Kim *et al.*, 1997a). The helicase activity occurs in a 3′ to 5′ direction along the RNA template and binds to single-stranded but not blunt-ended dsRNA (Tai *et al.*, 1996; Kim *et al.*, 1997a). The C-terminal 135 amino acids and the N-terminal 32 amino acids of the helicase domain are not required for RNA binding (Kim *et al.*, 1997a). The RNA helicase can unwind the native 5′NTR of HCV in a 5′ to 3′ direction. The directionality of the unwinding of the 5′NTR may be due to the presence of secondary structures or sequence within this region. The ability of the 5′NTR to direct the helicase in this direction means that in natural HCV infection the helicase can act in a bidirectional manner. The HCV NS3 helicase also could unwind DNA-RNA and DNA-DNA substrates and so exhibits DNA helicase activity (Tai *et al.*, 1996). This DNA helicase activity may affect cellular nucleic acid structures or activities and so may be relevant in viral pathogenesis. The HCV helicase is very similar to two cellular helicases, NDH II and NPH II, except that NDH II can also bind dsRNA and is stimulated by, rather than sensitive to, salt concentration. The presence of poly(U) and poly(A), which both stimulate NTPase activity, inhibit RNA helicase function, possibly by inhibiting the binding of the enzyme to substrate. Similarly the presence of polynucleotides inhibits the RNA helicase function of the NS34A complex possibly due to binding competition between poly(U) and duplex RNA for the same or overlapping sites within the NS3 region.

RNA helicases contain eight conserved motifs, four of which have been shown to play an important role. Closest to the N- terminus of RNA

helicases is the A motif. This motif has the basic amino acid sequence of AXXXXGKT/S and is required for ATP binding, ATP hydrolysis and RNA unwinding (Walker *et al.*, 1982). The second motif is the DEAD box element. The first aspartic acid of this element binds to Mg^{2+} through a water molecule (Pai *et al.*, 1990) and this metal ion binds to the ATP bound to the A motif. In the C-terminus of RNA helicases there are two important motifs. An S/TAT sequence is believed to play a role in RNA unwinding and the H/QRXGRXXR motif is believed to play a role in ATP hydrolysis and RNA unwinding.

Point mutations were introduced into these four motifs within an active NS3 construct (Kim *et al.*, 1997b). This construct encoded the C-terminal region of NS3 and produced a fully functional RNA helicase/ATPase. Mutation of the conserved lysine residue within motif A reduced NTPase activity by 95% in the absence of poly(U) and by 50% in the presence of poly(U). This mutation completely abolished RNA helicase activity.

Within the DECH motif of the HCV NS3 a variety of amino acid substitutions were carried out. The effect on activity depended on the charge of the amino acid introduced and the interactions of these residues with other amino acids.

The HCV NS3 contains a threonine residue at the first amino acid of the S/TAT motif. Substitution of this threonine residue with an alanine residue had little effect on the NTPase activity in the absence of poly(U) although this activity was not influenced by the presence of poly(U). The RNA helicase activity was approximately half of the wild type. The eIF-4a RNA helicase SAT motif was important in the RNA unwinding reaction and the contrast in these results implies that the TAT motif may play a lesser role in RNA unwinding than the SAT motif (Kim *et al.*, 1997b).

Enzymatic activity was demonstrated to be highly dependent on the presence of the glutamine residue at the start of the fourth motif. Further mutational analyses of additional residues throughout the conserved motifs demonstrated that the strictly conserved amino acids were of a higher importance for enzymatic activity than the less conserved amino acids.

1.8.5. A novel method for the inhibition of NS3 function.

A strategy developed for *in vitro* genetic selection has led to the isolation of nucleic acids that have the ability to bind to specific target proteins. Using this strategy, nucleic acid molecules, isolated from a pool of random nucleic acids, are obtained that have a high affinity with a target protein. After repeated rounds of selection and amplification, the end products isolated are tight-binding oligonucleotide ligands called aptamers (Kumar *et al.*, 1997).

A pool of 10^{14} random oligonucleotides (120 nucleotides in length) was analysed for binding to the HCV NS3 protein. During this selection cycle a non-specific competitor (tRNA) and a specific competitor (10G-1

RNA, a small aptamer that binds to NS3 isolated by Urvil *et al.*, 1997) were included to increase the stringency of the selection. The selection cycle was repeated six times until 14% of the aptamers present bound to the NS3 protein. All the NS3-binding aptamers were cloned and then were grouped together to give four classes of NS3-specific aptamers (Kumar *et al.*, 1997). The secondary structures of these NS3 specific aptamers are illustrated in Fig. 1.12.

Two aptamers, G6-16 and G6-19, inhibited processing *in vitro* by approximately 50%, moderate inhibition was observed with G6-8 and no inhibition of NS3 protease activity by G6-12. The lack of inhibition of protease activity by G6-12 is consistent with its low binding efficiency to the NS3 protease domain alone. G6-16 and G6-19 bound specifically to the NS3 protease domain.

None of the NS3 specific aptamers had any effect on the NS3 ATPase activity. In contrast G6-12, G6-16 and G6-19 completely inhibited the helicase activity of the NS3 protein (Kumar *et al.*, 1997).

1.8.6. The NS4A protein.

The NS4A protein is a short, 54 amino acid protein of approximately 6 kDa (Tomei *et al.*, 1993). The function of NS4A is primarily as a cofactor for catalysis by the NS3 serine protease. NS4A interacts with NS3 to form a stable complex which is subsequently localised to the ER membranes (Failla *et al.*, 1995; Satoh *et al.*, 1995). In the absence of NS4A, NS3 is located in the cytoplasm. Although NS4A is responsible for localising the protease complex to membranes its activation effect is essentially independent of membranes (Koch *et al.*, 1996).

The NS4A protein is composed of three distinct domains: an N-terminal hydrophobic domain, a central hydrophobic domain and an acidic C-terminal domain (Bartenschlager *et al.*, 1995b) as illustrated in Fig. 1.13. The central domain of NS4A is responsible for the interaction with NS3 (Koch *et al.*, 1996; Steinkuhler *et al.*, 1996a; Steinkuhler *et al.*, 1996b). The minimal domain required for this activation of NS3 was isolated to a central 12 amino acid stretch (Lin *et al.*, 1995; Bartenschlager *et al.*, 1995b; Butkiewicz *et al.*, 1996). Within this region hydrophobic residues, specifically Ile-29, are critical in the role of NS4A as the NS3 cofactor (Lin *et al.*, 1995; Butkiewicz *et al.*, 1996). Mutation of residues such as Val-24, Ile-25, Val-26, Val-30 and Leu-31 to Ala inhibited cofactor activity whereas the alanine substitution for Cys-22 and Arg-28 showed little or no effect (Butkiewicz *et al.*, 1996). The region exhibits a high propensity for the formation of a β strand-like structure (Butkiewicz *et al.*, 1996). The N-terminal domain of NS4A does not seem to be important in cofactor activity (Hamatake *et al.*, 1996; Koch *et al.*, 1996). This domain is very hydrophobic and has the potential to span the lipid bilayer once and so may be important in the membrane anchoring of NS3/NS4A complexes to the ER (Bartenschlager *et al.*, 1994). Overall a weak

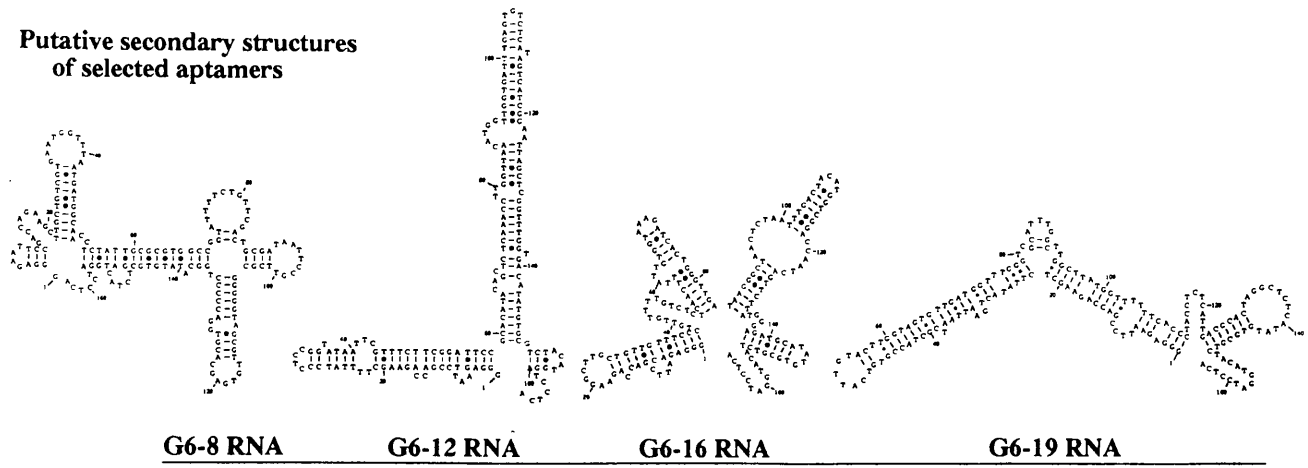


Fig. 1.12. Secondary structures of the aptamers, as drawn by the Mufold program (taken from Kumar *et al.*, 1997)

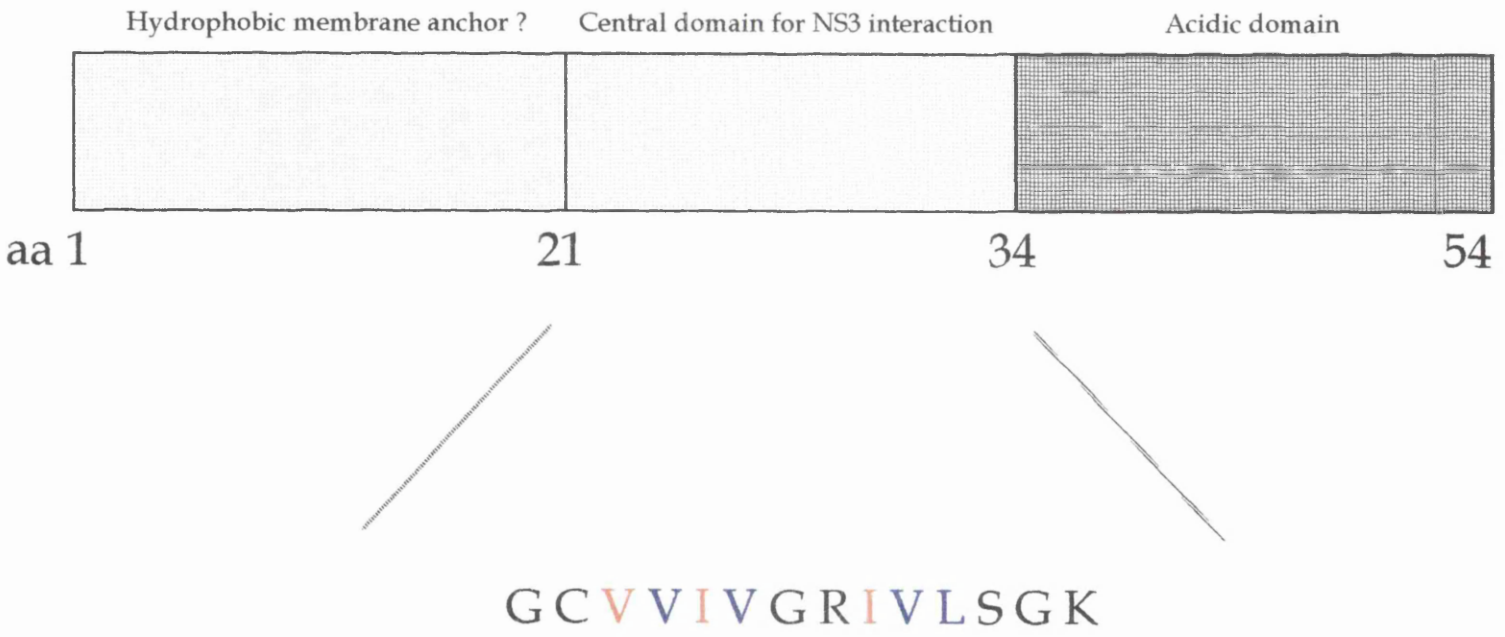


Fig. 1.13. Schematic diagram of NS4A. Residues coloured in red are essential for interaction with NS3 and those residues coloured in blue play a significant role in NS3 interaction.

interaction at multiple sites between NS4A and NS3 is sufficient for protease complex formation (Lin *et al.*, 1995; Koch *et al.*, 1996).

The presence of NS4A is essential for some of the processing events mediated by the NS3 protease, whereas at other sites NS4A merely increases the efficiency of NS3 mediated catalysis (Bartenschlager *et al.*, 1994; Lin *et al.*, 1994b). The presence of NS4A may induce conformational changes in NS3 that favour substrate hydrolysis, and NS4A may contribute directly to substrate binding by interacting with the polyprotein substrate (Lin *et al.*, 1997). NS4A forms a detergent-stable compex with the NS4B5A substrate and this complex formation requires key residues on NS4A (Lin *et al.*, 1997). NS4A may act as a mediator between the NS3 protease and the NS4B5A substrate to bring them into close contact with each other as binding occurs both with and without NS3. Three residues within the central domain of NS4A that are essential for interaction with NS3 (Val-23, Ile-25 and Iso-29) are also essential for the interaction with the NS4B5A substrate (Lin *et al.*, 1997). The multisubunit complex between NS4A and the NS4B5A substrate may also involve residues from NS3 and may also involve cellular components. It has been demonstrated that NS4A from subtype 1a can stimulate proteolysis by subtype 1b NS3 and vice versa (Butkiewicz *et al.*, 1996). The effect of NS4A on the NS3 protein is similar to that of NS2B in flaviviruses and p10 in pestiviruses (Lin *et al.*, 1994b).

1.8.7. The NS4B protein.

The NS4B protein is a 261 amino acid, relatively hydrophobic, 27kDa protein. It is located between amino acids 1712 and 1972 of the HCV Gla. polyprotein. The function of the NS4B protein of HCV is not known at this time. It plays no role in the processing of the polyprotein and thus may have a function in the replication of the HCV genome.

1.8.8. The NS5A protein.

The NS5A protein is encoded between amino acids 1973 and 2419. The NS5A gene is observed to produce two proteins of 56 and 58 kDa in size. Analysis has shown that both proteins are phosphorylated and that p58 is a hyperphosphorylated form of p56 (Tanji *et al.*, 1994a). There are three serine residues that are targetted for hyperphosphorylation at positions 2197, 2201 and 2204 and hyperphosphorylation at these residues is dependent on the presence of NS4A (Tanji *et al.*, 1995). Amino acids 2135 to 2139 of NS5A are important for NS4A-mediated hyperphosphorylation (Asabe *et al.*, 1997). There are two regions within the NS5A protein that are involved in basal phosphorylation (in the absence of NS4A). The first region falls between amino acids 2200 and 2250 and the second between amino acid 2350 and the C-terminus of the NS5A protein (Tanji *et al.*, 1995). Basal phosphorylation of

the 2200-2250 region may be greatly enhanced by deletion of the C-terminus of NS5A. Both these phosphorylation events take place after the NS5A has been processed from the viral polyprotein. A nuclear localisation-like signal has been identified at the C-terminus between amino acids 2326 and 2334 (PPRKKRTVV). A diagram showing the amino acid positions within the NS5A protein is illustrated in Fig. 1.14.

Both forms of NS5A are detected in the nuclear periplasmic membrane fraction although no trace of either form was detected within the nucleus (Tanji *et al.*, 1995). The phosphorylation pattern of the NS5A protein between diverse HCV types is maintained. Phosphorylation is mediated by cellular serine/threonine kinases with characteristics similar for both *in vivo* and *in vitro* experiments (Reed *et al.*, 1997). The kinase involved is believed to be a member of the CMGC kinase family (consisting of CKII and proline-directed kinases; CDK, MAPK and GSK3). Phosphorylation takes place mostly on serine residues although there is a low level of threonine phosphorylation. The C-terminus of NS5A is acidic, containing a large number of proline and serine residues, which agrees with the features observed for phosphorylation by CMGC kinases. There are a number of reasons why phosphorylation of the NS5A protein is believed to be mediated by cellular kinases. NS5A phosphorylation takes place in the absence of any viral proteins and NS5A itself does not contain any motifs associated with eukaryotic kinases. *In vitro* phosphorylation exhibits a preference for Mn^{2+} and is inhibited by Ca^{2+} (although Ca^{2+} may already be bound by the enzyme) (Reed *et al.*, 1997).

One of the proposed functions of NS5A is that it forms part of the replication complex for HCV. This theory is based on the observation that it can be coprecipitated with both NS2 and NS5B (the HCV polymerase). Alternatively, NS5A has also been linked with having a function in resistance to treatment with interferon. In 1995, Enomoto *et al.* showed that sensitivity to interferon-α therapy varied between different HCV isolates. Analysis of the interferon-resistant HCV types showed that a 40 amino acid stretch in the NS5A region, from amino acid 2209 to 2248, contained a unique amino acid change at amino acid 2218, that was exclusive to IFN-resistant HCV sequences. This region was identical to that observed in the prototype HCV type 1b sequence and was designated the interferon sensitivity determining region (ISDR). Further characterisation has shown that 13% of patients who had 1-3 amino acid changes from the wild type (HCV-J) sequence in this region were resistant to interferon treatment. One hundred percent of patients who had greater than 4 amino acid changes were complete responders (Enomoto *et al.*, 1996). Patients treated with interferon-β instead of interferon-α also showed a correlation between amino acid sequence in the ISDR of different HCV isolates and interferon-β resistance (Kurosaki *et al.*, 1997). NS5A has been shown to repress the IFN-induced protein kinase, PKR, which is a mediator of IFN-induced antiviral resistance through a direct interaction with the protein kinase catalytic domain (Gale *et al.*, 1997). This mechanism may be one by which HCV avoids the antiviral effects of IFN. However, one report by Zeuzem *et al.*

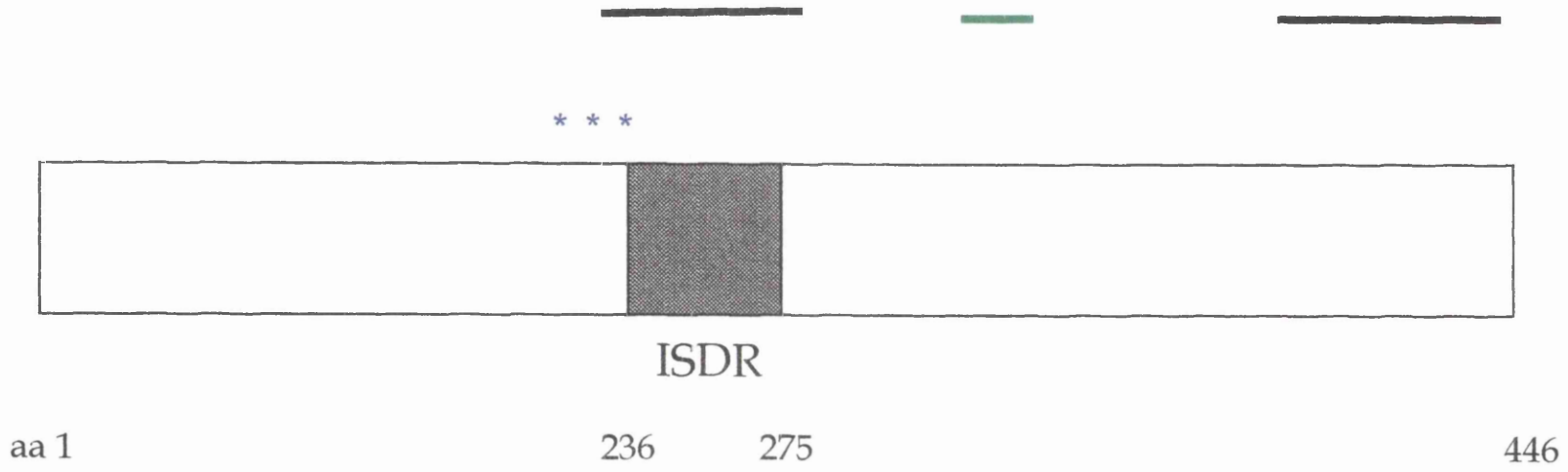


Fig. 1.14. Representation of the NS5A protein. The basal phosphorylation sites are represented by (▬▬), the hyperphosphorylated serine residues are represented by (*) and the nuclear localisation regions are represented by (▬▬).

(1997) did not find any link between the amino acid sequence in the ISDR and interferon resistance, as did a report by Khorsi *et al.* (1997). An NS5A fragment deleted at its N-terminus had the ability to activate transcription in yeast cells when fused to GAL4 DNA binding domain. Amino acids 130-352 of NS5A were critical. This region contained two acidic domains and a proline-rich region that have been shown to be involved in the function of several transcriptional activators (Tanimoto *et al.*, 1997).

1.8.9. The NS5B protein.

Analysis of the HCV polyprotein sequence by Miller and Purcell (1990) predicted that the C-terminal region of NS5B encoded an RNA-dependent RNA polymerase (RdRp) which was responsible for replication of the viral genome. Four motifs (A-D) within the NS5B protein have been shown to play a role in polymerase function (Lohmann *et al.*, 1997) and are shown positionally within NS5B in Fig. 1.15. Motif A, which probably plays a role in NTP binding and catalysis, is characterised by the presence of an invariant aspartic acid residue and is essential for enzymatic activity. Mutation of motif B, containing an invariant glycine residue, results in little or no RdRp activity. Motif C, consisting of the distinctive GDD sequence associated with polymerases, is sensitive to mutation at the central aspartic acid residue but tolerates mutation of the flanking residues, albeit with a lower enzymatic activity. The conservative mutation of arginine to lysine within motif D enhanced the activity of the enzyme by 50% (Lohmann *et al.*, 1997). The presence of the arginine residue that reduces RdRp catalytic activity within NS5B may explain the low replication levels of HCV *in vivo*. It should be noted that the polymerase assay used in these experiments did not involve any other of the viral proteins.

The RdRp activity of HCV was shown to be dependent on Mg^{2+} ions, had an optimum pH of 7.5 and was tolerant to variations in salt concentration (Behrens *et al.*, 1996). Polymerase activity was shown to act by a "copy-back" mechanism on the input RNA. It was also shown not to be restricted to RNA molecules containing the HCV 3'UTR (Behrens *et al.*, 1996; Lohmann *et al.*, 1997). The RdRp uses the 3'-terminal OH group as a primer for polymerisation producing the complementary strand copy of the input RNA. When the 3'-terminal OH was blocked, a DNA or RNA oligonucleotide primer could be used to initiate polymerisation (Behrens *et al.*, 1996; Lohmann *et al.*, 1997). Polymerisation was observed for both homo- and heteropolymeric RNA molecules of up to several hundreds of nucleotides in length. The "copy-back" mechanism may be a common feature of RNA polymerising enzymes. The observation that the HCV RdRp had the ability to replicate unrelated RNAs suggests that the NS5B able to carry out viral RNA replication and that there must be a specific feature that directs replication of the HCV genome. The restriction of RNA synthesis exclusively to HCV viral templates must require viral and/or cellular proteins. NS5B produced from the viral polyprotein was more

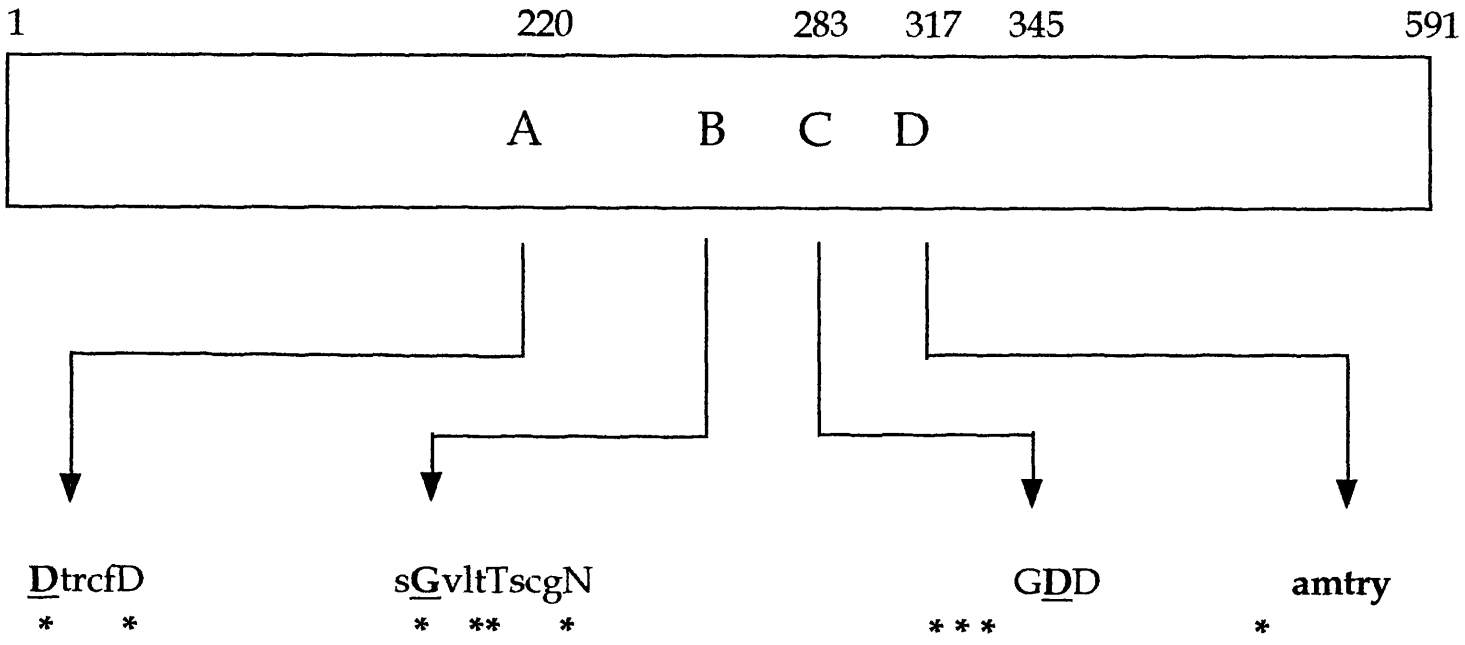


Fig. 1.15. Schematic representation of the NS5B protein. The positions of the amino acid motifs A to D are indicated.Residues marked by asterisks are highly conserved among HCV isolates. Capital letters indicate residues highly conserved among plus-strand RNA virus polymerases and residues underlined and bold are residues highly conserved amaong all viral polymerases. Diagram taken from Lohmann *et al.* (1997).

efficient at RNA synthesis than NS5B expressed as a single construct (Behrens *et al.*, 1996). This feature may be due to more accurate NS5B folding when it is expressed as part of the HCV polyprotein. Alternatively one or more of the viral non-structural proteins may act as cofactors for NS5B directed RNA synthesis.

The NS5B protein is also believed to encode a terminal nucleotide transferase (TNTase) which incorporates UMP preferentially at the 3′ end of the input RNA (Behrens *et al.*, 1996). Although this feature has also been observed for other viral RdRps the HCV NS5B protein was not purified to homogeneity. The TNTase activity may be due to a cellular enzyme that is co-purified or associated with NS5B. That the TNTase activity was approximately the same, irrespective of the level of the RdRp activity, backs up the observation that there may not be an endogeneous TNTase function (Lohmann *et al.*, 1997). The NS5B protein is unstable when expressed in cells (Lin *et al.*, 1994b). This observation may be a feature of regulation with NS5B probably being degraded through a cellular pathway as it is stable in reticulocyte lysates.

1.9. Polyprotein processing.

The polyprotein of HCV encodes both the structural and non-structural proteins. The structural proteins, located at the N-terminus of the polyprotein, are cleaved away from the polyprotein by the action of host-encoded signal peptidases located in the ER lumen (Hijikata *et al.*, 1991b; Tomei *et al.*, 1993). Host signal-peptidase mediated processing of the structural region is also seen for flaviviruses and pestiviruses with small differences between each genus (C.M. Rice, 1996). The non-structural region is processed by two proteases encoded within the virus polyprotein. The junction between the C-terminus of NS2 and the N-terminus of NS3 by a zinc-dependent metallo-protease encoded by the C-terminus of NS3 and the N-terminal third of NS3 (**1.8.3.**). The remaining cleavage sites within the polyprotein are processed by the NS3 serine protease (Grakoui *et al.*, 1993b; Bartenschlager *et al.*, 1993; Tomei *et al.*, 1993; Manabe *et al.*, 1994). The major cleavage pathway of the non structural region of the HCV polyprotein is shown in Fig. 1.16.

The presence of uncleaved NS2 sequence at the N-terminus of NS3 inhibits subsequent NS3 protease activity (D'Souza *et al.*, 1994). The helicase domain of NS3 can be deleted with no inhibitory effect on the serine protease activity (Bartenschlager *et al.*, 1994; Hahm *et al.*, 1995). Processing of the non-structural region in this way is similar to flaviviruses and pestiviruses. In flaviviruses NS3 is responsible for most of the processing within the non-structural regions with the exception of the NS4A/NS4B site which is processed by a signal peptidase (Chambers *et al.*, 1993; Lin *et al.*, 1994a; Preugschat *et al.*, 1991; Speight *et al.*, 1988). The pestivirus non-structural region is processed by p80 except for the p80 N-terminus. The enzyme responsible for this cleavage has not been determined. Both the

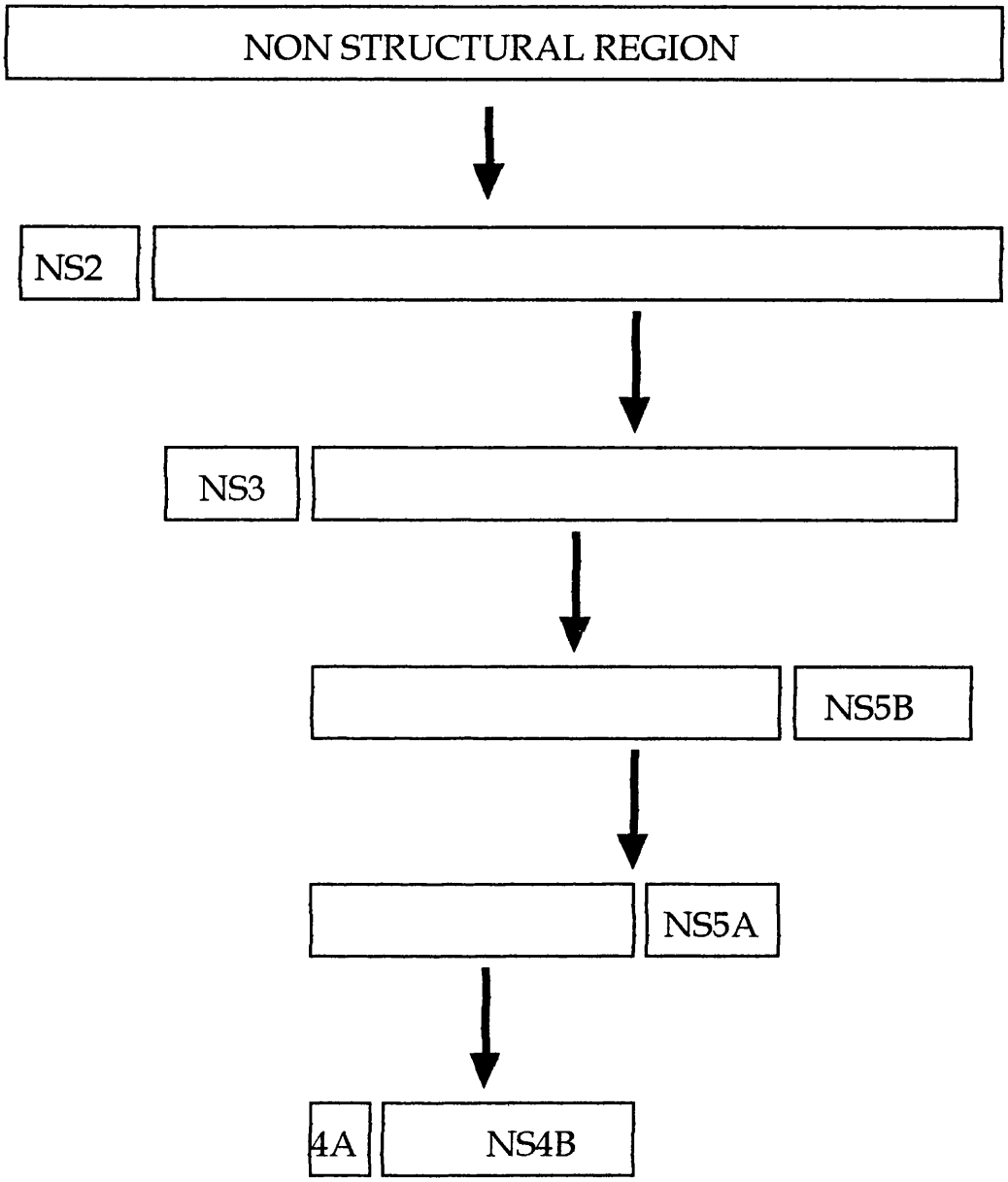


Fig. 1.16. Schematic illustration of the presumed major cleavage pathway of the non structural region of the HCV polyprotein (from Tanji *et al.*, 1994)

flavivirus NS3 and the pestivirus p80 are analagous to the NS3 protein of HCV. In flaviviruses a cofactor NS2B is required for processing by the protease. HCV NS3 has a cofactor, NS4A, whose effect on processing varies from site to site.

Processing at the NS3/NS4A site was shown to occur intramolecularly (*in cis*) whereas the remaining processing events at NS4A/NS4B, NS4B/NS5A and NS5A/NS5B occur intermolecularly (*in trans*) by Bartenschlager *et al.* and Tomei *et al.* in 1993. These processing features are similar to that seen for flaviviruses and pestiviruses. The NS2A/NS2B and NS2B/NS3 processing events take place *in cis* whereas the NS4B/NS5 cleavage site is processed *in trans* in flaviviruses. For pestiviruses both the N and C-termini of p80 are produced from intramolecular cleavage events while the remaining cleavage sites downstream of p80 are processed intermolecularly. Once released from the polyprotein all HCV proteins are associated, more or less, with membranes although a small amount of NS3 and NS5A can be isolated in the membrane-free fraction (Hijikata *et al.*, 1993b). The non-structural proteins may form a complex possibly containing the replicatory machinery and even the roots that hold the genome on the surface of the ER during replication.

Analyses of the NS3-mediated cleavage sites have given the following consensus sequences at each site:

	P6	P5	P4	P3	P2	P1	/	P1′
NS3/NS4A	**D**	L	E	V/I	V/M	**T**		**S**
NS4A/NS4B	**D**	E	M	E	E	**C**		**A/S**
NS4B/NS5A	**E**	C	X	X	P	**C**		**S**
NS5A/NS5B	**E/D**	D	V	V	C	**C**		**S**

This analysis gives a consensus cleavage site sequence as:

E/D X X X X C/T S/A

All the cleavage sites processed *in trans* contain a cysteine residue at P1 whereas the only cleavage site processed *in cis*, NS3/NS4A, has a threonine residue at P1. For all the cleavage sites an acidic residue is present at P6 with serine or alanine present at P1′. No basic residues are found between P7 and P2′ for any of the cleavage sites (Kolykhalov *et al.*, 1994). From further *in vitro* analysis the minimal domain required for processing falls between P4 and P1′ and residues immediately N and C-

terminal to these amino acids may be essential for efficient cleavage (Komoda *et al.*, 1994).

As for HCV the pestivirus structural proteins (N^{pro}, C, E^{rns}, E1 and E2) are released from the polyprotein mostly by cellular enzymes (Elbers *et al.*, 1996; Rumenapf *et al.*, 1993). The non-structural region downstream of the NS3 is processed by the NS3 serine protease at NS3/NS4A, NS4/NS4B, NS4B/NS5A and NS5A/NS5B (Xu *et al.*, 1997; Tautz *et al.*, 1997).

The only conserved residues at the NS3 mediated cleavage sites are a leucine at P1 and either a serine, alanine or asparagine residue at P1'. The residue at position 216 within the protease defines the depth of the P1 specificity pocket. For the pestivirus bovine viral diarrhoea virus (BVDV) the residue at position 216 is a glycine which results in a deeper binding pocket than that observed for HCV. An asparagine residue at position 190 limits the width of the specificity pocket (Xu *et al.*, 1997).

As for HCV, pestivirus NS3 protease mediated processing is stimulated by the presence of a cofactor, NS4A. Processing at the NS4B/NS5A and NS5A/NS5B sites requires the presence of NS4A. The pestivirus NS4A protein is a relatively acidic, 64 amino acids peptide. It has a hydrophobic N-terminus and a highly charged, acidic C-terminus similar to the HCV NS4A. The BVDV NS4A has a number of conserved acidic residues within the central region, differing from the hydrophobic central region of the HCV NS4A (Xu *et al.*, 1997).

The NS3 serine protease of flaviviruses processes the non-structural region of the polyprotein both upstream and downstream of the NS3 protein. The flavivirus NS3 protease processes at NS2A/NS2B, NS3/NS4A and NS4B/NS5. The consensus cleavage site for these processing events is G(A)RR / S(G). The P1 and P2 residues are highly conserved, basic amino acids and the P1' residue always contains a short side chain.

The flavivirus NS3 serine protease also uses a cofactor, located upstream of the NS3, NS2B (Zhang *et al.*, 1992). The NS2B cofactor is essential for cleavage at all the NS3 protease-dependent sites (Chambers *et al.*, 1993). The highly charged, central region of this protein is responsible for cofactor function (Chambers *et al.*, 1993).

For all three genera of the *Flaviviridae* the cleavage sites processed by the NS3 serine protease contain amino acids with short side chains at the P1' position.

1 10. HCV replication.

The replication of the HCV genome has not been well characterised due to the absence of an efficient tissue culture system for propagating the virus. The replication of RNA in other members of the *Flaviviridae* has been characterised (Fig. 1.17.). For flaviviruses, after the genomic RNA has been translated to produce the polyprotein, complementary negative-sense RNA strands are produced, which are subsequently used as templates to produce more genome-length positive-sense RNAs. These positive-sense strands are

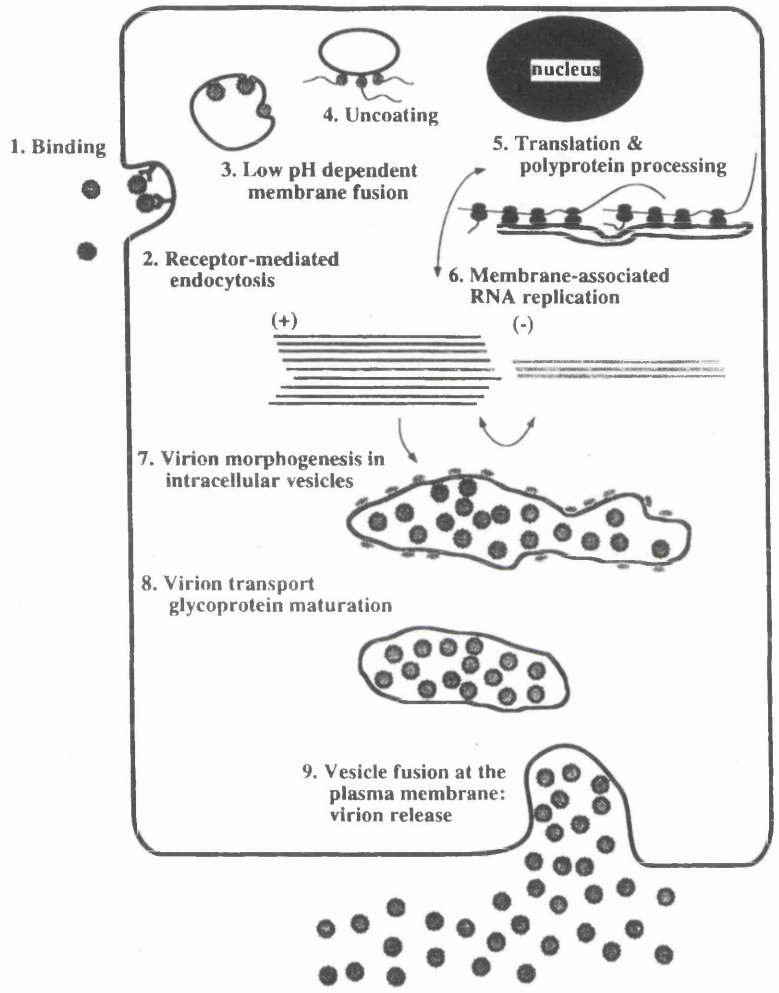


Fig. 1.17. The Flavivirus lifecycle (from C.M. Rice, 1996).

used for translation of the polyprotein, production of additional negative-sense RNA strands or are packaged into virions (Brinton *et al*, 1986; Trent and Naeve, 1980; Westaway, 1987). The synthesis of negative-sense RNA strands from the positive-sense genomic RNA occurs through a semi-conservative mechanism involving replicative intermediates (RIs) and replicative forms (RFs) (Chu and Westaway, 1985; Cleaves *et al.*, 1981). RIs are defined as duplex RNA molecules whereas RIs contain both double-stranded and single-stranded RNA regions (Chu and Westaway, 1985; Cleaves *et al.*, 1981). These replication intermediates and replication forms are detected in infected cells as well as *in vitro* in RNA polymerase reactions (Bartholomeusz and Wright, 1993; Chu and Westaway, 1985; Grun and Brinton, 1986). Regulation of positive and negative strands is believed to take place as at the peak of RNA synthesis in infected cells the ratio of positive-sense to negative-sense RNA molecules is 10:1 (Cleaves *et al.*, 1981). A similar replication process has been postulated for pestiviruses involving both RIs and RFs and the same ratio of positive-sense RNA to negative-sense RNA has been observed (Gong *et al.*, 1996).

HCV is assumed to replicate in the same manner as flaviviruses and pestiviruses. Full length genomic RNAs from flaviviruses (Kapoor *et al.*, 1995) and pestiviruses (Moorman *et al.*, 1996) have been shown to be sufficient to produce infection and infectious virus. Replication of HCV RNA transcribed from a molecular clone has been demonstrated by Kolykhalov *et al.* (1997) and Yanagi *et al.* (1997). Full length HCV transcripts containing both the 5′ and 3′ UTRs were produced under the control of a T7 promoter and were introduced by intrahepatic injection into chimpanzees. A consensus sequence within the open reading frame was required to give HCV replication. The observation that a consensus sequence was required suggests that many quasispecies are defective for replication. The length of the poly (U/UC) tract varied between different clones but all transcripts resulted in HCV replication although a longer tract was preferred (Kolykhalov *et al.*, 1997). The HCV RNA was resistant to RNase digestion implying that the RNA was packaged. HCV RNA alone is sufficient to cause disease in chimpanzees.

1.11. Aims of this project.

At the time of beginning this project (1994) the NS3 protein was known to be responsible for most cleavage events within the non-structural region of the HCV polyprotein. The positions of the NS3-mediated cleavage sites were known and the basic motif determined.

The objectives of this project were:

1) To develop a simple and efficient assay to study the processing of an HCV substrate *in trans*.
2) To determine the substrate specificity for cleavage *in trans* by the NS3 serine protease.
3) To investigate the protease requirements for cleavage.

Chapter 2 - Materials and Methods

2.1. Materials.

2.1.1. Enzymes and inhibitors.

Most restriction enzymes, calf intestinal phosphatase (CIP) and *Taq* DNA polymerase were purchased from Boehringer Mannheim. T4 DNA polymerase and T4 polynucleotide kinase were purchased from New England Biolabs. T7 RNA polymerase and RNasin were obtained from Promega. Sigma supplied RNaseA and lysozyme.

All enzyme reactions were carried out under the reaction conditions specified by the manufacturer unless otherwise stated.

2.1.2. Radiochemicals.

Both radiochemicals listed below were produced by Amersham.

[^{35}S] L-methionine (>800 Ci/mmol)
[α^{35}S] dATP (>1000 Ci/mmol)

2.1.3. Synthetic oligonucleotides.

M13/pUC sequencing primer (-40) 17-mer and M13/pUC reverse sequencing primer (-48) 24-mer were supplied by New England Biolabs. Other oligonucleotides were synthesised in the Division of Virology on a Cruachem PS250 automated synthesizer.

2.1.4. Plasmids.

pTF7.5 NS34A produces a single transcript encoding the NS3 gene followed by the gene for NS4A, pTF7.5 NS3 expresses NS3 alone, pTM1 NS5 and pTZ18 NS5 both express the full length NS5 region of HCV. All four constructs were supplied by Dr M. McElwee.

Cloning vectors pTZ18 and pUC119 were purchased from USB, and pTM1 was obtained from Dr. B. Moss, NIH. Cloning vector pGEM-1 was supplied by Promega and pQE30 was supplied by Qiagen Inc.

The plasmid ΔLL NS3-5_{S-A} encodes the NS3-NS5 region of HCV A (containing an inactive serine protease) and was obtained from Dr B.E. Clarke at GlaxoWellcome plc.

2.1.5. Bacterial strains.

E. coli strain DH5α (ϕ80d*lac*ZΔM15, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17 (r_K^-, m_K^+) *sup*E44, *rel*A1, *deo*R, Δ(*lac*ZYA-*arg*F) U169) was used for propagation of recombinant plasmids.

E. coli strains M15[pRep4] and SG13009[pREP4] were used for the expression of 6-His tagged proteins and were both obtained from Qiagen Inc. Both strains are Nal^S, Str^S, rif^S, lac^-, ara^-, gal^-, mtl^-, F^-, $recA^+$, uvr^+.

2.1.6. Virus stocks.

vTF7.3 is a recombinant vaccinia virus that expresses T7 RNA polymerase (Fuerst *et al.*, 1986) and was obtained from Dr. B. Moss.

2.1.7. Mammalian cell lines and tissue culture growth media.

CV-1 cells (an African green monkey kidney cell line) were used to grow vTF7.3 and for the transfection of plasmid DNAs. The cell line was maintained in Dulbecco's modified Eagle's medium (DMEM, from Gibco-BRL) supplemented with 10% foetal calf serum (FCS), 100 units/ml penicillin and 100 units/ml streptomycin (all from Gibco-BRL). The cell line was grown at 37°C in a 5% CO_2 humidified incubator.

2.1.8. Antibodies and antisera.

The monoclonal antibody SV5-P-k (Hanke *et al.*, 1992) was obtained from Dr R. Randall of The University of St. Andrews.

The NS3 antisera (1245 and 717) and NS5 antisera (1040 and Baculo) were supplied by GlaxoWellcome.

2.1.9. Chemicals.

Unless stated otherwise, all chemicals and reagents were obtained from Sigma Chemicals Co. or BDH Chemicals Ltd.

En^{3}hance was supplied by DuPont. Ammonium persulphate and temed were supplied by Bio-Rad.

2.1.10. Miscellaneous materials.

GeneClean III and MerMaid nucleic acid purification kits were purchased from Bio 101 Inc. X-Omat S X-ray film was obtained from Kodak Ltd. And all tissue culture plasticware was supplied by Nunc. Pansorbin cells were purchased from Calbiochem-Novabiochem and OptiMem 1 was obtained from Gibco BRL.

2.2. Manipulation of DNA.

2.2.1. Purification of synthetic oligonucleotides.

Oligonucleotides synthesised "in-house" were supplied as protected molecules still coupled to the support matrix. They were de-protected and eluted from the column by pushing 1.5ml of ammonia through the column between two 2ml syringes at a rate of 0.5ml every 20 minutes. The oligonucleotide/ammonia solution was then heated for 5 hours at 55°C to remove the base protecting groups and then removed to ice for 5 minutes. The oligonucleotide was dried under vacuum to remove the ammonia and resuspended in 200µl dH_2O. One millilitre of butanol was added and the mixture vortexed for 15 seconds before centrifugation at 13,000 rpm for 1 minute. The supernatant was then removed and the pellet dried under vacuum for 5 minutes. The purified oligonucleotide was then resuspended in 200µl dH_2O.

2.2.2. Polymerase Chain Reaction (PCR) amplification of DNA.

Amplification of DNA was carried out using *Taq* DNA polymerase and 10x PCR buffer, 50mM $MgCl_2$ and detergent W-1 supplied with the polymerase by the manufacturer (Boehringer Mannheim). The PCR reaction mix consisted of the components shown below, the order in which they are displayed being that in which the components were added to the reaction tube (an 0.5ml microfuge tube).

Nuclease free water (to a final volume of 50µl)	xµl
10x PCR buffer	5.0µl
Detergent W-1	2.5µl
50mM $MgCl_2$	2.0µl
4 dNTPs (10mM)	1.0µl
Upstream primer	100 pmol
Downstream primer	100 pmol
Taq DNA polymerase	0.5µl
Template DNA (10-50ng/µl)	1.0µl

The reaction mixture was vortexed, centrifuged for 5 seconds before the addition of 2 drops of nuclease-free light mineral oil (to overlay the top of the reaction mixture). The reaction was then carried out using a Perkin Elmer Model 480 thermal cycler. The reaction was carried out at the following temperatures, strand separation at 94°C for 30 seconds, primer annealing at 50°C (temperature varies depending on primer) for 40 seconds and strand elongation from primers at 72°C for 120 seconds (time varies with length of PCR product) for 30 cycles.
PCR was also carried out using the proof reading enzyme *Pfu* using 10x *Pfu* PCR buffer supplied with the enzyme (Stratagene). The reaction was carried out as illustrated below.

Nuclease free dH_2O	40µl
10x *Pfu* PCR buffer	5µl
4 dNTPs(10mM)	1µl
100pmol upstream primer	1µl
100pmol downstream primer	1µl
Pfu DNA polymerase	1µl
Template DNA (10-50ng)	1µl

The actual reaction was carried out in the same manner as for *Taq* PCR reactions with a strand separation temperature of 95°C and a strand elongation temperature of 74°C.

PCRs were electrophoresed through a 1% TAE gel and purified using a silica matrix (see 2.2.9).

2.2.3. Small scale plasmid DNA preparation by boil lysis.

A 20ml tube containing 5ml of L-broth with 100µg/ml ampicillin was inoculated with a single bacterial colony and incubated with shaking overnight at 37°C. One and a half ml of the overnight culture were transferred to an 1.5ml microfuge tube and centrifuged at 13,000 rpm for 30 seconds. The supernatant was gently removed and the pellet resuspended in 400µl of STET solution containing 500µg/ml lysozyme. The sample was kept on ice for 5 minutes before boiling for 30 seconds and subsequent centrifugation at 13,000 rpm for 15 minutes. The pelleted bacterial debri were carefully removed using a sterile toothpick, and 400µl of isopropanol added to the supernatant. The sample was mixed thoroughly before incubation at -20°C for 30 minutes. The microfuge tube was then centrifuged at 13,000 rpm for 10 minutes, the supernatant removed and the pellet washed with ice cold 70% ethanol. The DNA pellet was dried under vacuum for 5 minutes before resuspension in 100µl dH_2O.

L-broth: 10g NaCl, 10g Bactopeptone and 5g yeast extract per litre.

STET solution: 100mM NaCl, 10mM Tris-HCl (pH 8.0), 1mM EDTA (pH 8.0), 5% Triton X-100.

2.2.4. Large scale preparation of plasmid DNA.

A 20ml universal containing 10ml of L-broth (2.2.3) with 100µg/ml ampicillin was inoculated with either a single bacterial colony or 20µl of a glycerol stock. The universal was incubated with shaking at 37°C for 8 hours. The medium was then transferred to a 250ml bottle containing 100ml L- broth/ampicillin (see above) and the bottle shaken vigorously overnight at 37°C. The bacterial culture was then centrifuged at 3,000 rpm for 10 minutes, the supernatant removed and the pellet vortexed before resuspension in 5ml Solution 1. Ten millilitres of freshly made Solution 2 were slowly added, the tube inverted 10 times and left to sit at room temperature for 10 minutes before vigorous mixing with 5ml Solution 3.

After incubation on ice for 5 minutes the tube was centrifuged at 2,500 rpm for 5 minutes and the supernatant carefully removed to a fresh

tube before the addition of an equal volume of isopropanol and incubation at -20°C for 10 minutes. The tube was centrifuged at 3,000 rpm for 10 minutes, the supernatant decanted and the remaining pellet dissolved in 1.5ml TE buffer. Two ml of ice cold 5M LiCl were added to precipitate the RNA and the mixture kept in ice for 5 minutes before centrifugation at 2,500 rpm for 5 minutes. The supernatant was removed to a fresh tube containing 2 volumes of ethanol, the solutions mixed and then incubated at -20°C for 10 minutes. After a further centrifugation for 10 minutes at 2,500 rpm the supernatant was removed and the pellet washed with 70% ethanol before resuspension in 0.5 ml TE buffer and transferral to a 1.5ml microfuge tube. Two µl of RNase A solution and 2µl of Proteinase K solution were added and the sample incubated at 37°C for 15 minutes. The DNA was precipitated by the addition of 0.5 volumes of 20% PEG/2.5M NaCl and incubation on ice for a further 15 minutes. The DNA was pelleted by centrifugation at 13,000 rpm for 10 minutes, the supernatant removed and the pellet subsequently dissolved in 0.6ml TE buffer.

An equal volume of phenol/chloroform was added to the DNA solution and the mixture vortexed thoroughly for 30 seconds before centrifugation at 13,000 rpm for 1 minute. The upper aqueous layer was transferred to a microfuge tube containing an equal volume of chloroform. The mixture was again mixed thoroughly and subsequently centrifuged at 13,000 rpm for 1 minute. The upper aqueous layer was again removed to a microfuge tube and 2 volumes of ethanol and 0.1 volumes of 3m NaOAc were added before incubation at -20°C overnight. The sample was then centrifuged for 10 minutes at 13,000 rpm, the supernatant removed and the pellet dried under vacuum for 5 minutes before resuspension in 100µl dH_20. The concentration of the plasmid DNA was determined by measuring the absorbance at a wavelength of 260nm (A_{260}) assuming that 1 A_{260} = 50µg/ml for double stranded DNA.

Solution 1: 50mM glucose, 25mM Tris-HCl (pH8.0), 10mM EDTA (pH8.0).

Solution 2: 0.2M NaOH, 1% (w/v) SDS, freshly made.

Solution 3: 250g KOAc, 150g glacial acetic acid made up to 1 litre with dH_2O.

TE buffer: 10mM Tris-HCl (pH 8.0), 1mM EDTA (pH 8.0).

RNaseA solution: 10mg/ml pancreatic RNaseA in 10mM Tris-HCl (pH 7.5), 15mM NaCl, heated at 100°C for 15 minutes to inactivate DNases, cooled and stored at -20°C.

Proteinase K solution: 10mg/ml proteinase K dissolved in dH_2O.

2.2.5. Restriction enzyme digestion of DNA.

Restriction enzyme digestion of plasmid DNA was done in 20µl containing 2µl of miniprep/maxiprep DNA (approximately 1µg), 2µl relevant 10x reaction buffer, 1 unit of restriction enzyme and 0.5µl of 10mg/ml RNase A. The reactions were carried out at the temperature appropriate for the restriction enzyme for 2-4 hours. After digestion 2µl of loading dye were added and the samples analysed by agarose gel electrophoresis through a non-denaturing 1% TBE gel.

Digestion of DNA for cloning purposes, creation of vector backbones, or of inserts, was carried out in larger volumes, 50-100µl, with the same relative concentrations of enzyme and buffer. Digestion was carried out for a similar time period as illustrated above. The digestion products were fractionated through a 1-1.5% agarose TAE gel, depending on the size of the DNA band in question. DNA fragments were subsequently purified as in 2.2.9.

Partial digestion of DNA was occasionally necessary when an insert or vector contained more than one restriction enzyme site when digestion at one site only was desired. A 100µl reaction was set up containing all the necessary ingredients with the exception of the relevant restriction enzyme. Five 1.5ml microfuge tubes were placed on ice and numbered 1-5. Thirty microlitres of the reaction mix were transferred to microfuge tube 1, 20µl into each of microfuge tubes 2 to 4 and the remaining 10µl removed to microfuge tube 5. All 5 microfuge tubes were returned to ice. The required quantity of restriction enzyme (2-5 units/µg DNA) was added to microfuge tube 1, the tube mixed thoroughly and returned to ice. Ten microlitres from the sample in tube 1 was added to tube 2, tube 2 was mixed thoroughly and then returned to ice. This procedure was repeated until tube 5, after which each microfuge tube would contain a 20µl reaction sample. Each tube was incubated at the relevant temperature for restriction enzyme digestion for 15 minutes before fractionation through a TAE agarose gel and purification of the required DNA fragment.

TAE: 40mM Tris-acetate (pH 8.0), 1mM EDTA.

TBE: 90mM Tris-HCl (pH 8.0), 90mM boric acid, 1mM EDTA (pH 8.0).

2.2.6. Dephosphorylation of linearised plasmid DNA.

Removal of the 5′ terminal phosphates of digested vectors was carried out using calf intestinal alkaline phosphatase (CIP). Reactions were typically carried out in a total volume of 50µl containing 5µl 10x CIP

buffer (Boehringer Mannheim), 1 unit of CIP per 5µg of DNA and made up to volume with dH_2O. The reaction was incubated at 37°C for 60 minutes before electrophoresis through a 1% TAE agarose gel, purification of linearised, dephosphorylated vector by silica matrix adsorption (2.2.9),and stored at a concentration of 100ng/µl for ligation reactions.

2.2.7. End repair of DNA fragments.

Linearised vectors and DNA restriction enzyme fragments were made blunt-ended using T4 DNA polymerase. For PCR products the presence of T4 polynucleotide kinase was included to phosphorylate the ends of the DNA product. A standard protocol for the blunt-ending and kinasing reaction used a reaction volume of 100µl containing 10µl T4 polynucleotide kinase buffer, 10 units T4 polynucleotide kinase, 10 units T4 DNA polymerase, 5µl of a 10mM mix of the 4 dNTPs, 1µl 100mM ATP and made up to volume with dH_2O. The reaction mix was mixed thoroughly and incubated at 37°C for 60 minutes before analysis by agarose gel electrophoresis through a 1-1.5% agarose TAE gel and purification of the desired DNA fragment by silica matrix adsorption (2.2.9).

T4 polynucleotide kinase buffer (10x): 700mM Tris-HCl (pH7.6), 100mM $MgCl_2$, 50mM DTT (NEB).

2.2.8. Agarose gel electrophoresis of DNA.

Electrophoresis of DNA was carried out in horizontal slab gels [BRL gel electrophoresis apparatus Model H5 (14 x 11 x 0.5cm) and BRL gel electrophoresis Horizon 58 (5.7 x 8.3 x 0.3cm)] containing 1-1.5%(w/v) agarose in 1 x TAE or 1 X TBE containing 0.5µg/ml ethidium bromide. DNA samples contained 0.1 volumes of loading dye. Electrophoresis was carried out at 50-150V in 1 x TAE or 1 x TBE buffer (2.2.5) containing 0.5µg/ml ethidium bromide.

Agarose gel loading dye: 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 10% (w/v) Ficoll.

2.2.9. Purification of DNA from agarose gels by silica matrix adsorption.

The solutions used in this procedure were either part of the GeneClean (>200bp DNA fragments) or MerMaid (<200bp DNA fragments) kits. DNA fragments resolved by agarose gel electrophoresis were visualised by long-wave UV illumination and the DNA fragments excised in a gel slice, transferred to an 1.5ml microfuge tube containing 3 volumes of NaI solution and incubated at 55°C until the agarose slice has completely melted. A silica matrix solution (Glassmilk for DNA fragments >200bp or Glassfog for DNA fragments <200bp) was added, 5µl for up to 5µg of DNA and 1µl for every additional µg of DNA and the mixture mixed thoroughly for 10 minutes at room temperature to bind the DNA to the silica matrix. The sample was centrifuged at 13,000 rpm for 5 seconds, the supernatant removed and the pellet washed 3 times with an ethanol-based wash solution through repeated resuspension/centrifugation steps. The DNA was eluted from the pellet in 10-100µl dH_2O, incubated at 55°C for 5 minutes and centrifuged for 2 minutes to pellet the silica matrix. The DNA solution was removed to a microfuge tube and stored at -20°C.

2.2.10. Ligation of DNA fragments.

Ligation reactions were carried out in a reaction volume of 20µl containing 4µl 5x ligation buffer (Gibco BRL), 1 unit T4 DNA ligase, 100ng linearised vector, DNA insert and made up to final volume with dH_2O. DNA inserts were added in 1-14µl volumes depending on insert size and concentration (a 1:1 molar ratio of vector:insert was calculated and insert volumes to achieve this value used) and the reactions incubated at room temperature overnight before transformation of competent cells.

2.2.11. Preparation of competent bacterial cells.

Ten millilitres of L-broth (2.2.3) were inoculated with a single colony of *E. coli* strain DH5α and incubated with shaking at 37°C overnight. One ml of this overnight culture was added to 50ml of L-broth in a 250ml conical flask and shaken at 37°C until the cells reached an A_{550} of 0.4. The cells were centrifuged at 2,500 rpm for 5 minutes, the supernatant removed and the pellet gently resuspended in 20ml of ice cold 0.1M $CaCl_2$. The suspension was placed on ice for 30 minutes, centrifuged at 2,500 rpm for 5 minutes after which the pellet was resuspended in 5ml of ice cold 0.1M $CaCl_2$ and stored at 4°C.

2.2.12. Transformation of competent cells with plasmid DNA.

DNA was added to a 1.5ml microfuge tube containing 200µl of competent cells (see 2.2.11) and the solution gently mixed before incubation on ice for 30 minutes. The tube was then placed at 42°C for 90 seconds before the addition of 600µl of L-broth (2.2.3) and placed at 37°C for 60 minutes. Four hundred µl of the transformation mix were then spread onto an LB agarose plate containing 100µg/ml ampicillin and incubated overnight at 37°C to allow the formation of single colonies.

LB agarose: L-broth plus 1.5% (w/v) agarose.

2.2.13. Double stranded DNA sequencing.

Sequencing of double stranded DNA was based on the dideoxynucleotide chain termination method (Sanger *et al*, 1977) as modified by Tabor and Richardson (1987). A modified T7 DNA polymerase lacking 3' to 5' exonuclease activity was used in the method described below.

Eight µl (2-5µg) of plasmid DNA were added to a 1.5ml microfuge tube containing 2µl of 2M NaOH and the mixture incubated at room temperature for 10 minutes. Three microlitres of 3M NaOAc, 7µl dH_2O and 60µl ethanol were added, the mixture mixed thoroughly and subsequently placed on dry ice for 15 minutes. The sample was then centrifuged at 13,000 rpm for 10 minutes, the supernatant removed and the pellet washed with ice cold 70% ethanol before re-centrifugation at 13,000 rpm for 5 minutes. The supernatant was removed and the pellet dried briefly under vacuum before resuspension in 10µl dH_2O. Two µl of annealing buffer and 2µl of primer solution (5ng/µl) were added and the sample incubated at 37°C for 20 minutes, then allowed to cool to room temperature for at least 10 minutes. Six µl of labelling reaction mix (containing 2µl labelling mix, 1µl 0.3M DTT, 0.5µl $\alpha^{35}S$ dATP and 2 units of T7 DNA polymerase diluted in enzyme dilution buffer supplied with the enzyme) were added to the DNA-primer mix and incubated at room temperature for 5 minutes. During the labelling reaction 2.5µl of the four dideoxy termination mixtures (G, A, T, C) were added to four separate wells respectively in a 96-well, round bottomed microtitre plate and pre-heated at 37°C for at least 1 minute. Four and a half microlitres of the labelled DNA-primer mix were added to each of the four termination mixes, briefly mixed and centrifuged before incubation at 37°C for 5

minutes after which the reactions were arrested by the addition of 5µl of STOP mix.

The samples were heated at 75-80°C for 2 minutes and then 3µl of each were loaded onto a 6% polyacrylamide sequencing gel (80ml 6% sequencing solution, 800µl APS, 80µl TEMED) in the order G, A, T, C. The samples were electrophoresed at a constant power of 70 watts, the gel transferred to 3MM paper and dried under vacuum for 2 hours at 80°C. The dried gel was then exposed directly onto an X-Omat S film and incubated at room temperature for 24-48 hours.

Sequencing annealing buffer: 280mM Tris-HCl (pH7.5), 100mM $MgCl_2$, 350mM NaCl.

Sequencing dideoxy termination mix: 150mM each dNTP, 10mM $MgCl_2$, 40mM Tris-HCl (pH 7.5), 50mM NaCl, 15µM of relevant dideoxynucleotide.

Sequencing labelling mix: 2µM dGTP, 2µM dCTP, 2µM dTTP.

Sequencing STOP mix: 95% (v/v) deionized formamide, 20mM EDTA (pH 7.5), 0.005% (w/v) xylene cyanol, 0.005% (w/v) bromophenol blue.

6% sequencing acrylamide solution: 6% (w/v) acrylamide/0.3% (w/v) bis-acrylamide, 7M urea, 1x TBE (ratio 19:1) [Scotlab].

2.2.14. Oligonuceotide directed mutagenesis.

Complementary oligonucleotides were designed so that when annealed they would form a *Kpn* I/*Bst* EII cassette for ligation into cut vector. One hundred µg of each Oligonucleotide (1µl) were added to a 0.5ml microfuge tube containing 30µl 150mM NaCl and incubated at 65°C for 10 minutes and slowly cooled to room temperature. The annealed oligonucleotides were then phosphorylated in a 100µl reaction containing 10µl annealed oligonucleotides, 10µl 10x T4 polynucleotide kinase buffer, 10µl 100mM ATP, 5µl T4 polynucleotide kinase and 65µl dH_2O for 1 hour at 37°C. The phosphorylated partially double stranded oligonucleotide was then purified by silica matrix adsorption (2.2.9) and ligated into cut plasmid vector (2.2.10).

2.3. Expression and purification of 6-His tagged proteins.

The protocols in this section were carried out based on those found in *The QIAexpressionist*, **2nd edition**, 1992 (Qiagen Inc.).

2.3.1. Rapid screening and small-scale expression cultures.

The production of competent *E. coli* strains M15[pREP4] and SG13009[pREP4] and their subsequent transformation was undertaken as described in 2.3.11 and 2.3.12. A 20ml tube containing 1.5ml L-broth (2.2.5) supplemented with 100mg/ml ampicillin and 25mg/ml kanamycin was inoculated with a single colony and grown overnight at 37°C with shaking. One and a quarter millilitres of the overnight culture were added to a 20ml tube containing 8.75 ml, pre-warmed L-broth (plus ampicillin/kanamycin as above) and incubated at 37°C with shaking for 60 minutes until the A_{600} was 0.7-0.9. Subsequently 1ml was removed and stored (uninduced fraction) before the remaining culture was induced with IPTG and returned to 37°C for 3-5 hours. Another 1ml sample was removed (induced fraction) before centrifugation at 3,000 rpm for 3 minutes and removal of the supernatant.

The cell pellet was resuspended in 200µl buffer A (2.3.4) and the cells lysed by gentle mixing. The lysate was subsequently centrifuged at 13,000 rpm for 10 minutes before removal of the supernatant to a fresh tube. Fifty microlitres of a 50% slurry of Ni-NTA resin were added to the supernatant and the sample mixed at room temperature for 30 minutes. The sample was then centrifuged at 13,000 rpm to pellet the resin and the supernatant discarded. The resin was washed 3 times with buffer C (2.3.4) before the addition of 20µl buffer C and gentle mixing for 2 minutes to elute the protein from the resin. The mixture was pelleted at 13,000 rpm for 10 seconds before careful removal of the supernatant to a fresh tube containing 5µl 5x SDS PAGE sample buffer. Five x SDS PAGE buffer was added to the previous samples taken and all samples analysed on a denaturing, 12% SDS polyacrylamide gel (see 2.6.3).

5x SDS PAGE buffer: 15% (v/v) β-mercaptoethanol, 15% (w/v) SDS, 1.5% (w/v) bromophenol blue, 50% (v/v) glycerol.

2.3.2. Large scale expression of protein.

A 20ml tube containing 10ml L-broth (plus 100µg/ml ampicillin and 25µg/ml kanamycin) was inoculated with a single bacterial colony or 100µl

of a glycerol stock and grown at 37°C overnight with shaking. Fifty millilitres of L-broth (containing ampicillin/kanamycin as above) were inoculated with 1ml of the uninduced overnight culture and grown at 37°C with shaking until the A_{600} was 0.7 - 0.9. IPTG was added to a final concentration of 1-2mM after which the culture was returned to 37°C for 5 hours. The cells were pelleted at 4,000g (3,000 rpm) using an SW 27 rotor for 10 minutes, the supernatant removed and the pellet stored at -70°C.

2.3.3. Determination of cytosolic or periplasmic location of expressed protein.

A 100ml culture was grown as described in 2.3.2 except 1ml of the culture was removed before induction with IPTG (uninduced control) and 1ml removed before centrifugation (induced control). These two samples were centrifuged at 13,000 rpm and the pellets resuspended in 50µl of 5x SDS-PAGE buffer. The remainder of the 100ml culture were separated into two aliquots and the cells pelleted as in 2.3.2.

To check for cytosolic location one sample was resuspended in 5ml sonication buffer, frozen in dry ice/ethanol and subsequently thawed in cold water. The cells were sonicated briefly to induce lysis before centrifugation at 9,000 rpm for 20 minutes. The supernatant was removed and stored (soluble protein extract) and the pellet resuspended in 5ml sonication buffer (insoluble protein extract).

To check for periplasmic location the remaining pellet was resuspended in 10ml 30mM Tris-HCl (pH 8.0)/20% sucrose before addition of EDTA to 1mM and incubation at room temperature for 5-10 minutes with gentle mixing. The sample was centrifuged at 8,000 rpm for 10 minutes at 4°C, the supernatant removed and the pellet resuspended in 10ml ice-cold 5mM $MgSO_4$ before stirring for 10 minutes in an ice/water bath. The sample was recentrifuged at 8,000 rpm for 10 minutes at 4°C and the supernatant stored (periplasmic extract). Ten µl of all three extracts were removed to fresh tubes containing 5µl 5x SDS-PAGE buffer and the samples heated at 100°C for 5 minutes (as well as the uninduced and induced samples) before loading on a 12% SDS polyacrylamide gel (see 2.6.3).

Sonication buffer: 50mM Na-phosphate (pH 7.8), 300mM NaCl

2.3.4. Denaturing purification of insoluble proteins.

A 500 ml culture was grown and induced as described in 2.3.2 before centrifugation at 9,000 rpm for 20 minutes. The supernatant was removed

and the pellet resuspended in Buffer A at 5ml per gram before stirring at room temperature for 60 minutes. The lysate was then centrifuged at 9,000 rpm for 15 minutes at 4°C and the supernatant removed to a fresh tube containing 8ml of a 50% slurry of Ni-NTA resin previously equilibrated in Buffer A. The sample was mixed at room temperature for 45 minutes before loading into a 4ml column pre-equilibrated with buffer A. The column was then washed with 10 column volumes (40ml) of Buffer A and 5 column volumes of Buffer B. The column was then washed with Buffer C until the A_{280} was <0.01.

The protein was eluted with 15ml Buffer D followed by 15ml Buffer E and then 15ml Buffer F, collecting 3ml fractions during each elution. Twenty µl of each sample were added to a 1.5ml microfuge tube containing 50µl 5X PAGE buffer and half of this sample analysed by SDS-PAGE (2.7.3)

Buffer A : 6M guanidine hydrochloride, 100mM NaH_2PO_4, 10mM Tris-HCl, pH adjusted to 8.0.

Buffer B: 8M urea, 100mM NaH_2PO_4, 10mM Tris-HCl, pH adjusted to 8.0.

Buffer C: Same composition as Buffer B except pH 6.3.

Buffer D: Same composition as Buffer B except pH 5.9.

Buffer E : Same composition as Buffer B except pH 4.5.

Buffer F: 6M guanidine hydrochloride, 0.2M acetic acid.

2.4. Culture of cell lines and viruses.

2.4.1. Cell Culture.

Confluent monolayers of CV-1 cells were washed with 20ml of versene followed by 25ml trypsin/versene (1:5). The cells were subsequently incubated at 37°C for 5 minutes before suspension in 10ml DMEM-10% FCS. One millilitre (approximately 1 x 10^6 cells) was added to a 175cm^2 cell culture flask containing 50ml of DMEM-10%FCS and incubated at 37°C until confluent.

Trypsin solution: 0.25% (w/v) Difco trypsin dissolved in Tris-saline solution plus 0.005% (w/v) phenol red.

Tris-saline: 140mM NaCl, 30mM KCl, 280mM Na_2HPO_4, 1mg/ml dextrose, 25mM Tris-HCl (pH7.4).

Versene solution: PBS supplemented with 0.6mM EDTA and 0.0015% (w/v) phenol red.

2.4.2. Growth and purification of vTF7.3.

A confluent 175cm^2 flask of CV-1 cells was inoculated with 10µl of vTF7.3 stock (1x10^6 pfu/µl) in 5ml DMEM-5% FCS and incubated at 37°C for 1 hour with agitation every 15 minutes. The medium was replaced with 20ml DMEM-5% FCS and returned to 37°C for approximately 2 days until cytopathic effects were visible. The infected cells were centrifuged at 3,000 rpm and resuspended in 4ml 10mM Tris-HCl pH 9.0 at 4°C. The cells were disrupted by 3 cycles of freeze thawing (5 minutes in dry ice, 5 minutes at 37°C) before the nuclei were pelleted at 10,000 rpm for 5 minutes at 4°C and the supernatant removed to a sterile tube. The pellet was resuspended in 4ml 10mM Tris-HCl pH 9.0, recentrifuged and the two supernatants combined. One tenth volumes of trypsin solution (2.5mg/ml) was added and the sample incubated at 37°C for 30 minutes with frequent mixing. Four millilitre aliquots were then layered onto 12ml 36% (w/v) sucrose in 10mM Tris-HCl pH9.0 and centrifuged at 13,000 rpm in an SW29 rotor for 80 minutes at 4°C. The supernatant was discarded and each pellet was resuspended in 1ml of 1mM Tris-HCl pH 9.0, aliquoted and stored at -70°C.

2.4.3. Titration of vTF7.3 stock virus.

Thirty-five mm petri dishes containing 1 x 10^6 CV-1 cells were infected with dilutions of vTF7.3. Dilutions were carried out in a total volume of 1ml PBS and duplicate 100µl volumes (for each dilution from 10^{-5} to 10^{-9}) were used for titration. The cells were incubated at 37°C for 60 minutes to allow the virus to adsorb with gentle agitation every 15 minutes. The inoculum was removed, 2ml DMEM-5%FCS added and the cells incubated at 37°C for 2 days. Then the medium was removed and the cells gently covered with CIDEX (a commercial glutaraldehyde preparation diluted 1:1 with PBS) for 3-4 hours to fix the cells. The CIDEX was removed and the cells stained with Giemsa for 10 minutes before washing with tap water. Virus plaques were visible as clear regions within the purple-stained cell monolayer.

PBS: 170mM NaCl, 3.4mM KCl, 10mM HPO_4, 1.8mM KH_2PO_4 (pH 7.2), 6.8mM $CaCl_2$, 4.9mM $MgCl_2$.

Giemsa stain: 1.5% (w/v) Giemsa in glycerol, heated at 50°C for 2 hours and diluted with an equal volume of methanol.

2.5. Introduction of nucleic acids into mammalian cells.

2.5.1. Preparation of cationic liposomes.

Cationic liposomes were prepared using a method based on that described by Rose *et al.* (1991). Dimethyldioctadecyl ammonium bromide (DDAB) was diluted in chloroform to 4mg/ml and 1ml added to a 20ml tube containing 1ml of dioleoyl-α-phopsphatidyl ethanolamine (DOPE: 10mg/ml in chloroform). After mixing, the chloroform was removed by evaporation with a gentle stream of nitrogen. Lyophilization was carried out for a further 4 hours to remove any further traces of chloroform and the dried lipids resuspended in 10ml sterile dH_2O through sonication in a sonicating water bath. The lipids were further sonicated on ice using a soniprobe set at maximum power until the suspension cleared. The final lipid solution had a ratio 1:2.5 wt/wt for DDAB:DOPE and was stored at 4°C for up to one month.

2.5.2. Liposome mediated transfection of nucleic acids into mammalian cells.

Sub-confluent monolayers of CV-1 cells in 35mm petri dishes (approximately $5x10^5$ cells per dish) were infected with the recombinant vaccinia virus vTF7.3 using a m.o.i. of 5 p.f.u. per cell for 30 minutes at 37°C. The inoculum was removed and the cells washed with 1ml of OptiMem before addition of the DNA transfection mix. The required plasmid DNA (1-20 μg) in 250 μl OptiMem was combined with 15μl of liposomes in 250μl OptiMem and the tube mixed thoroughly. This transfection mix was left to stand for 10-20 minutes before addition to the cell monolayer and incubation at 37°C for 3 hours. One ml of DMEM-10% FCS was added and the cells returned to 37°C for 12 hours.

The medium was replaced with 2ml of medium lacking methionine for 30 minutes at 37°C before replacement with 1ml of methionine free medium containing 50μCi of [^{35}S] L-methionine/ml for 2 hours at 37°C. After radio-labelling the cells were washed in ice cold PBS (2.4.3) before the addition of either 200μl protein dissociation mix (cell

lysate sample) or 100µl 0.5% SDS to lyse the cells. The cell lysate was then removed using a 2ml syringe and transferred to a 1.5ml microfuge tube for storage at -20°C or immunoprecipitation.

Methionine free medium: Eagles A, Eagles B without methionine.

Eagles A: 0.23g $CaCl_2$, 0.23g $MgSO_4$, 0.1ml conc. HCl in 1 litre dH_2O.

Eagles B without methionine: 500ml salts, 400ml amino acids minus methionine, 32ml vitamins made up to 1 litre with dH_2O.

Salts: 102.4g NaCl, 6.4g KCl, 2.4g, NaH_2PO_4, 72g glucose, 16ml 0.01% ferric nitrate, 4.68g L-glutamine, 16ml (1,600,000 units) penicillin, 1.6g streptomycin, 16ml 0.02% antimycotic made up to 1 litre with dH_2O.

Amino acids minus methionine: 0.84g L-arginine mono HCl, 0.384g L-histidine mono HCl, 1.048g L-isoleucine, 1.048g L-leucine, 1.462g L-lysine mono HCl, 0.66g L-phenylalanine, 0.952g L-threonine, 0.16g L-tryptophan, 0.724g L-tyrosine, 0.936g L-valine, 0.07g inositol, 0.48g L-cysteine, 55g $NaHCO_3$ dissolved in 700ml dH_2O and heated to 50°C before addition of 30ml 1% phenol red and made up to a final volume of 1 litre with dH_2O.

Vitamins: 2g choline chloride, 2g folic acid (dissolved in 200ml 1M NaOH), 2g nicotinamide, 2g pantothenic acid/Ca salt, 2g pyridotal HCl, 2g thiamine HCL (aneurine), 0.2g riboflavine, made up to 4 litres with dH_2O.

Protein dissociation mix: 100mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 200mM β-mercaptoethanol, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue.

2.5.3. Polyethlyleneimine mediated transfection of nucleic acids into mammalian cells.

This transfection protocol is taken from that described in Boussif *et al.* (1995). The transfection of nucleic acids into mammalian cells using polyethyleneimine (PEI) used the same basic method as illustrated above (2.5.2) with the only difference being the preparation of the DNA transfection mix. The required plasmid DNA (1-20µg) was added to a 1.5ml microfuge tube containing 100µl 150mM NaCl. Sterilised PEI solution was added to another 1.5ml microfuge tube also containing 100µl 150mM NaCl. PEI 800kDa (Fluka) was used as a 10mM monomer aqueous stock solution [9mg of the 50%(w/v) commercial solution diluted in 10ml of dH_20] and was neutralised with HCl before filtration through a 0.2µm Millipore. The amount of PEI added was directly proportional to the amount of plasmid DNA to be introduced (3µl of PEI per 1µg plasmid DNA). The two microfuge tubes were mixed thoroughly and left for 10 minutes at room temperature, the contents combined, mixed thoroughly and left for 10 minutes at room temperature. The DNA transfection mix was then added to vTF7.3 infected cells (after the 1ml OptiMem wash) with 500µl OptiMem and the cells incubated at 37°C for 4 hours. The rest of the method is exactly as described above (2.5.2).

2.6. Protein Analysis.

2.6.1. Immunoprecipitation of proteins.

This method of immunoprecipitation was based on a method by Kolykhalov *et al.* (1994). The cell lysate (2.5.2) was heated at 100°C for 5 minutes, diluted 1:5 in immunoprecipitation (IP) buffer and stored on ice for 30 minutes to 2 hours. The sample was then centrifuged at 13,000rpm for 15 minutes to pellet the cell debris and the supernatant removed to a 1.5ml microfuge tube. Fifty microlitres of formalin-fixed *S. aureus* bearing protein A (Pansorbin) was added to the supernatant and the sample mixed with rotation at 4°C for 30 minutes before re-centrifugation at 13,000rpm for 5 minutes. The supernatant was gently removed, transferred to a microfuge tube containing 2µl of specific monoclonal antibody and rotated overnight at 4°C.

Sixty microlitres of Protein A Sepharose suspension (100mg in 800µl IP buffer) was added to the sample and a further 30 minute rotation at 4°C carried out The sample was then centrifuged at 13,000rpm for 30 seconds, washed three times in 0.5M LiCl/0.1M Tris pH 7.5 and 70µl protein

dissociation mix added. The sample was stored at -20°C or analysed by SDS-PAGE (2.6.3).

Immunoprecipitation (IP) buffer: 50mM Tris-HCl (pH 7.6), 200mM NaCl, 1mM EDTA, 0.1% (v/v) Triton X-100, 1mg/ml BSA.

2.6.2. Coupled *in vitro* transcription/translation reactions.

Cell free protein synthesis was carried out using the TnT coupled transcription/translation system from Promega. The reactions were set up on ice in accordance with the protocol supplied.

TnT Rabbit reticulocyte lysate	12.5µl	
TnT Reaction buffer		2.0µl
TnT T7 RNA polymerase	0.5µl	
Amino acid mixture minus methionine		0.5µl
[^{35}S] L-Methionine	2.0µl	
RNasin (40 units/µl)	0.5µl	
DNA template (1µg)	yµl	
Nuclease free dH_2O (to a final volume of 25µl)		xµl

The reaction was carried out at 30°C for 60 minutes and then 70µl protein dissociation mix (2.5.2) were added. The sample was stored at -20°C or analysed by SDS-PAGE (2.6.3)

2.6.3. SDS polyacrylamide gel electrophoresis (SDS-PAGE) of proteins.

Proteins were separated by electrophoresis through a polyacrylamide gel containing SDS with the discontinuous buffer system of Laemmli (1970). The gel apparatus and glass plates were assembled following the manufacturers instructions. A 12% acrylamide resolving gel mix was used containing 14.4ml 30% acrylamide solution, 16.8ml 1M Tris-HCl pH 8.8, 12.9ml dH_20, 450µl 10% SDS, 450µl 10% APS and 50µl TEMED, with the TEMED added immediately before pouring the gel. A gap at the top of the gel plate was left to allow for addition of the stacking gel (approximately twice the depth of the comb). The resolving gel was immediately overlaid with 1ml isopropanol to leave a smooth interface after polymerisation. After the gel had polymerised the surface isopropanol was poured off, the stacking gel (1ml 30% acrylamide solution, 630µl 1M Tris-HCl pH 6.8, 8ml

dH_2O, 100µl 10% APS, 100µl 10% SDS and 10µl TEMED) was added on top of the set resolving gel surface and a comb inserted to form loading wells for the protein samples. After polymerisation the comb was removed, the gel placed in the electrophoresis apparatus and the reservoirs filled with running buffer.

Protein samples were heated at 100°C for 5 minutes before loading into the wells, as were low range molecular weight protein markers (to allow determination of the protein sizes of the samples). The gel was run at 45mA until the bromophenol blue reached the bottom of the resolving gel after which the gel was removed from the apparatus and immersed in Gel Fix for 1 hour. The Gel Fix was then replaced by En^{3}hance for 45 minutes and then the gel washed 3 times with dH_2O before a final 45 minute soak in dH_2O. The gel was then dried under vacuum for 60 minutes and placed in direct contact with X-Omat film S and exposed at -70°C for 24-96 hours.

For non-radioactively labelled gels the gel was first immersed in 0.05% (w/v) Coomassie blue in Gel Fix and destained in Gel Fix for 3-4 hours before drying under vacuum.

30% acrylamide solution: 30% (w/v) acrylamide/0.8% bis-acrylamide (ratio 37.5:1) [Scotlab].

Gel Fix: 50% (v/v) methanol, 10% (v/v) acetic acid, 40% (v/v) dH_2O.

2.6.4. Quantitation of cleavage efficiency.

A dried SDS-PAGE gel was placed in direct contact with a Molecular Dynamics phosphor screen and placed at room temperature overnight. The screen was loaded into a Molecular Dynamics PhosphorImager and the image incorporated into a computer file. The ImageQuant software was used to analyse the scanned image. A background value was taken before determination of both uncleaved substrate and cleavage product band intensities. The size of the background value box was the same size as the boxes used for substrate and cleavage products. The background value was subtracted from the experimental values before any calculations were carried out. The efficiency of cleavage was calculated as a percentage as illustrated below.

$$\frac{\text{Cleavage product A + cleavage product B}}{\text{Cleavage products + uncleaved substrate}} \times 100$$

Chapter 3 - Construction of HCV protein expression clones

3. Introduction.

The aim of this study was to characterise the HCV NS3 serine protease. To study the catalytic activity and requirements of the NS3 protease an efficient system for NS3 mediated cleavage was necessary. Once an assay of this nature had been developed the characterisation of substrate and protease requirements could be carried out.

Before a protease assay could be developed sources of the HCV serine protease and a suitable HCV substrate were required. The HCV serine protease is encoded within the N-terminal region of the NS3 protein (Hijikata *et al.*, 1993a; Bartenschlager *et al.*, 1994; Lin *et al.*, 1994b; Tanji *et al.*, 1994b; Failla *et al.*, 1995). It had been demonstrated that the NS3 protein associates with the NS4A protein to form a complex with optimal catalytic activity (Failla *et al.*, 1994; Bartenschlager *et al.*, 1994). A construct that expressed the NS3 serine protease was required and in addition, it was considered that inclusion of the NS4A protein would result in a more efficient protease.

The first cleavage site processed by the NS3 protease *in trans* during polyprotein processing is the NS5A/NS5B cleavage site. This processing event is efficient and quick. For these reasons the NS5 protein was chosen as a substrate for the *trans*-cleavage assay.

This chapter describes approaches to generate the reagents required for a *trans*-cleavage assay for the NS3 serine protease. The protease was expressed in bacteria, by *in vitro* transcription/translation and transiently in transfected mammalian cells. The substrate was expressed by *in vitro* transcription/translation or transiently in transfected mammalian cells.

3.1. HCV cDNA clones available in the laboratory.

A number of HCV plasmid constructs were available in the laboratory when I started the project. Two plasmids encoding the NS3 protease region had been constructed: the full length NS3 gene (encoding amino acids 1027 to 1657 of the HCV Gla polyprotein) and the NS34A gene (encoding amino acids 1027 to 1709) had been cloned, under control of the T7 promoter, into pTF7.5. The NS34A construct contained the full NS3 gene and almost the entire NS4A region with just 2 amino acids missing from the C-terminus of NS4A. The gene encoding the full length NS5 protein (consisting of amino acids 1973 to 3011 of the HCV Gla polyprotein), under

control of the T7 promoter, was cloned into pTZ18 and pTM1. pTM1 contains the EMCV IRES element that directs cap-independent translation and results in a higher level of protein expression. These constructs are illustrated in Fig. 3.1.

3.2. Construction of substrate clones.

It had been reported that the NS5B protein of HCV was unstable when expressed in mammalian cells (Lin *et al.*, 1994b). The NS5B protein is not further processed by HCV-encoded enzymes, and thus may be degraded by a cellular pathway. To determine the level of processing with respect to the uncleaved substrate (2.6.4) it would be more accurate if both cleavage products were identifiable. Quantitation of the level of processing was calculated by dividing the total of the cleavage product intensities by the total of cleaved and uncleaved band intensities. The intensity of each band was determined by the number of (radioactively labelled) methionine residues within the protein. The sum of the intensities of the two cleavage products would equal the intensity of the single substrate band (as the uncleaved substrate contains the same number of methionine residues as the two cleavage products). This method of quantitation is "quasi-quantitative" as the calculation uses steady state levels of substrate and products. If only one cleavage product was identifiable the estimation of processing would be increasingly inaccurate.

It was thought that an NS5 substrate truncated at both the N- and C-termini might produce a stable NS5B product that was less susceptible to degradation. A truncated NS5 substrate (519 amino acids in length) consisting of residues 2105 to 2623 of the HCV Gla polyprotein was constructed. This substrate still contained the NS5A/NS5B cleavage site but, after processing by the NS3 serine protease, would give smaller cleavage products (including perhaps a stable NS5B protein).

No efficient antiserum against the NS5 region was available in the laboratory at the time so an epitope tag derived from the SV5 P and V proteins (Hanke *et al.*, 1992) was introduced at both termini of this truncated NS5 substrate. A specific monoclonal antibody (SV5-P-k) recognises this sequence and would enable identification of both cleavage products. The epitope tag consists of the 15 amino acid sequence:- **Ser Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr**

PCR primers were designed containing a suitable restriction enzyme site for cloning (*Xba* I or *Nco* I for the 5'primers and *Pst* I for the 3' primer) followed by the epitope tag sequence and 15 nucleotides complementary to the terminal sequence of the desired NS5 region. The nucleotide sequence encoding the epitope tag was changed between the 5' and 3' PCR primers (maintaining the required amino acid sequence) to ensure that no mispriming took place during PCR. The truncated, epitope-tagged NS5 substrate (designated tNS5T) was cloned into pTZ18 and subsequently into pTM1. The 3' PCR primer used was identical for production of both PCR

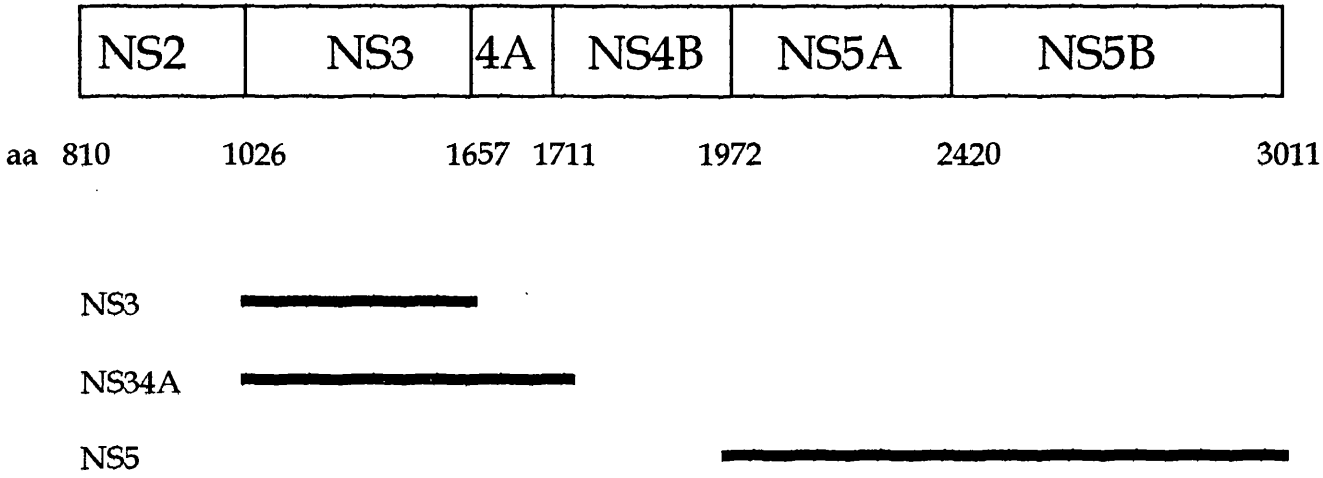


Fig. 3.1. Schematic representation of the HCV constructs available within the laboratory. The NS3 and NS34A cDNAs were cloned into pTF7.5. The NS5 cDNA was cloned into pTM1 and pTZ18. The amino acid positions above indicate the N-termini of the non-structural proteins. The black lines represent the regions encoded within the available HCV constructs .

products and contained a *Pst* I restriction enzyme site for cloning into both pTZ18 and pTM1. An *Xba* I site was used in the 5′ PCR primer for cloning into pTZ18 and an *Nco* I site used in the 5′ PCR primer for cloning into pTM1. The *Nco* I site contains an ATG codon that is used as the initiation codon for translation by the EMCV IRES element in pTM1. As the PCR producing the pTM1 tNS5T insert was carried out using pTZ18 tNS5T as a template a smaller 5′ PCR primer was sufficient. The PCR primers used are illustrated in Fig. 3.2.

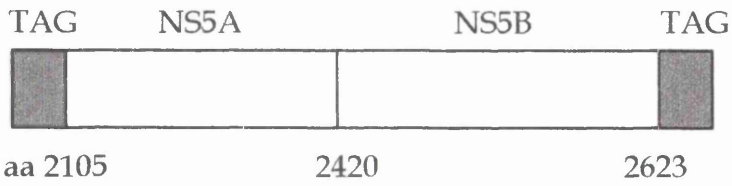
After cloning of correctly sized inserts the tNS5T constructs were sequenced from both the 5′ and 3′ ends to confirm that no mutations had been introduced during PCR

Attempts were made to produce a full length NS5 substrate with the epitope tag identified by the SV5-P-k antibody at either terminus. This would enable detection of a full length NS5B protein and confirm whether this protein was unstable. Attempts to produce this substrate were unsuccessful.

Having developed a truncated substrate (tNS5T) identifiable using the specific monoclonal antibody (SV5-P-k) two antibodies against the NS5 region were supplied by Dr B.E. Clarke (GlaxoWellcome). These antibodies could be used to identify both the truncated NS5 substrate and the full length NS5 substrate. These antibodies were used for immunoprecipitation of cell lysates expressing the NS5 construct. The ability of these antibodies to identify the HCV Gla NS5 protein and that of SV5-P-k to identify the tNS5T substrate is illustrated in Fig. 3.3. The NS5 protein was immuneprecipitated by the NS5(Baculo) antisera only and tNS5T was detected using the SV5-P-k antibody. Both NS5 proteins could be used for *trans*-cleavage analysis as the baculovirus-derived NS5 antibody (which recognises the C-terminus of NS5A) identifies the full length HCV Gla NS5 substrate and the SV5-P-k antibody identifies the tNS5T substrate.

3.3. Construction of protease expressing clones.

Although the HCV serine protease was shown to be encoded within the N-terminal region of the NS3 protein (Hijikata *et al.*, 1993a; Bartenschlager *et al.*, 1994; Lin *et al.*, 1994b; Tanji *et al.*, 1994b; Failla *et al.*, 1995) it was not known whether this region, expressed without the C-terminus of NS3, folded in a manner that still permitted catalytic activity. The DNA sequence encoding the protease domain (designated NS3P) would produce a protein consisting of the first 190 amino acids of NS3 (amino acids 1027 to 1216 of the HCV Gla polyprotein). This sequence was cloned into pTM1 and the bacterial expression vector pQE30. The NS3P sequence cloned into pTM1 would be suitable for expression *in vitro* or in mammalian cells. The NS3P sequence cloned into pQE30 would be used to express the protein fragment in bacterial cells and the protein product purified. This purified NS3P protein could be added to substrate expressed either *in vitro* or *in vivo* to determine whether it had suitable catalytic activity. The PCR primers



6653

GGGTCTAGACCATGGGAAAGCCGATCCCAAACCCTTTGCTGGGATTGGACTCCACCATGACTACTGACAAT

5′ tNS5T PCR primer (pTZ18) - *Xba I* site

GCGGCGCCATGGGAAAGCCGATCCCAAAC

5′ tNS5T PCR primer (pTM1) - *Nco I* site

8211

GGGCTGCAGCTACGTAGAATCTAACCCCAACAGGGGATTGGGAATGGGTTTACCTGAGAATTCAACCCGCTG

3′ tNS5T PCR primer - *Pst I* site

Fig. 3.2. PCR primers used for the amplification of a truncated NS5 region, tNS5T. The top diagram illustrates the tNS5T protein with amino acid positions from the HCV Gla polyprotein indicated. The nucleotides corresponding to restriction enzyme sites are illustrated in red, epitope tag in blue and NS5 sequence in green. The nucleotide position within the HCV Gla nucleotide sequence is illustrated for theNS5 sequence.

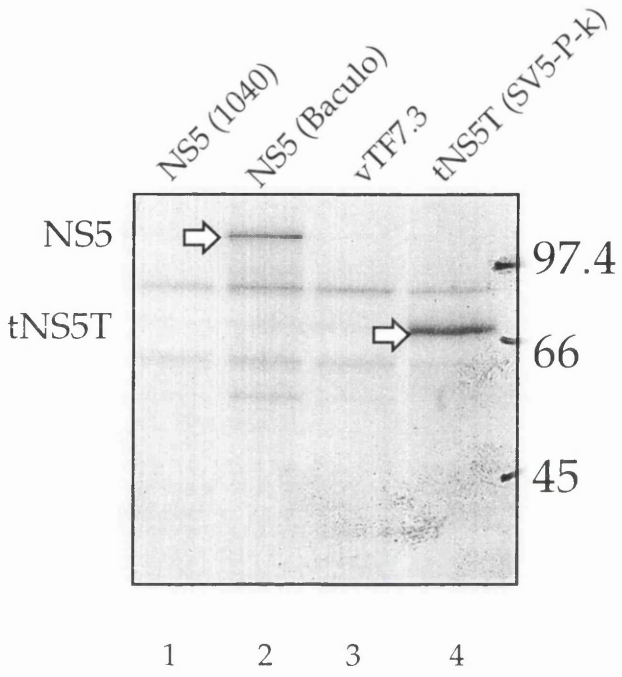


Fig. 3.3. Identification of full length and truncated NS5 proteins using different antisera. vTF7.3 infected CV-1 cells were transfected with pTM1 NS5 (lanes 1 and 2), pTM1 tNS5T (lane 4) or left untransfected (lane 3). Radiolabelled cell extracts were prepared 11 hours post-transfection and reacted with various antisera. The NS5 (1040) antisera (lane 1) failed to detect the full length NS5, whereas NS5 (Baculo) detected the full length NS5 (lane 2). The truncated NS5 substrate was precipitated by the SV5-P-k antibody (lane 4). The unstransfected cell lysate (lane 3) was immuneprecipitated with the SV5-P-k antibody, illustrating the background protein products remaining after immuneprecipitation. The NS5 proteins are indicated by an arrow. Molecular weight markers are located at the right of the gel.

used in the cloning of these constructs are illustrated in Fig. 3.4. The 3′ PCR primer was identical for both reactions and contained a *Pst* I site for cloning. The 5′ PCR primer used for cloning into pTM1 contained an *Nco* I site whereas a *Kpn* I site was used in the 5′ primer for cloning into pQE30.

Cloning of the NS3P PCR product into pTM1 required partial digestion with *Nco* I as an internal site was present within the HCV Gla NS3P nucleotide sequence. The PCR product was initially digested with *Pst* I to linearise the plasmid DNA and the vector purified by silica matrix adsorption (2.2.9.) before partial digestion with *Nco* I. The partial digest is illustrated in Fig. 3.5. The partial digest containing the 1:400 dilution of the enzyme shows a very faint band corresponding to the desired band. In the subsequent two reactions the band is more prominent. Once cloned the 5′ and 3′ ends of the insert were sequenced to ensure that no PCR-induced mutations had been introduced.

3.3.1. Expression of NS3P in bacteria.

After cloning of the NS3P gene into pQE30 the ends were sequenced to check that no PCR-induced mutations had been introduced. The construct was checked for protein expression in bacterial cells. After induction with IPTG the NS3P protein was expressed in both strains used (SG13009[pREP4] and M15[pREP4]). Both these bacterial strains allow high level expression. SG13009[pREP4] gives levels of protein expression that can be toxic to the cells, but can be useful for the expression of proteins that are expressed poorly in M15[pREP4]. The NS3P protein was expressed in both strains with no observable difference in the level of expression (Fig. 3.6.).

Proteins expressed in bacterial cells may be soluble in the cytoplasm, located within cytoplasmic inclusion bodies (insoluble) or secreted into the periplasmic space. Many proteins expressed at high levels in bacteria form inclusion bodies, while others are well tolerated by the cell and remain in the cytoplasm in their native configuration.

After expression of NS3P in bacteria, half the cells were pelleted, resuspended in sonication buffer (50mM Na-phosphate pH 7.8, 300mM NaCl) and freeze/thawed before lysis by sonication. The sample was pelleted and the supernatant removed. This supernatant would contain the protein if it was soluble within the cell. Once resuspended the pellet would contain the protein if it was insoluble. In addition to this analysis, the remaining cells were pelleted and resuspended in 30mM Tris-HCl, 20% sucrose (pH8.0) before addition of EDTA to 1mM and incubation with rotation at room temperature for 10 minutes. The sample was pelleted, the supernatant removed and the pellet resuspended in ice-cold 5mM $MgSO_4$ before stirring at 4°C for 10 minutes. The sample was pelleted and the supernatant removed. This supernatant would contain the protein if it was secreted into the periplasm (this protein containing sample is called the osmotic shock fluid). When all three samples were analysed by SDS-PAGE the protein was found to be insoluble (Fig. 3.7.).

3421

GCGCATGCCATGGCGCCCATCACAGCG

5′ NS3P primer (pTM1) - *Nco I* site

3421

GCGGGGGTACCGCGCCCATCACAGCG

5′ NS3P primer (pQE30) - *Kpn I* site

3990

GCGCCAATGCATTGGTTCTGCAGTGGAGAGGAGTTGTC

3′ NS3P primer - *Pst I* site

Fig. 3.4. PCR primers for the production of the protease domain of HCV NS3, designated NS3P. Restriction enzyme sequences are illustrated in red and sequences corresponding to HCV NS3 are illustrated in blue. The sequence preceding the restriction enzyme site was included to ensure efficient digestion by the enzyme. The numbers shown above the start of the blue nucleotide sequence indicate the HCV Gla nucleotide positions.

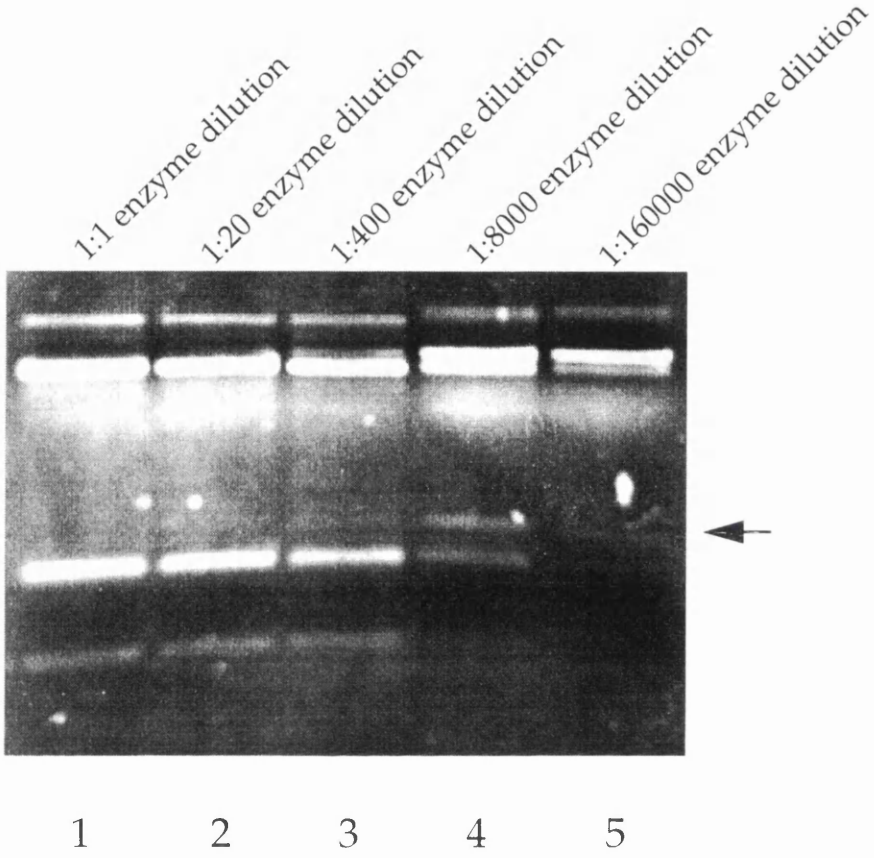


Fig. 3.5. Partial *Nco I* digest of the NS3P PCR fragment. The original PCR product was blunt-ended before ligation into *Sma* I cut pTZ18. This construct (pTZ18 NS3P) was subsequently digested with *Pst* I (to linearise the vector) and then purified by silica gel matrix adsorption. Partial digests using *Nco* I were set up and the reactions incubated at 37°C for 15 minutes. After incubation the digests were analysed through a 1% TAE gel. The reactions were loaded corresponding to their decreasing enzyme concentration from left to right. The top insert band (lanes 4 and 5), indicated by an arrow, corresponds to the full length *Nco* I/*Pst* I NS3P insert.

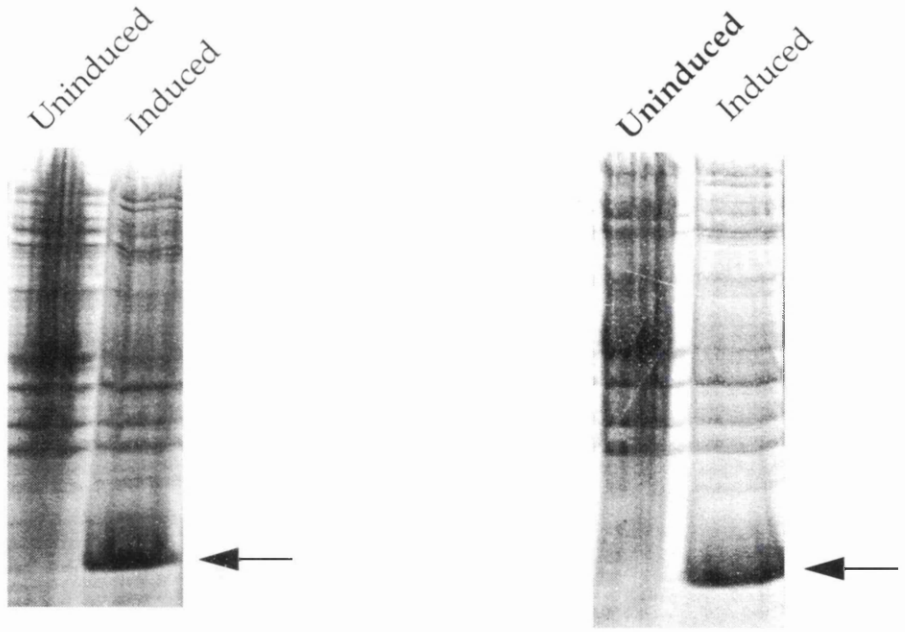


Fig. 3.6. Expression of the NS3 protease domain in two bacterial strains. After IPTG was added to a final concentration of 1-2mM the cells were incubated for 5 hours at 37°C. Samples of both uninduced and induced cell lysates were analysed through a 12% SDS polyacrylamide gel. The NS3P product was expressed both in SG13009 [pREP4]cells (left) and M15[pREP4] cells (right). The induced NS3P band is indicated by an arrow.

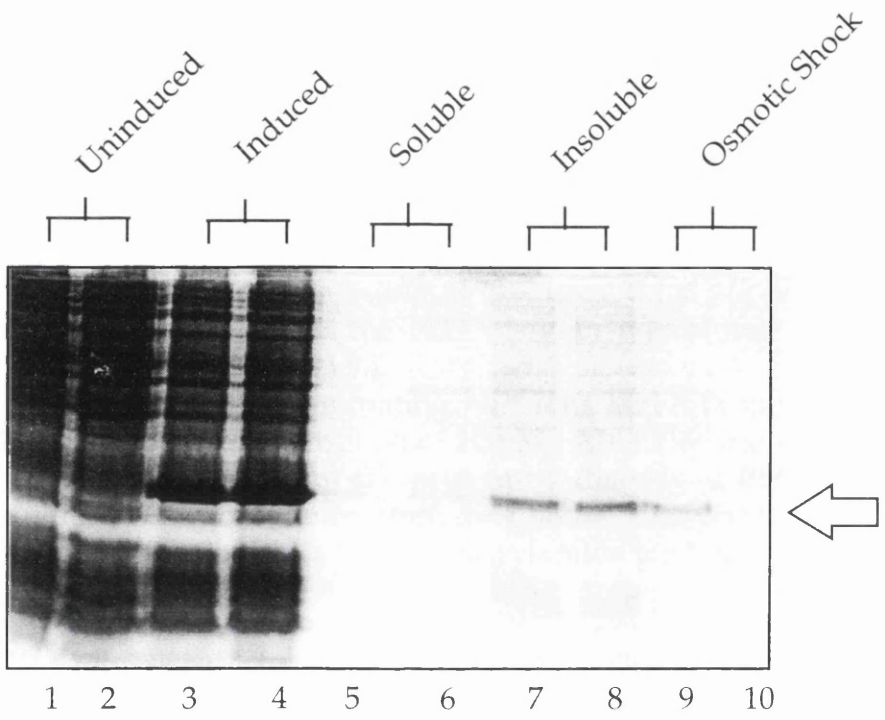


Fig. 3.7. Expression of the NS3 protease domain in *E. coli* strains SG13009[pREP4] and M15[pREP4]. For each pair of lanes the first lane corresponds to expression in the M15[pREP4] strain and the second lane corresponds to expression in the SG13009[pREP4] strain. The uninduced cells were loaded in lanes 1 and 2 and the induced band was seen for both cell lines in lanes 3 and 4. The nature of the expressed protein was investigated by determining whether the protein was found in the soluble fraction (lanes 5 and 6), the insoluble fraction (lanes 7 and 8) or associated with the periplasm (lanes 9 and 10). For both strains the NS3P protein product was found in the insoluble fraction. The faint band in the M15[pREP4] osmotic shock lane (lane 9) was probably spill over from the adjacent lane (insoluble fraction) .

After expression in both bacterial strains the NS3P protein was tested for binding to the Ni^{2+} resin (**2.3.1.**). The cells were pelleted and resuspened in Buffer A (6M GuHCl, 0.1M Na-phosphate, 0.01M Tris-HCl, pH 8.0) before mixing with a 50% slurry of Ni-NTA resin. After mixing for 30 minutes the sample was pelleted and the supernatant removed. The sample was washed with Buffer C (8M urea, 100mM NaH_2PO_4, 10mM Tris-HCl, pH adjusted to 6.0) before gentle mixing with BufferC/100mM EDTA. The EDTA chelates the Ni^{2+} ions from the NTA resin and elutes the protein. The samples obtained after elution of the protein from the pelleted Ni^{2+} resin were analysed by SDS-PAGE (Fig. 3.8). The NS3P protein expressed in bacteria was bound to the Ni-NTA resin and could be purified after expression.

A large scale (500ml) expression of NS3P was carried out. The cells were pelleted and resuspened in Buffer A (6M GuHCl, 0.1M Na-phosphate, 0.01M Tris-HCl, pH 8.0) before mixing with a 50% slurry of Ni-NTA resin. After mixing for 45 minutes the resin was loaded onto a column and washed with Buffers A-F. The protein was eluted from the column with buffer E (8M urea, 0.1M Na-phosphate, 0.01M Tris, pH4.5). Although there was a small amount of NS3P eluted with buffer D (8M urea, 0.1M Na-phosphate, 0.01M Tris, pH 5.9) the majority of the NS3P protein was eluted in the first two fractions with buffer E (Fig. 3.9.).

To remove any contaminating Ni^{2+} ions EDTA (a metal ion chelator) was added to a final concentration of 10mM. After chelation of the Ni^{2+} ions the EDTA was removed from the protein by dialysis in Buffer E (the buffer used to elute NS3P from the column) for 4 hours. The purified NS3P protein was analysed through a 12% SDS polyacrylamide gel (Fig. 3.10.).

3.4. HCV protease cofactor clone, NS4A.

The cofactor for NS3 serine protease activity was demonstrated to be the 54 amino acid NS4A protein (Failla *et al.*, 1994; Bartenschlager *et al.*, 1994). The NS4A gene (consisting of amino acids 1658 to 1711 of the HCV Gla polyprotein) from HCV Gla was amplified by PCR and cloned into pTM1 using the PCR primers shown in Fig. 3.11. An *Nco* I site was used in the 5′ primer and a *Pst* I site used in the 3′ primer. The cloned gene was sequenced completely and shown to be correct. The NS4A protein is predicted to have a molecular weight of 6 kDa (Tomei *et al.*, 1993) and hence after expression *in vitro* the protein was not observed by SDS-PAGE on either a 12% or a 15% SDS polyacrylamide gel.

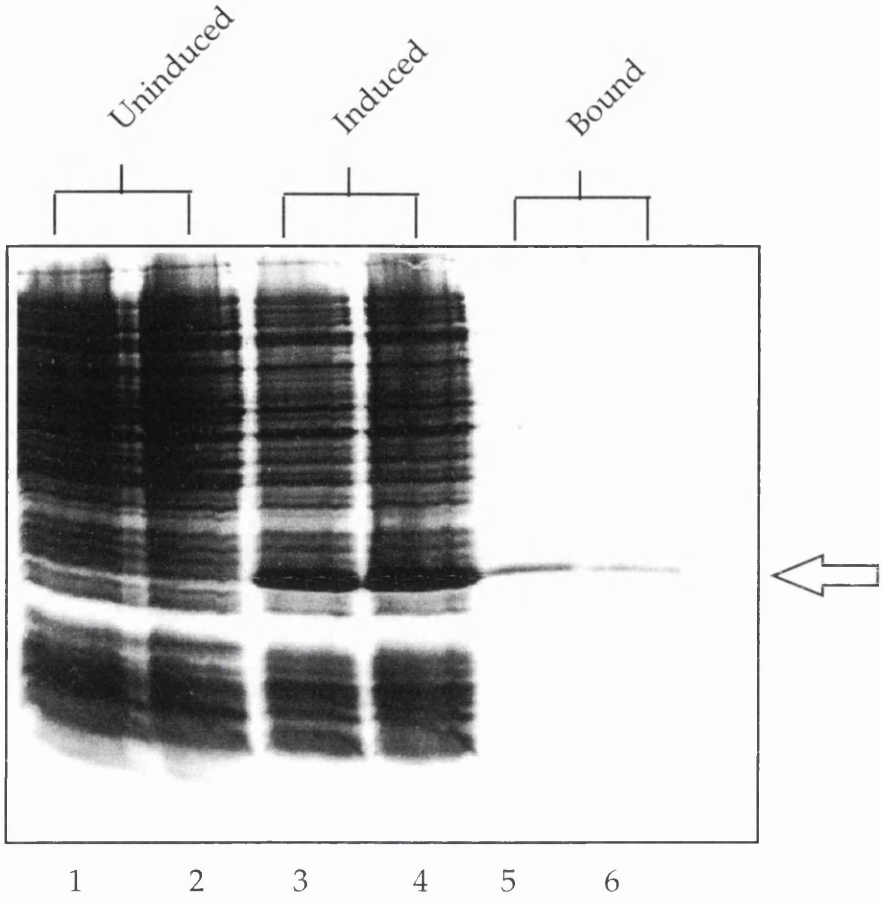


Fig. 3.8. Expression of the NS3 protease domain (NS3P) in bacteria and analysis of binding to Ni^{2+} resin. The first lane of each pair corresponds to expression in the M15[pREP4] strain and the second lane corresponds to expression in the SG13009[pREP4] strain. The induced NS3P band (lanes 3 and 4) was seen in addition to the background protein bands observed in the uninduced samples (lanes 1 and 2). The expressed NS3P is indicated by an arrow. When expressed in either cell line the NS3P protein was observed to bind to the nickel resin (lanes 5 and 6) after the cells had been lysed.

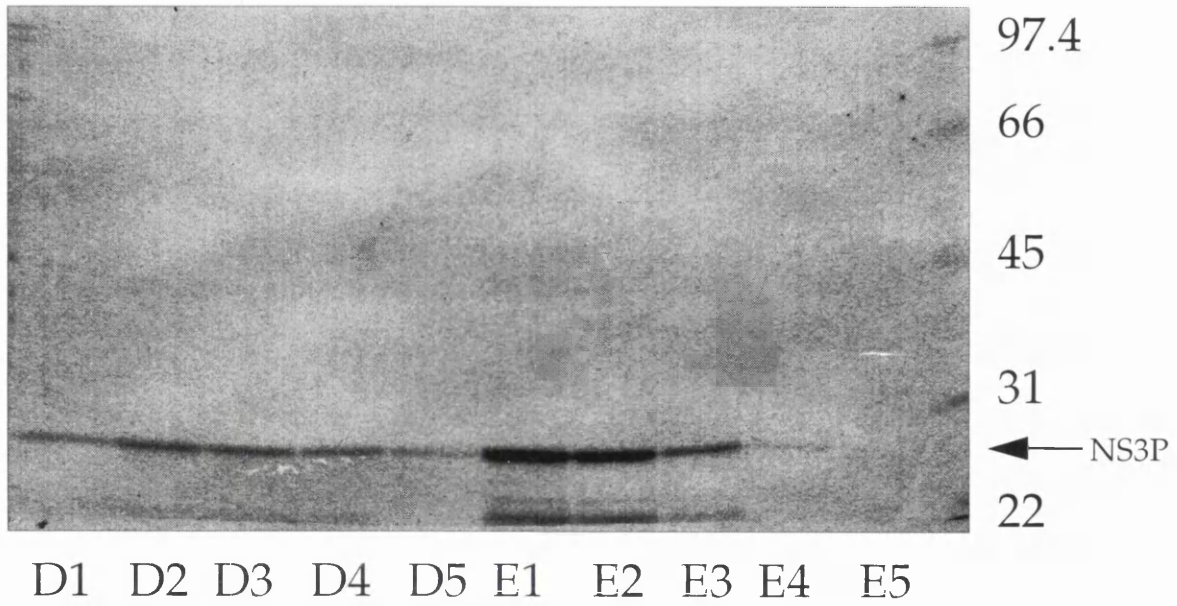


Fig. 3.9. Elution of bacterially expressed NS3P from the Nickel resin. After binding to Ni-NTA, bound protein was eluted with 5x3ml aliquots of buffer D (8M urea, 0.1M Na-phosphate, 0.01M Tris, pH 5.9) and 5x3ml aliquots of buffer E (8M urea, 0.1M Na-phosphate, 0.01M Tris, pH4.5). Forty microlitres of each eluate were mixed with 40µl Protein Dissociation Mix and half of this sample analysed by SDS-PAGE. The majority of the NS3P protein was eluted with the first two washes with buffer E. The NS3P band is identified with an arrow. Molecular weight markers are indicated on the right.

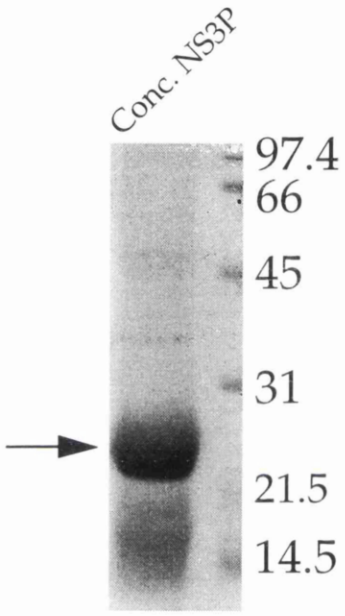


Fig. 3.10. Bacterially expressed NS3P after purification and concentration. The concentrated NS3P was analysed by SDS-PAGE through a 12% gel. The NS3P product is a distinct band of approximately 25kDa and is identified by an arrow. Molecular weight markers are located at the right of the gel.

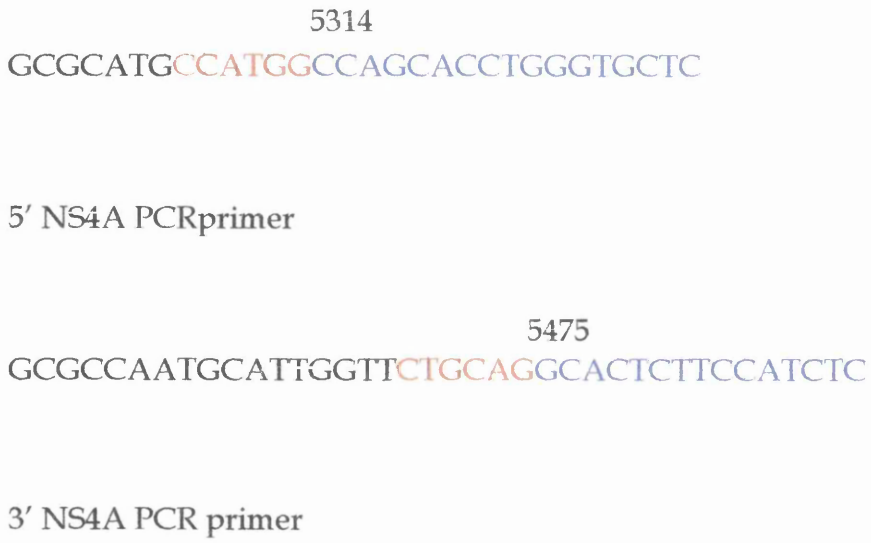


Fig. 3.11. PCR primers for the amplification of the HCV Gla NS4A gene. Restriction enzyme sites are illustrated in red and HCV NS4A sequence illustrated in blue. The sequence illustrated in black was included to ensure efficient restriction enzyme digestion. The numbers shown above indicate the nucleotide position within the HCV Gla sequence of the start and finish of the NS4A region.

3.5. Summary of HCV expression clones.

Two NS5 clones (full length NS5 and the truncated tNS5T) were constructed to provide a substrate in an HCV serine protease assay. Three protease constructs (pTF7.5 NS34A, pTF7.5 NS3, pTM1 NS3P) were produced for expression in a T7-RNA polymerase driven system, as well as a construct expressing the cofactor, pTM1 NS4A. All the HCV constructs (pTF7.5 NS34A, pTF7.5 NS3, pTM1 NS3P, pTM1 tNS5T, pTZ18 tNS5T and pTM1 NS5) were tested for expression by *in vitro* transcription/translation and analysed by SDS-PAGE through a 12% polyacrylamide gel (Fig. 3.12.).

All the HCV constructs expressed well *in vitro*. In this system the NS34A construct gave two separate bands, a 75 kDa band corresponding to uncleaved NS34A and a smaller 70 kDa product corresponding to NS3. The NS3P product appeared to be approximately 19kDa. The full length NS5 substrate band was estimated to be approximately 120 kDa in weight. The truncated, epitope tagged NS5 substrate (tNS5T) was seen as a 75-80 kDa band. There are a number of additional bands present that are either non-specific by-products of the TNT system or breakdown products.

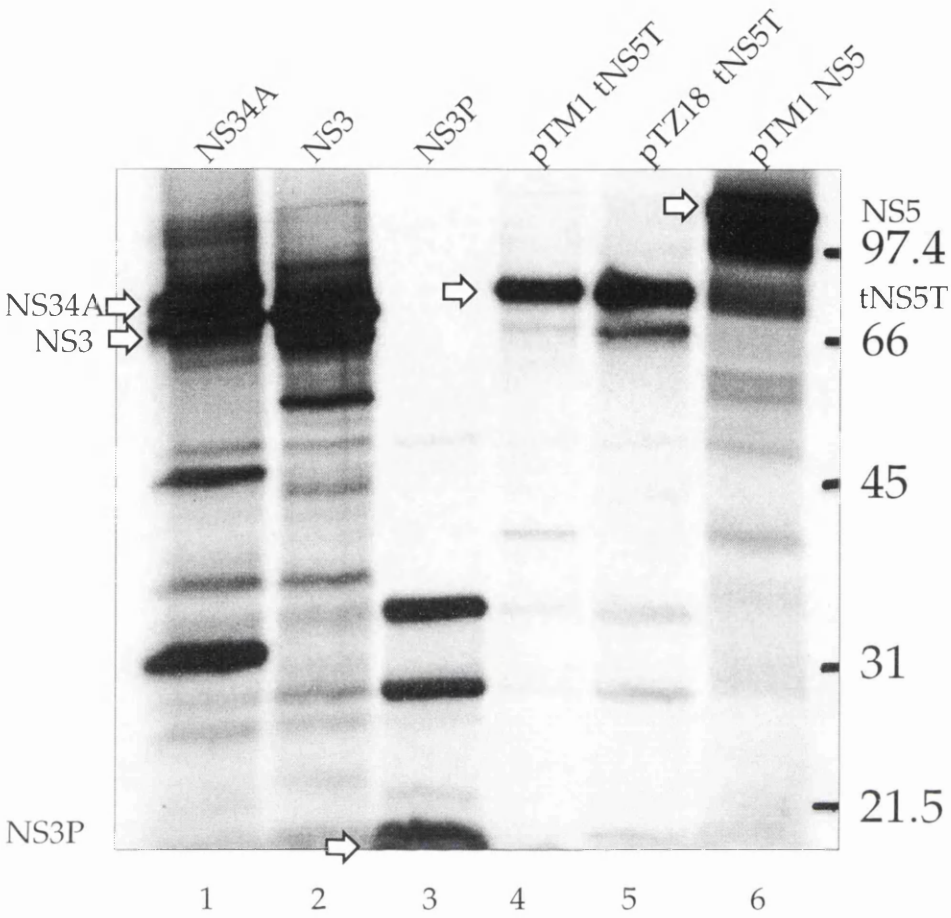


Fig. 3.12. Expression of HCV proteins by *in vitro* transcription/translation. The HCV protein cDNAs were cloned within a variety of plasmids, all of which gave a high level of protein expression *in vitro*. Three substrate reactions were loaded at the right of the gel (lanes 4, 5, 6). The protease reactions were loaded at the left of the gel (lanes 1, 2, 3). The NS3P band was observed to have a slightly smaller molecular weight than predicted. There were a number of bands in addition to the HCV protein bands (the HCV bands are indicated by an arrow). These additional protein bands were either produced non-specifically during the transcription/translation reaction or were breakdown products. The molecular weight markers are located at the right of the gel.

Chapter 4 - Development of an assay for protease activity

4.1. Proteolytic activity of NS3 serine protease using *in vitro* transcription/translation systems.

The use of the coupled *in vitro* transcription/translation system (TNT) from Promega was found to be efficient for the expression of both substrate and protease constructs separately (Fig. 3.12.). It had not been determined whether, having expressed a suitable substrate *in vitro*, a separately produced NS3 protein could process *in trans*. The ability of the NS3 serine protease to process *in trans* could be investigated either (i) by combining lysates after separate transcription/translation or (ii) by programming the TNT reactions with both plasmids.

Separate 25µl *in vitro* transcription/translation reactions (2.6.2) were carried out for the NS34A construct and both the full length and truncated substrate constructs. After the reactions had been incubated at 30°C for 60 minutes, 10µl of the substrate reaction were mixed with 10µl of the protease reaction and the combined lysates incubated at 37°C for 2 hours. Then 70 µl of protein dissociation mix were added and 40µl of the sample analysed by SDS-PAGE (12% gel). The results of this experiment are shown in Fig. 4.1. No cleavage of either substrate, as evidenced by the appearance of smaller cleavage products, was observed. The length of the incubation and the incubation temperature were varied but no cleavage products were observed.

One microgram of both substrate and protease (NS34A) constructs was introduced into the same TNT reaction and incubated at 30°C for 180 minutes. The reaction was terminated by addition of 70µl protein dissociation mix and half the sample analysed by SDS- PAGE through a 12% SDS polyacrylamide gel (lanes 6 and 7, Fig. 4.1.). As for the combination of substrate and protease TNT lysates no evidence of cleavage was observed.

As coupled transcription/translation *trans*-cleavage reactions had not been successful it was investigated whether separate transcription and translation reactions would enable processing *in trans*. mRNAs were transcribed from linearised substrate and protease vectors and purified before separate translation reactions were carried out in rabbit reticulocyte lysates. Ten microlitres of the substrate reaction were mixed with 10µl of the protease reaction and the sample incubated at 37°C for 2 hours. Seventy microlitres of protein dissociation mix were added to the reaction and 40µl of the sample analysed by SDS-PAGE through a 12% SDS polyacrylamide

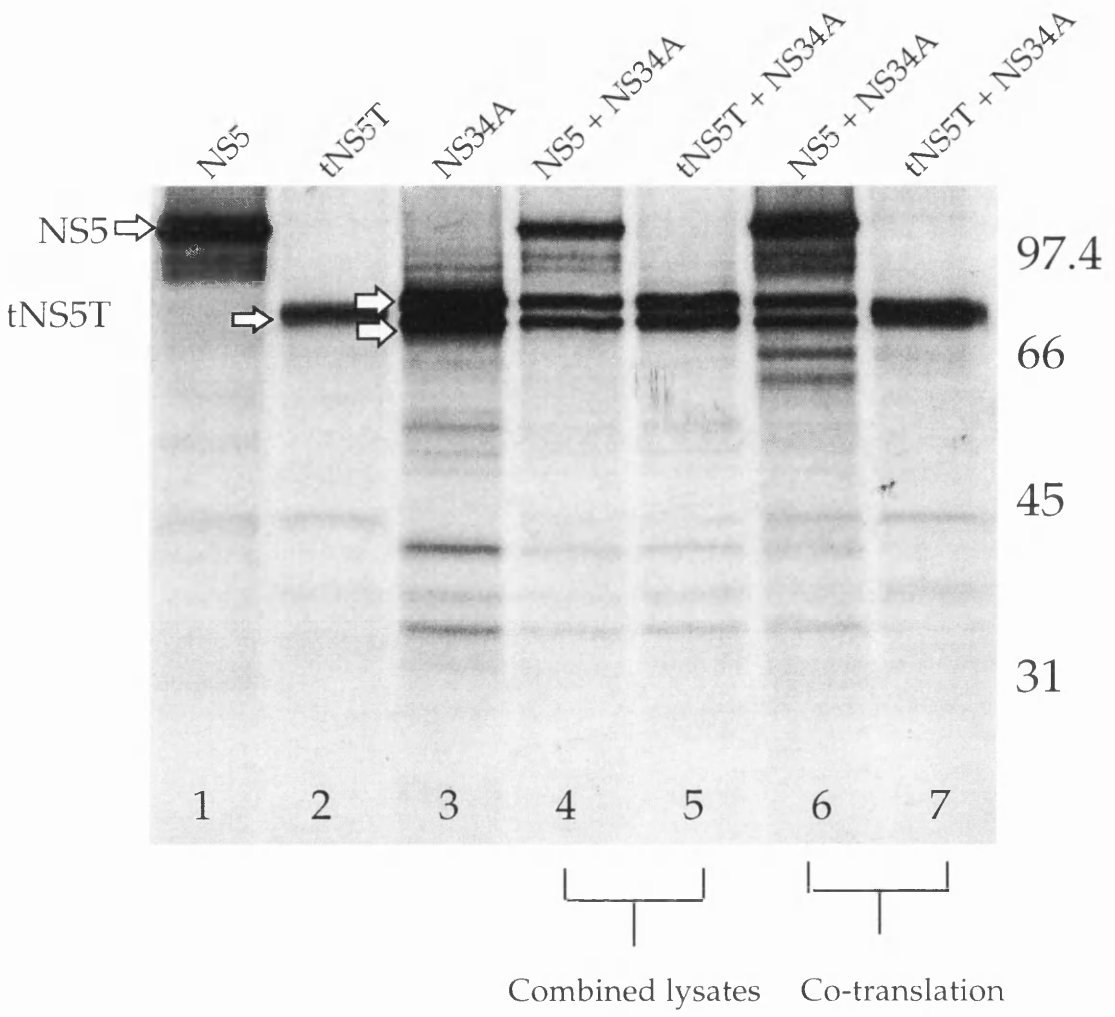


Fig. 4.1. *In vitro trans* -cleavage assay using coupled transcription/translation system (TNT). The HCV protease and substrate proteins were expressed *in vitro* by coupled transcription/translation. Both the full length NS5 (lane 1) and tNS5T (lane 2) were expressed individually and are indicated by a black arrow. The NS34A construct produced a full length NS34A and a smaller, cleaved NS3 product (lane 3). The protease bands are indicated by white arrows. When protease and substrate were produced by separate TNT reactions and the lysates combined no cleavage products were observed (lanes 4 and 5). When both protease and substrate were produced in the same TNT reaction no cleavage products were observed (lanes 6 and 7). Molecular weight markers are illustrated at the right of the gel.

gel. The quality of the proteins produced by this method was lower than was observed for coupled transcription/translation. The level of expression of the HCV proteins was lower than the coupled transcription/translation reactions and there was also a significantly higher background. Processing of the substrate was not visible using separate transcription and translation reactions (lanes 4 and 5, Fig. 4.2.).

An alternative approach was to use the bacterially expressed NS3 protease domain, which had been purified and concentrated (3.3.), to cleave substrates produced by *in vitro* transcription/translation. The catalytic activity of this concentrated protease fragment was not known. Introduction of this protein to a lysate containing *in vitro* expressed substrates could result in *trans*-cleavage. The purified, bacterially expressed NS3P was added in increasing amounts to TNT reactions of both the full length and the truncated NS5 substrates. After the addition of the purified protease the samples were incubated at 37°C for 2 hours before the addition of 70 µl protein dissociation mix and analysis by SDS-PAGE. The results (Fig. 4.3.) show that neither substrate was processed when the purified protease was added to substrate expressed *in vitro*.

Trans-cleavage of the HCV NS5A/NS5B substrates by the HCV NS3 serine protease was not observed using any *in vitro* protein expression system. Co-expression of substrate and protease proteins in a TNT reaction, combination of protease and substrate-containing lysates produced by TNT or RRL reactions or the addition of bacterially expressed NS3P to *in vitro* expressed substrates all failed to produce *trans*-cleavage. As no assay involving *in vitro* expression of HCV proteins had been successful an alternative method was investigated.

4.2. Proteolytic activity of the NS3 serine protease in mammalian cells.

vTF7.3 is a recombinant vaccinia virus encoding T7 RNA polymerase. After infection of mammalian cells vTF7.3 expresses a large amount of T7 RNA polymerase in the cytoplasm of the infected cell. It has been demonstrated that introduction of plasmids containing genes under control of the T7 promoter can result in a high level of expression of the encoded proteins (Fuerst *et al.*, 1986). Mammalian cells (CV-1 or BHK) were infected with vTF7.3 and subsequently transfected with plasmids encoding the substrate and protease cDNAs under control of the T7 promoter. Successful introduction of plasmid DNA by this method resulted in a high level of expression of the HCV proteins (Fig. 4.4.). When the truncated NS5 substrate was cotransfected with the NS34A serine protease two additional protein bands were produced of the approximate size expected of tNS5T cleavage products. As these protein bands were only produced after cotransfection of the substrate with NS34A they were identified as being tNS5AT and tNS5BT. The 549 amino acid tNS5T substrate was processed into a 331 amino acid NS5A product of approximately 50 kDa (tNS5AT) and a smaller, 218 amino

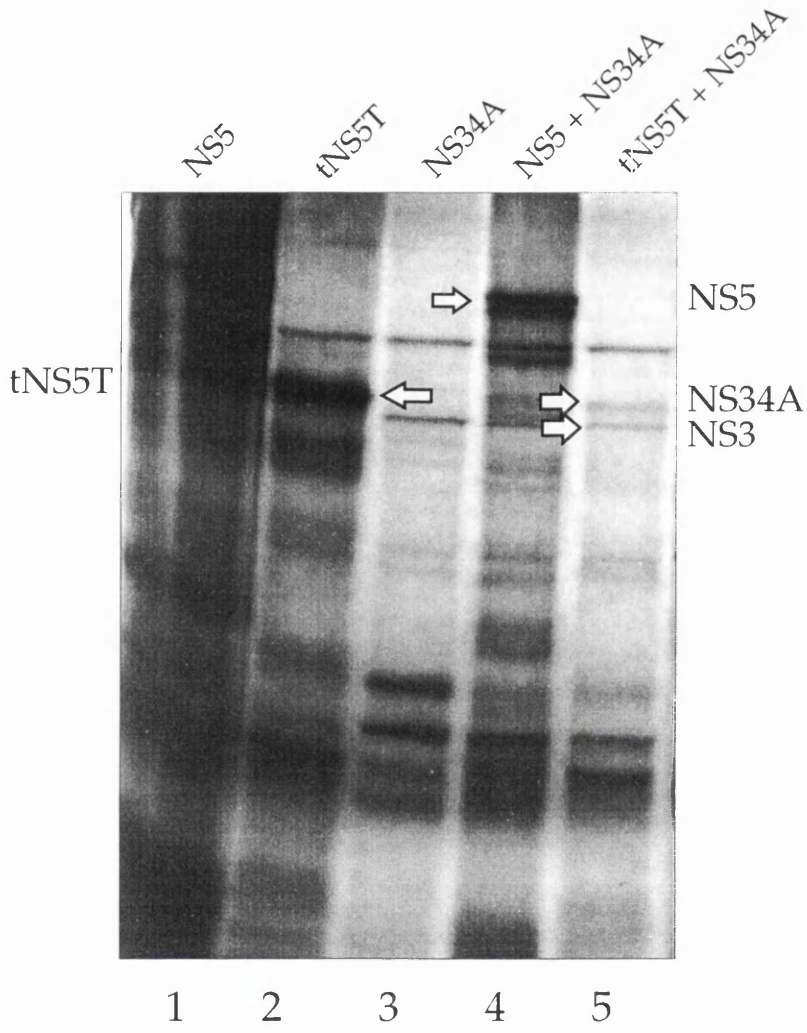


Fig. 4.2. *In vitro trans*-cleavage assay using separate transcription and translation reactions to produce the HCV proteins. The full length NS5 (lane 1), tNS5T (lane 2) and NS34A complex (lane 3) were expressed poorly after separate transcription and translation from cDNAs. When substrate and protease lysates were combined (lanes 4 and 5) no cleavage products were observed. The quality of protein production was significantly lower than for coupled transcription/translation reactions.

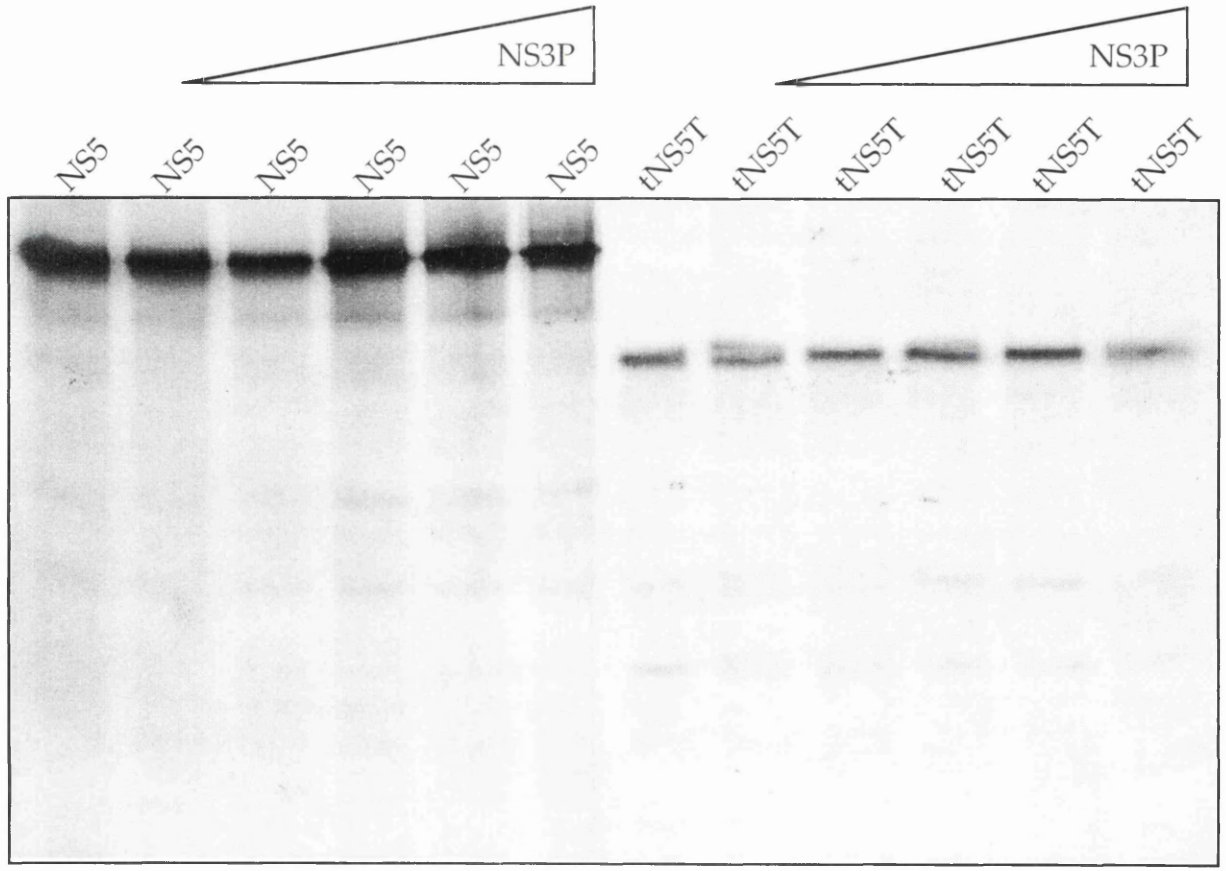


Fig. 4.3. *In vitro trans*-cleavage assay using bacterially-expressed, purified NS3P. Both substrates were expressed *in vitro* by coupled transcription/translation. Increasing amounts (0.5, 1, 2, 5 and 10 μg) of bacterially expressed, purified NS3P were added to both substrates and after incubation at 37°C the lysates were analysed by SDS-PAGE. No cleavage products were observed

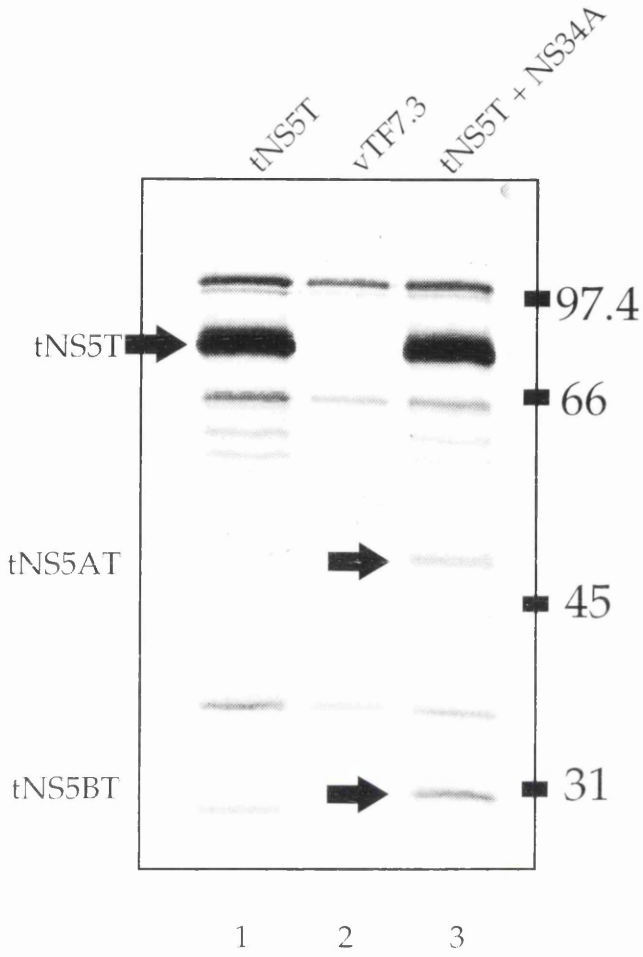


Fig. 4.4. *Trans*-cleavage of the truncated NS5 substrate (tNS5T) by coexpressed NS34A. Plasmid DNAs were transfected into CV-1 cells infected with vTF7.3. At 11 hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody. This protocol was carried out in all subsequent *trans*-cleavage experiments. When expressed alone in mammalian cells the tNS5T substrate was not processed (lane 1). When tNS5T was cotransfected with NS34A two cleavage products were produced (the 50kDa tNS5AT and the 30kDa tNS5BT). Despite the evidence of *trans*-cleavage a significant amount of unprocessed substrate remained(lane 3). In lane 2, unstransfected cells were immuneprecipitated to determine the background protein bands observed for this procesdure. The positions of molecular weight markers are located at the right of the gel.

acid, 28 kDa NS5B protein. In this initial *trans*-cleavage experiment 5μg of substrate (tNS5T) DNA was cotransfected with 5μg of the NS34A construct.

Although the NS34A protease complex processed the tNS5T substrate there was still a significant quantity of unprocessed substrate remaining. To attempt to optimise the level of processing mediated by the protease construct the protease construct was titrated against a constant amount of substrate. 5μg of substrate (tNS5T) was cotransfected with increasing amounts of the protease construct (Fig. 4.5.). As the amount of NS34A DNA increased the level of processing increased until reaching an optimal level of processing when 5μg of NS34A plasmid DNA was transfected. Any further increase in the quantity of protease DNA transfected had little effect on the amount of uncleaved substrate in comparison to the cotransfection with 5μg protease DNA. The efficiencies of processing observed with different quantities of NS34A are tabulated below.

Cotransfection	Relative % Cleavage efficiency
tNS5T + 2μg NS34A	6
tNS5T + 5μg NS34A	17
tNS5T + 10μg NS34A	20
tNS5T + 15μg NS34A	24

Table 4.1. Titration of NS34A when cotransfected with 5μg tNS5T cDNA. Increasing quantities of pTF7.5 NS34A were cotransfected with pTM1 tNS5T and after protein coexpression the relative % of substrate cleaved determined. As the amount of NS34A coexpressed increased the level of processing increased however there was very little increase in processing efficiency when more than 5μg NS34A was cotransfected.

Quantitation of the level of processing was calculated by dividing the total of the cleavage product intensities by the total of cleaved and uncleaved band intensities. The intensity of each band was determined by the number of (radioactively labelled) methionine residues within the protein. The sum of the intensities of the two cleavage products would equal the intensity of the single substrate band (as the uncleaved substrate contains the same number of methionine residues as the two cleavage products). This method of quantitation is "quasi-quantitative" as the calculation uses steady state levels of substrate and products. The level of processing efficiency was estimated as a percentage (cleaved substrate as a percentage of total substrate). The calculations for the values obtained in Table 4.1. are illustrated in Table 4.2.

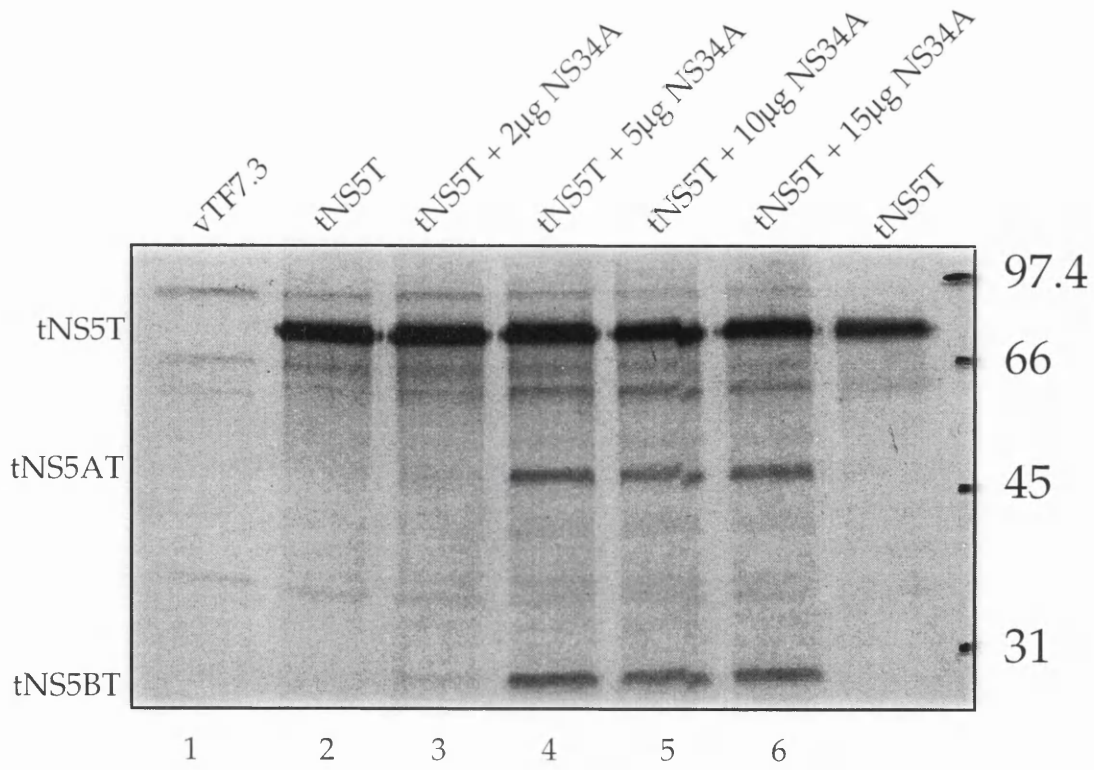


Fig. 4.5. *Trans*-cleavage analysis using different amounts of protease DNA. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with 5µg tNS5T (lanes 2 and 7), 5µg tNS5T+ 2µg NS34A (lane3), 5µg tNS5T+ 5µg NS34A (lane 4), 5µg tNS5T +10µg NS34A (lane 5), 5µg tNS5T+15µg NS34A (lane 6). Eleven hours post-transfection the cells were labelled with ^{35}S methionine for 2 hours, lysed and then immunepreciptated with the SV5-P-k antibody before analysis by SDS-PAGE. As the quantity of protease DNA introduced was increased, the efficiency of processing increased. Little increase in processing efficiency was observed when cotransfection with 5µg NS34A (lane 4) was increased to 10 or 15µg of NS34A (lanes 5 and 6). Molecular weight markers are indicated on the right.

Cotransfection	+2μg NS34A	+5μg NS34A	+10μg NS34A	+15μg NS34A
tNS5T	78408	76354	51917	36397
tNS5AT	3146	8821	6845	6213
tNS5BT	1588	6484	6346	5278
Total Substrate	83142	91659	65108	47888
Relative % Cleavage Efficiency	6	17	20	24

Table 4.2. Calculation of relative % cleavage efficiencies. The intensities of the uncleaved substrate and cleavage products were determined using ImageQuant. The relative % cleavage efficiency was calculated by dividing the sum of the cleavage product intensities by the total substrate intensity and the multiplying this value by 100.

It was considered that the transfection protocol might be inefficient at introducing two plasmids into the same cell when large quantities of DNA were used. Alternatively, the level of processing observed could be the most efficient possible and any further increase in protease would have no effect on cleavage efficiency.

The amount of protease DNA that gave the highest level of processing (5μg) was maintained as a constant and the substrate DNA titrated. Cotransfections were carried out using different quantities of substrate DNA from 1-5μg (Fig. 4.6.). As the quantity of substrate cDNA transfected decreased the amount of substrate protein expressed decreased. The amount of NS34A remained constant, but in comparison with the substrate the ratio of protease:substrate increased. The lower the quantity of substrate DNA transfected the more cleavage products were produced relative to unprocessed substrate. Thus there was an increase in the efficiency of processing when the quantity of the substrate present fell (when the quantity of protease was unchanged). The relative % cleavage for these assays is illustrated below.

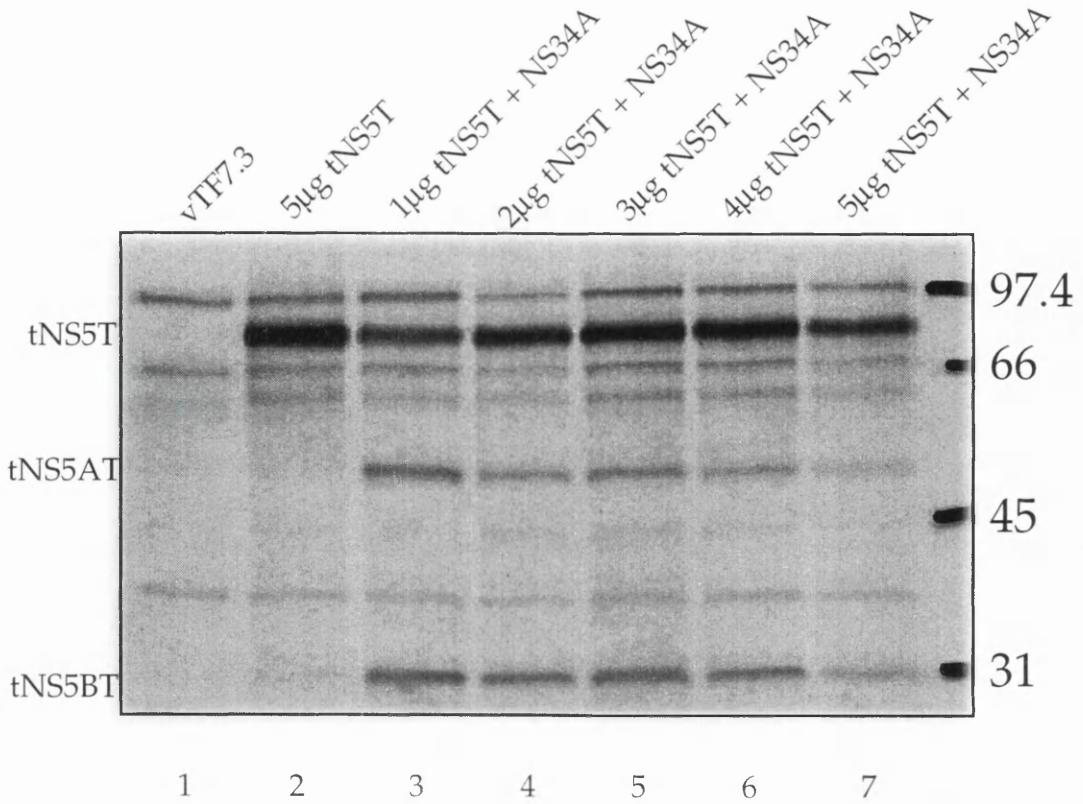


Fig. 4.6. Cotransfection of different amounts of substrate with protease. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with 5µg tNS5T alone (lane 2), 1µg tNS5T + NS34A (lane 3), 2µg tNS5T + NS34A (lane 4), 3µg tNS5T + NS34A (lane 5), 4µg tNS5T + NS34A (lane 6), 5µg tNS5T + NS34A (lane 7). Five micrograms of NS34A were used in each cotransfection. Eleven hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immunoprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. As the amount of substrate DNA was increased the efficiency of processing decreased, with an increasing ratio of unprocessed substrate relative to processed. The relative % cleavage efficiency was highest when 1µg of substrate was cotransfected with 5µg NS34A (lane 3). Molecular weight markers are indicated on the right

Cotransfection	Relative % Cleavage Efficiency
1μg tNS5T + NS34A	46
2μg tNS5T + NS34A	17
3μg tNS5T + NS34A	22
4μg tNS5T + NS34A	17
5μg tNS5T + NS34A	14

Table 4.3. Titration of tNS5T when cotransfected with 5μg NS34A cDNA. Increasing quantities of substrate cDNA were cotransfected with a fixed amount of pTF7.5 NS34A. The less substrate cotransfected, the less substrate expressed and the higher the ratio of protease:substrate. As the ratio of protease:substrate increased a higher relative % of cleavage efficiency was obtained.

Further titration of the assay was carried out using 1μg of substrate DNA and varying the amount of protease DNA (Fig. 4.7.). When 5μg of protease DNA was cotransfected the substrate was processed with the highest level of efficiency. Any further decrease in the amount of protease DNA cotransfected resulted in a fall in the efficiency of processing. The processing efficiencies are tabulated below.

Cotransfection	Relative % Cleavage Efficiency
tNS5T + 1μg NS34A	12
tNS5T + 2μg NS34A	32
tNS5T + 3μg NS34A	29
tNS5T + 4μg NS34A	32
tNS5T + 5μg NS34A	42
5μg tNS5T + 5μg NS34A	21

Table 4.4. Titration of NS34A when cotransfected with 1μg tNS5T. Increasing quantities of pTF7.5 NS34A were cotransfected with an unchanging amount of pTM1 tNS5T. As the ratio of protease:substrate increased the relative % cleavage efficiency increased.

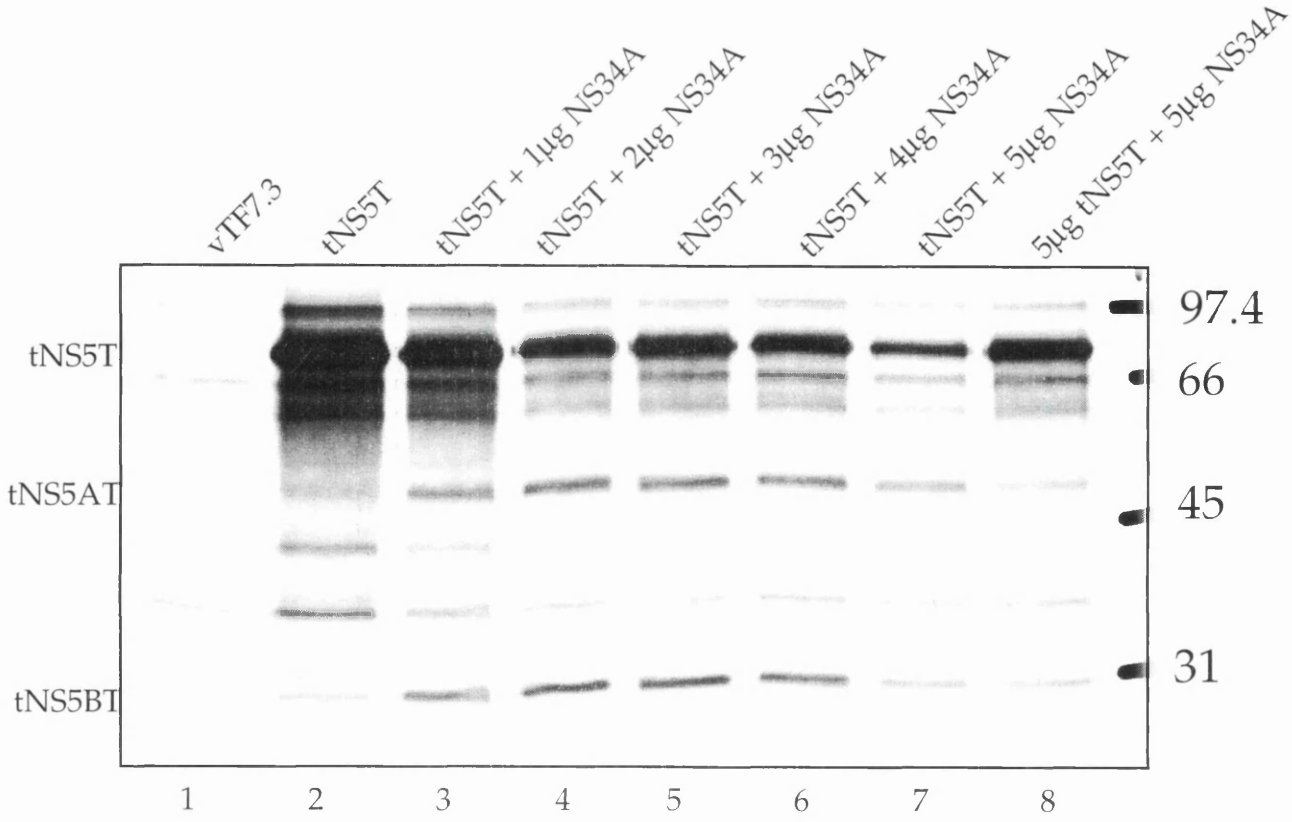


Fig. 4.7. Cotransfection of substrate with increasing amounts of protease DNA. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with 1μg tNS5T (lane 2), 1μg tNS5T + 1μg NS34A (lane 3), 1μg tNS5T + 2μg NS34A (lane 4), 1μg tNS5T + 3μg NS34A (lane 5), 1μg tNS5T + 4μg NS34A (lane 6), 1μg tNS5T + 5μg NS34A (lane 7) or 5μg tNS5T + 5μg NS34A (lane 8) . Eleven hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immunoprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. Cotransfection of tNS5T with 5μg NS34A (lane 7) gave the highest efficiency of processing as there was more processed substrate relative to unprocessed. Molecular weight markers are indicated on the right.

Standard parameters for the cotransfection protocol were established with 1μg of substrate DNA cotransfected with 5μg of protease DNA. This protocol gave an optimal level of substrate processing with approximately 40-45% of the substrate processed in different experiments.

To determine if the efficiency of processing by the protease could be improved the duration of the incubation step of the transfection protocol (after addition of the plasmid DNA to the cells) was extended from 11 to 15 hours after transfection (Fig. 4.8.).

When the length of the incubation step after transfection was increased, the efficiency of processing increased by a significant amount. The increase in the efficiency of processing of the tNS5T substrate increased from approximately 45% to 75%, resulting in significantly less substrate remaining uncleaved.

To confirm that processing of the substrate was mediated by the NS3 protease, an inactive NS3 serine protease construct, ΔLL NS3-$5_{S\text{-}A}$ was used as a negative control. This construct was supplied by Dr B. E. Clarke (GlaxoWellcome) and encodes the NS3-NS5 region of the type 1b HCV-A genome with a serine residue in the catalytic triad of the NS3 serine protease changed to an alanine residue, that abolishes protease activity (D'Souza *et al.*, 1994, 1995). When the full length or the truncated NS5 substrate was cotransfected with the inactive serine protease no cleavage products were observed (Fig. 4.9.). All the previous experiments with respect to the characterisation of the *trans*-cleavage assay were carried out using the tNS5T substrate only as this was the only substrate that could be detected by immunoprecipitation. The NS5 antibody required for the identification of the full length NS5 and NS5A was unavailable until this time. The previous characterisation of the *trans*-cleavage assay had only been possible with the tNS5T substrate.

A comparison of different transfection protocols was carried out to see whether any further optimisation of the level of processing could be achieved. The standard transfection protocol used cationic liposomes whereas it was reported that the positively charged polyethylene imine (PEI) was a highly efficient vector for the transfection of plasmid DNA into cells (Boussif *et al.*, 1995). Transfections were carried out with 1μg of tNS5T + 5μg NS34A cotransfected into vTF7.3 infected CV-1 cells. For one reaction the DNA was transfected with cationic liposomes and for the other PEI was used. The experiments were carried out in tandem, the lysates immunoprecipitated and then analysed through a 12% SDS polyacrylamide gel (Fig. 4.10.). No significant difference in processing efficiency was observed when either cationic liposomes or PEI were used to introduce the plasmid DNA into the mammalian cells.

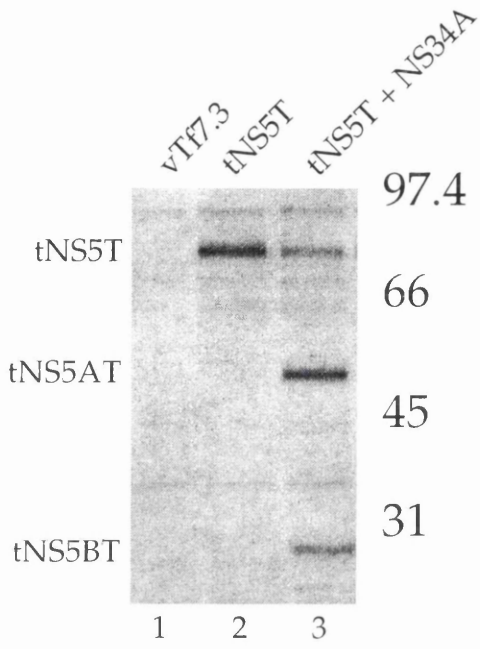


Fig. 4.8. Optimisation of processing. vTF7.3 infected CV-1 cells were left untransfected with DNA (lane 1) or transfected with tNS5T (lane 2) or tNS5T + NS34A (lane 3). For the original protocol the cells were labelled with ^{35}S-methionine 11 hours post-transfection , lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The incubation period of the protocol (after introduction of the plasmid DNA) was increased from 11 to 15 hours and this resulted in a significantly lower amount of unprocessed substrate. The efficiency of processing was increased from 45% to 75%. Molecular weight markers are indicated on the right.

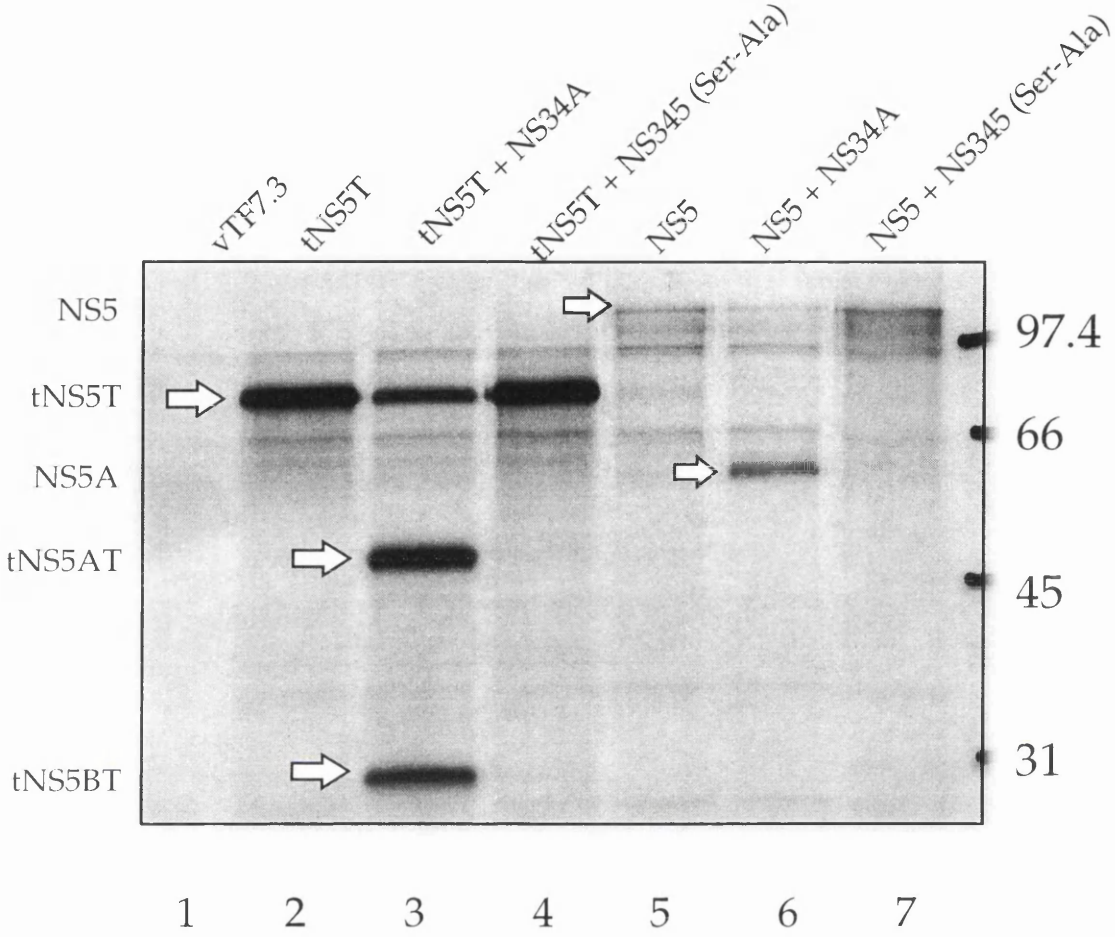


Fig. 4.9. Comparison of processing between an active and an inactive NS3 protease. CV-1 cells infected with vTF7.3 were transfected with various combinations of HCV plasmid DNAs. The tNS5T substrate was transfected alone (lane 2), with 5µg NS34A (lane 3) or with 5µg NS234 [Ser-Ala] (lane 4). The NS5 substrate was transfected alone (lane 5), with 5µg NS34A (lane 6) or with 5µg NS34A [Ser-Ala] (lane 7). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours , lysed and then immuneprecipitated with the relative antibodies. Both substrates were processed when they were coexpressed with NS34A (lanes 3 and 6). When the substrates were coexpressed with the inactive protease NS234[Ser-Ala] no cleavage products were observed (lanes 4 and 7). The tNS5T protein bands are indicated by white arrows and the NS5 protein bands indicated by black arrows. Molecular weight markers are illustrated on the right.

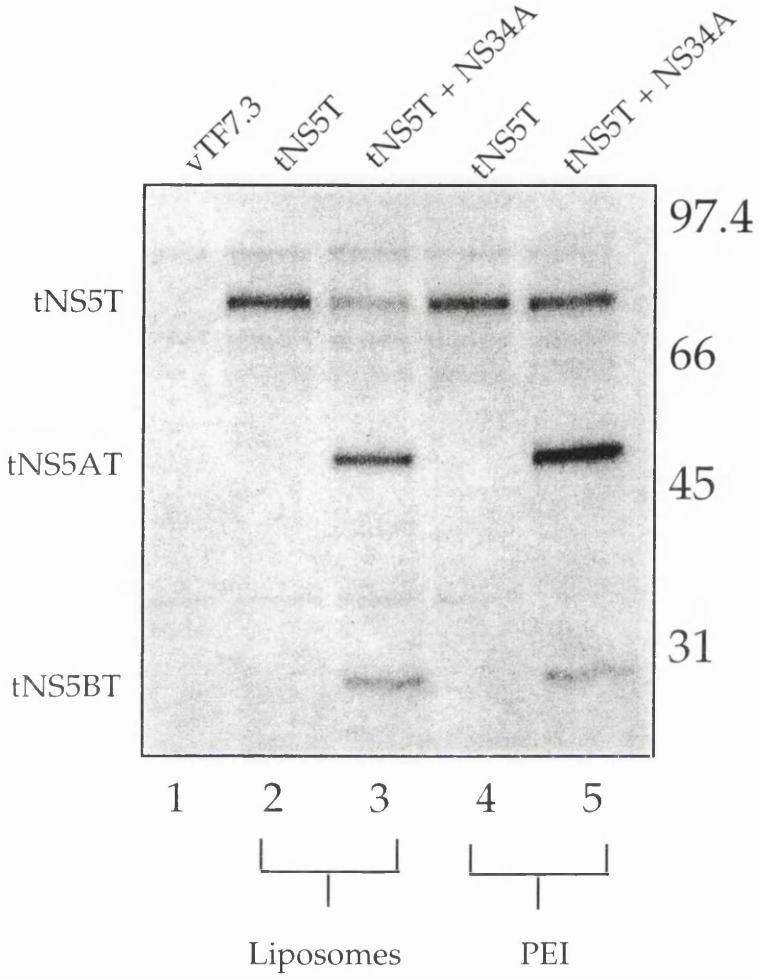


Fig. 4.10. Comparison of different transfection methods. The HCV plasmids were introduced into cells using either cationic liposomes or polyethylene imine (PEI). vTF7.3 infected CV-1 cells were transfected with plasmids and 15 hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours. The cell lysates were immunoprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The immunoprecipitated control (vTF7.3 infected cell lysate) was loaded in lane 1. The tNS5T substrate was transfected alone (lane 2) or with NS34A (lane 3) using cationic liposomes. tNS5T was also transfected either alone (lane 4) or with NS34A (lane 5) using PEI. Using either transfection medium the same level of processing was observed (lanes 3 and 5). Molecular weight markers are located at the right of the gel.

4.3. Summary.

An assay was developed for the analysis of cleavage mediated *in trans* by the NS3 serine protease. One microgram of the relevant HCV substrate cDNA was cotransfected with 5μg of NS34A cDNA into mammalian cells. The method of DNA introduction (Liposomes or PEI) did not appear to affect the efficiency of the cleavage reaction. After transfection of the plasmid DNAs the time of incubation before labelling was optimised resulting in a significantly increased quantity of cleavage products produced relative to the unprocessed substrate. The cleavage products obtained were confirmed as being produced by NS3 serine protease-mediated cleavage.

Chapter 5 - Mutagenesis of the NS5A/NS5B cleavage site

5.1. Development of a cassette for mutagenesis within the tNS5T substrate.

Mutagenesis of the NS5A/NS5B cleavage site required the development of an efficient system for the introduction of specific mutations. The alteration of amino acids within the cleavage site would provide an insight into the substrate requirements for processing by the NS3 serine protease. To study the importance of amino acids close to the NS5A/NS5B cleavage site a small cassette encompassing the cleavage site was introduced. This cassette, encoded in the tNS5T cDNA, would allow the production of tNS5T substrates containing mutated NS5A/NS5B cleavage sites, altered at the nucleic acid level to produce specific amino acid mutations, in an easy and straight-forward manner.

Two naturally occurring restriction enzyme sites, one either side of the cleavage site, were present within the tNS5T cDNA. A *Sal* I site situated 45 nucleotides upstream of the cleavage site and a *Sac* II site situated 47 nucleotides downstream of the cleavage site were contained within the nucleotide sequence of the tNS5T cDNA. When the tNS5T cDNA was digested with both these enzymes a double-stranded 86bp insert was removed. Construction of a corresponding mutant cleavage site cassette (for introduction into the digested vector) required the ligation of two synthetic oligonucleotides of considerable size. The quality of synthesis of oligonucleotides decreases as the size of the oligonucleotide increases. The deprotection of the bases used becomes inefficient and a mixed population of oligonucleotides would be created as opposed to the desired oligonucleotide. This fact, coupled with the expense, would make the construction of a cassette of this size infeasible.

By introducing silent mutations in the nucleotide sequence (that would not alter the amino acid sequence of the encoded polyprotein) two restriction enzyme sites closer to the NS5A/NS5B cleavage site could be created. A *Kpn* I site could be introduced 19 nucleotides upstream of the cleavage site and a *Bst* EII site could be introduced 30 nucleotides downstream. These two restriction enzyme sites, when digested, would produce a double-stranded 49bp insert with sticky ends.

To introduce the silent mutations by PCR two primers were designed. The 5' primer contained the *Sal* I site in the tNS5T cDNA and the new *Kpn* I site. The 3' primer contained the *Sac* II site in the tNS5T cDNA and the new *Bst* EII site (Fig. 5.1.). After PCR amplification the product was digested with *Sal* I and *Sac* II. The digested product was ligated into *Sal* I/*Sac* II digested pTM1 tNS5T (Fig. 5.2.). The PCR product was sequenced to

Fig. 5.1. PCR primers used to introduce silent mutations into the HCV NS5 sequence. For the nucleotide sequences illustrated the sequence in green is a non-specific sequence to protect the primer from degradation during PCR, the sequence in black corresponds to HCV sequence, the sequence in blue represents the restriction enzyme sites present within the HCV sequence and the sequence in red represents the introduced restriction enzyme sites. The conserved amino acid sequence is illustrated above the nucleotide sequences. The nucleotide number is illustrated below the 5′ terminal nucleotide of each oligonucleotide.

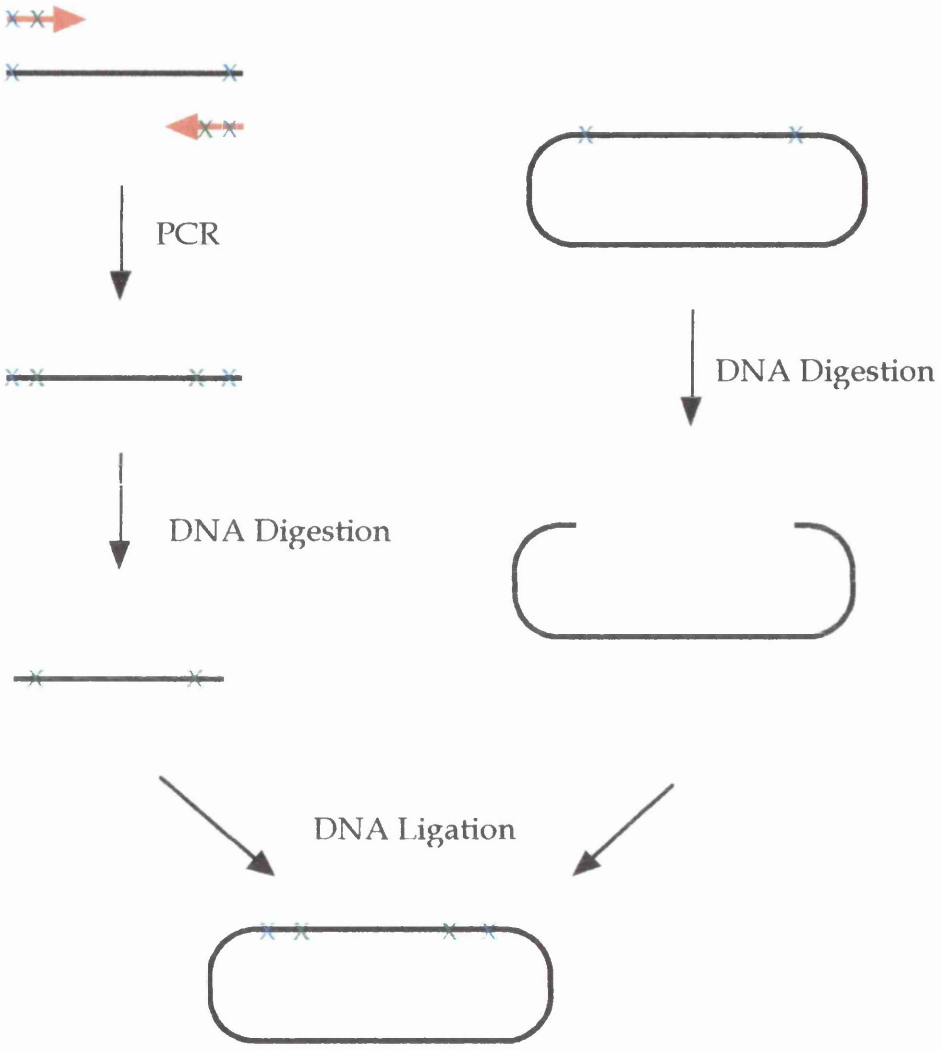


Fig. 5.2. PCR mutagenesis of the NS5A/NS5B cleavage site. The conserved restriction enzyme sites are illustrated by a blue X and the introduced restriction enzyme sites by a green X

check that no additional mutations had been introduced during the PCR reaction. Once the insert had been ligated into the tNS5T gene the presence of a compression made it difficult to determine whether the nucleotide sequence had been amplified correctly. This region was re-sequenced using the Sequenase DNA sequencing kit, using dITP instead of dCTP to overcome the problems associated with compression. When the region was sequenced in this manner the compressed sequence was readable and this confirmed that the sequence of the PCR product was correct (Fig. 5.3.). The DNA sequence determined by the Sequenase method of DNA sequencing contained an unreadable region of sequence but this region was confirmed as correct by the standard method of sequencing.

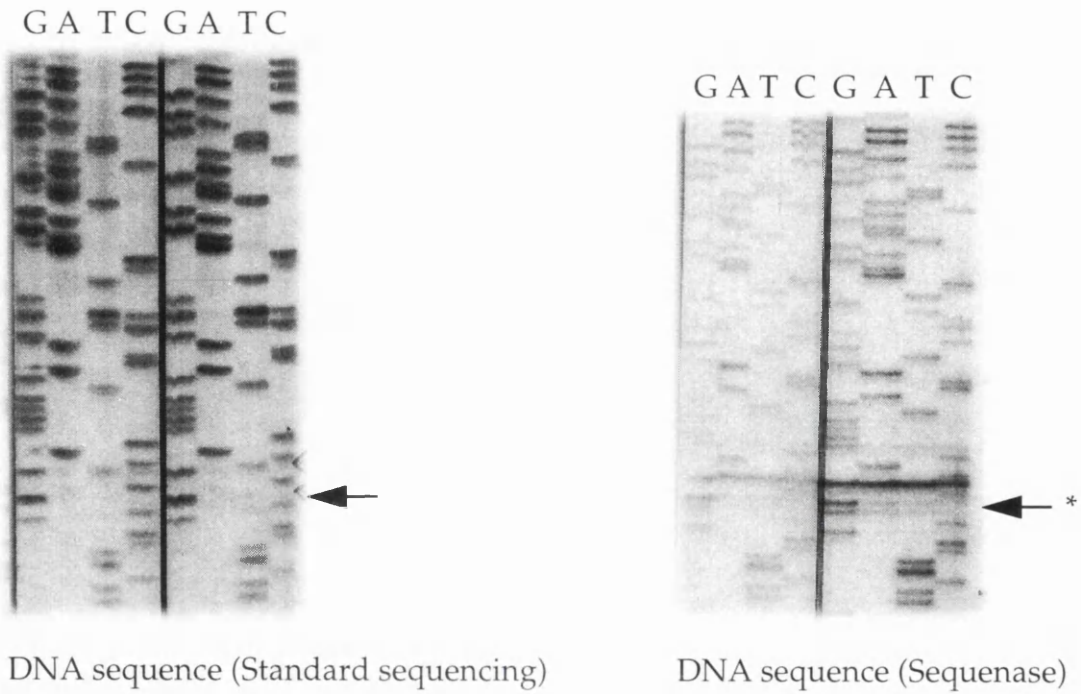
pTM1 contains a *Kpn* I site in the vector outwith the multiple cloning site and so could not be used for the initial mutagenesis step without carrying out a partial digest. Despite varying the DNA and restriction enzyme concentrations the partial digest always favoured the *Kpn* I site in pTM1 and so made cloning into this vector impossible. pTZ18 also contains a site within its multiple cloning site and so pTZ18 tNS5T could not be used for the initial cloning step. Therefore the tNS5T cDNA containing the additional restriction enzyme sites was subcloned into pGEM-1, a vector lacking a *Kpn* I site and so suitable for the introduction of the mutated NS5A/NS5B cleavage sites. Hence mutagenesis was initially carried out using pGEM tNS5T and then the mutated cleavage site subcloned into pTM1 tNS5T (as protein expression from cDNAs cloned within pTM1 was significantly higher). The tNS5T cDNA contains a *Bam* HI site 75 nucleotides upstream of the NS5A/NS5B cleavage site and an *Eco* RI site 605 nucleotides downstream. The pGEM tNS5T containing the mutated cleavage site was digested with *Bam* HI/*Eco* RI to produce a 680bp insert that was ligated into *Bam* HI/*Eco* RI digested pTM1 tNS5T.

The expression of the three tNS5T constructs (pTZ18 tNS5T, pTM1 tNS5T, pGEM tNS5T) containing the cleavage site with the *Kpn* I/*Bst* EII sites introduced was checked. The introduction of the additional restriction enzyme sites, which had no effect on the amino acid sequence, also had no effect on protein expression. All three constructs were also co-expressed with NS34A to ensure that the "altered" tNS5T substrate was still processed by the NS3 serine protease (Fig. 5.4.). Although the tNS5T substrate was expressed from all the vectors the level of expression from the pTM1 tNS5T plasmid was significantly higher than for the other vectors.

5.2 Mutagenesis of conserved residues within the cleavage site motif

The cleavage sites processed by the NS3 serine protease all essentially have the same amino acid motif. For the HCV Gla polyprotein this motif has been shown to consist of a conserved acidic P6 residue (aspartic or glutamic acid), a cysteine or threonine residue (both these amino acids contain the sulfhydryl group -SH) at P1 and a serine residue at P1'.

Sequence - TTCTTCCGC**GG**CGCACGGGGTGACCAGTGCGCGCCTGTCCAAGAGTAAGACATT



DNA sequence (Standard sequencing) DNA sequence (Sequenase)

Fig. 5.3. Comparison of different sequencing methods for determination of the DNA sequence at the NS5A/NS5B cleavage site. Two clones containing the altered cleavage site were sequenced by the standard DNA sequencing method (left) and this sequence resulted in a compressed area (◀—) whose sequence could not be determined. Using the Sequenase method (right) the compressed area was shown to contain the correct guanidine doublet (◀— *). A band in the T track was observed (left) two nucleotides after the doublet but this band is an artifact.

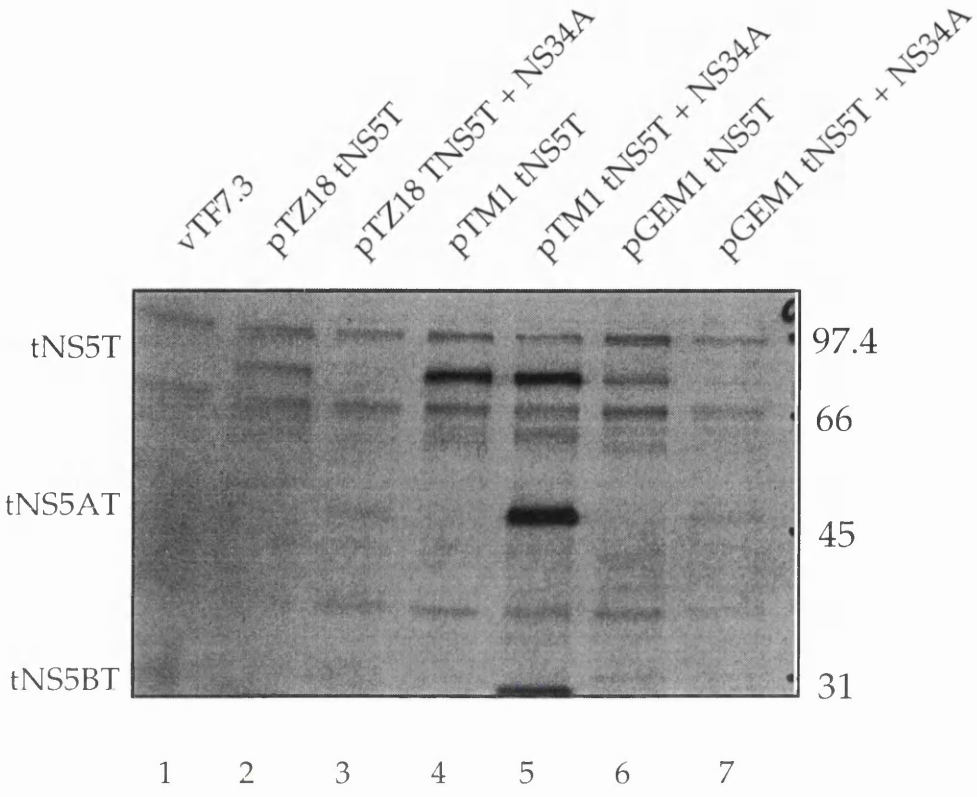


Fig. 5.4. Comparison of substrate expression from different plasmids transfected in mammalian cells. vTF7.3 infected CV-1 cells were transfected with different plasmids encoding the tNS5T substrate. Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours before lysis and then immuneprecipitation with the SV5-P-k antibody. In addition to untransfected cells (lane 1), the tNS5T substrate was expressed from pTZ18 (lanes 2 and 3), pTM1 (lanes 4 and 5) and pGEM-1 (lanes 6 and 7). The levels of expression of the substrate varied depending on the vector. The tNS5T substrate was expressed at the highest level when expressed from pTM1 (lanes 4 and 5). When the substrate was cotransfected with NS34A cleavage products were produced(lanes 3, 5 and 7). As the level of expression was higher from the pTM1 tNS5T construct (lane 5) processing was observed more clearly than for the other vectors (lanes 3 and 7). Molecular weight markers are located at the right of the gel.

```
P6          P1P1'
D(E) X X X X C(T) / S
```

The cleavage site is processed between the P1 residue and the P1′ residue. The importance and/or essential nature of these amino acids was not known. The threonine residue at P1′ is only observed for the NS3/NS4A cleavage site. This site is the only one processed *in cis* by the NS3 serine protease within the HCV polyprotein

For mutagenesis of amino acids at the NS5A/NS5B cleavage site two complementary oligonucleotides containing the necessary nucleotide changes were annealed together. Annealing of these oligonucleotides produced the cassette required for ligation into the tNS5T gene. For each of the single amino acid changes at P6, P1 or P1′ the 5′ to 3′ oligonucleotide was designed to give three possible amino acid changes at each codon. This was achieved by using a mixture of three nucleotide precursors in the synthesis at a specific position within the nucleotide sequence. For each codon targetted, the first nucleotide was replaced by a mixture of the other three nucleotides. As the first nucleotide of the codon primarily determines the amino acid encoded this mutagenesis method would produce three alternative amino acids at each position. A complementary oligonucleotide encoding the wild type sequence was annealed to the mutant oligonucleotides to form the cassette containing the mutant cleavage site (Fig. 5.5.).

All nine possible mutants were obtained using this procedure, and the DNAs were checked to ensure that no extra mutations had been introduced. These clones were transfected into mammalian cells both with and without the protease construct to determine expression and processing by the NS3 serine protease (Figs. 5.6.-5.12.). Each set of positional mutants (P6, P1 and P1′) was analysed in duplicate to ensure that the results obtained were consistent and that there was no aberrations.

The effect of mutating the P6 residue of the NS5A/NS5B cleavage site was investigated (Figs. 5.6. and 5.7.). The replacement of the wild type P6 glutamic acid with glutamine or lysine produced a cleavage site that was still processed by the NS3 serine protease but with a lower level of efficiency. The introduction of a STOP codon at P6 produced a truncated NS5A substrate that did not contain the NS5A/NS5B cleavage site and was not processed by NS3.

The replacement of the wild type cysteine residue at P1 produced cleavage sites that were unsuitable for cleavage by the NS3 serine protease (Figs. 5.8. and 5.9.). When a serine residue was introduced at P1 there was a low level of processing but replacement with arginine or glycine resulted in the abolition of processing.

The effect of mutating the P1′ residue on processing depended on the amino acid introduced (Figs. 5.10. and 5.11.). Cleavage sites containing threonine and alanine were processed whereas the cleavage site with proline at P1′ was not cleaved. In Fig. 5.12. all nine mutants were analysed

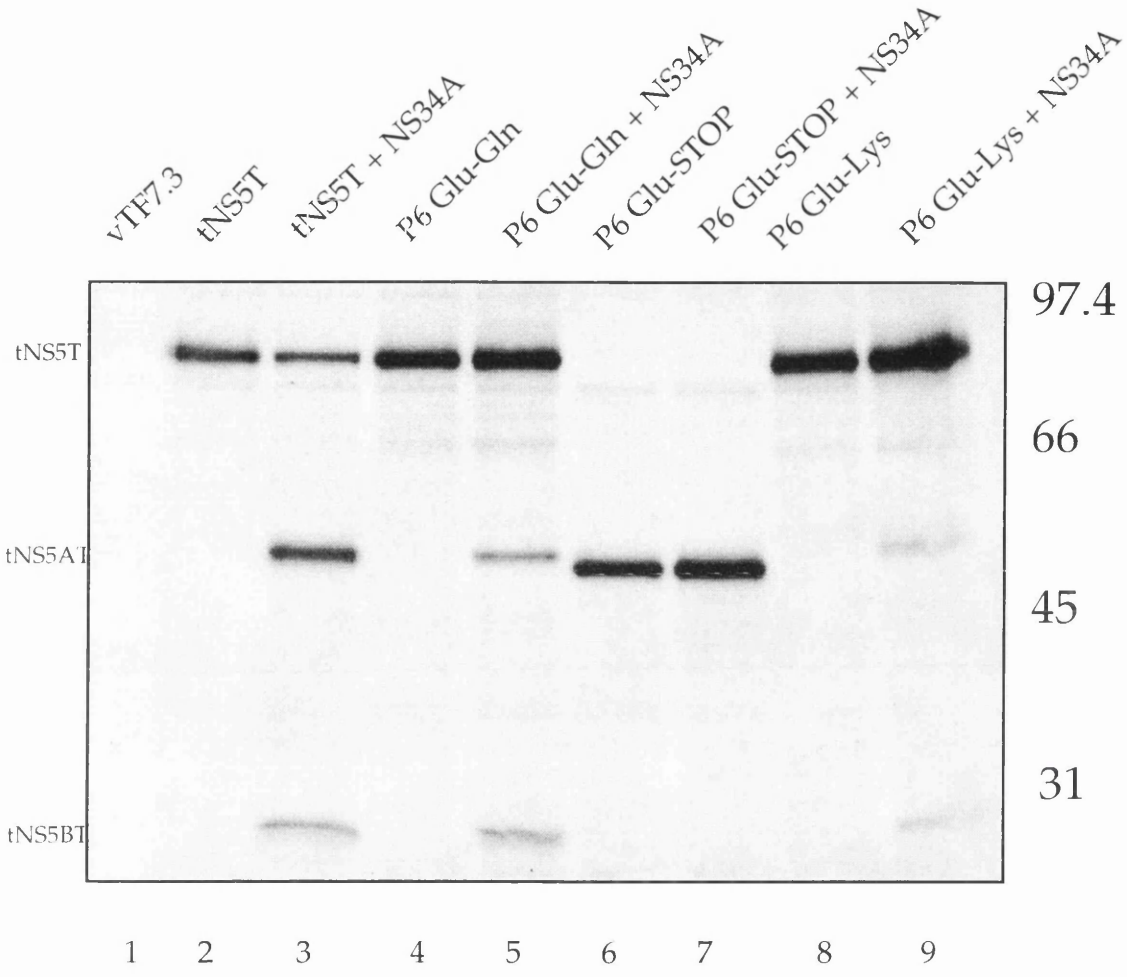


Fig. 5.6. Mutation of the P6 amino acid in the NS5A/NS5B cleavage site within the tNS5T substrate. CV-1 cells were infected with vTF7.3 before transfection either with HCV plasmids (lanes 2 to 9) or without DNA as a control (lane 1). In lane 2 the wild type tNS5T substrate was expressed alone and in lane 3 the substrate was coexpressed with NS34A. A tNS5T substrate containing a glutamine residue at P6 was transfected alone (lane 4) or with NS34A (lane 5). A tNS5T substrate with a STOP codon at P6 was transfected with (lane 7) or without (lane 6) NS34A. A tNS5T substrate containing a lysine at P6 was transfected alone (lane 8) or with NS34A (lane 9). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours before cell lysis and immunoprecipitation with the SV5-P-k antibody. Both the substrates containing P6 Gln and P6 Lys were processed by NS34A (lanes 5 and 9) albeit with different efficiencies. The tNS5T substrate containing the STOP codon at P6 produced a truncated substrate that did not contain the NS5A/NS5B cleavage site. This substrate was not processed by NS34A. Molecular weight markers are located at the right of the gel.

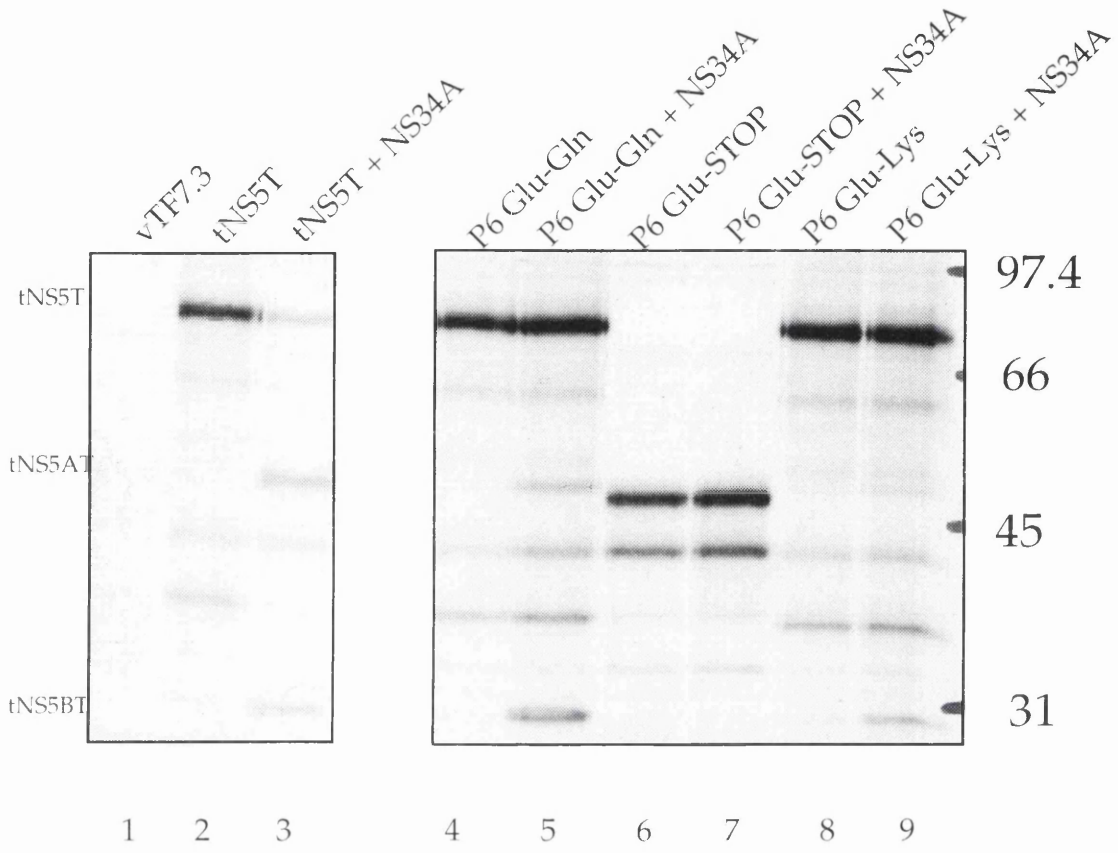


Fig. 5.7. Repeat analysis of mutation of the P6 amino acid in the NS5A/NS5B cleavage site within the tNS5T substrate. CV-1 cells were infected with vTF7.3 before transfection either with HCV plasmids (lanes 2 to 9) or without DNA as a control (lane 1). In lane 2 the wild type tNS5T substrate was expressed alone and in lane 3 the substrate was coexpressed with NS34A. A tNS5T substrate containing a glutamine residue at P6 was transfected alone (lane 4) or with NS34A (lane 5). A tNS5T substrate with a STOP codon at P6 was transfected with (lane 7) or without (lane 6) NS34A. A tNS5T substrate containing a lysine at P6 was transfected alone (lane 8) or with NS34A (lane 9). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours before cell lysis and immuneprecipitation with the SV5-P-k antibody. Both the substrates containing P6 Gln and P6 Lys were processed by NS34A (lanes 5 and 9) albeit with different efficiencies. The tNS5T substrate containing the STOP codon at P6 produced a truncated substrate that did not contain the NS5A/NS5B cleavage site and so was not processed by NS34A. Molecular weight markers are located at the right of the gel.

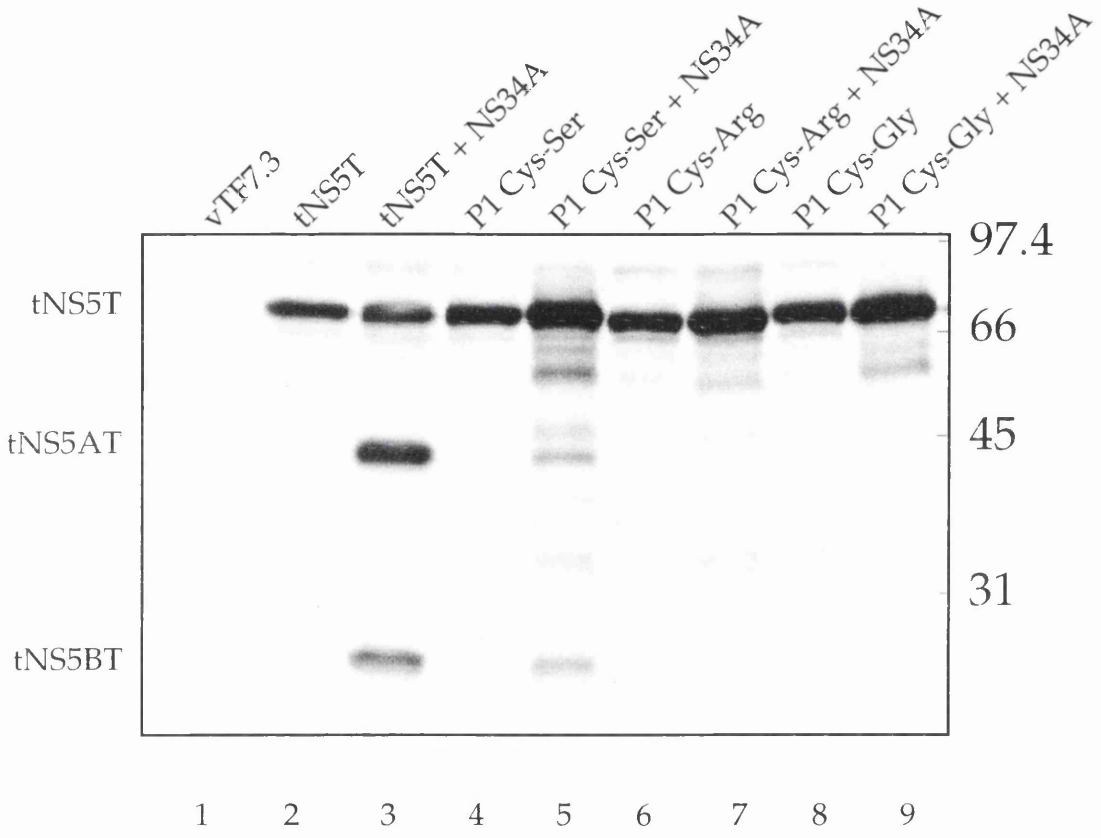


Fig. 5.8. Mutation of the P1 residue in the NS5A/NS5B cleavage site within tNS5T. CV-1 cells were infected with vTF7.3 before transfection either with HCV plasmids (lanes 2 to 9) or without DNA as a control (lane 1). In lane 2 the wild type tNS5T substrate was expressed alone and in lane 3 the substrate was coexpressed with NS34A. A tNS5T substrate containing a serine residue at P1 was transfected alone (lane 4) or with NS34A (lane 5). A tNS5T substrate containing an arginine residue at P1 was transfected with (lane 7) or without (lane 6) NS34A. A tNS5T substrate containing a glycine at P1 was transfected alone (lane 8) or with NS34A (lane 9). Fifteen hours postransfection the cells were labelled with ^{35}S-methionine for 2 hours , lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. None of the mutant substrates were processed by NS34A (lanes 7 and 9) with the exception of the P1 Ser tNS5T substrate that was processed to a small degree (lane 5). Molecular weight markers are indicated on the right.

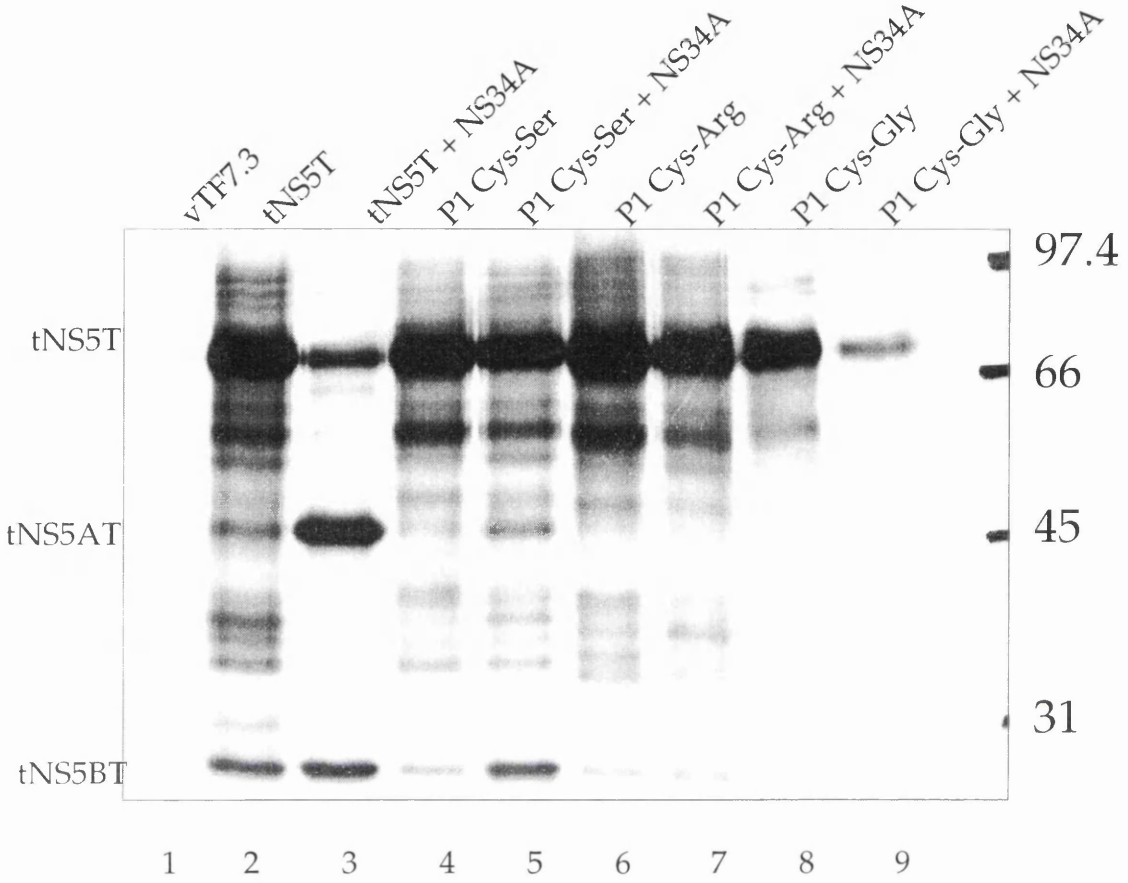


Fig. 5.9. Repeat of the mutational analysis of the P1 residue in the NS5A/NS5B cleavage site within tNS5T. CV-1 cells were infected with vTF7.3 before transfection either with HCV plasmids (lanes 2 to 9) or without DNA as a control (lane 1). In lane 2 the wild type tNS5T substrate was expressed alone and in lane 3 the substrate was coexpressed with NS34A. A tNS5T substrate containing a serine residue at P1 was transfected alone (lane 4) or with NS34A (lane 5). A tNS5T substrate containing an arginine residue at P1 was transfected with (lane 7) or without (lane 6) NS34A. A tNS5T substrate containing a glycine at P1 was transfected alone (lane 8) or with NS34A (lane 9). Fifteen hours postransfection the cells were labelled with ^{35}S-methionine for 2 hours , lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. None of the mutant substrates were processed by NS34A (lanes 7 and 9) with the exception of the P1 Ser tNS5T substrate that was processed to a small degree (lane 5). Molecular weight markers are indicated on the right.

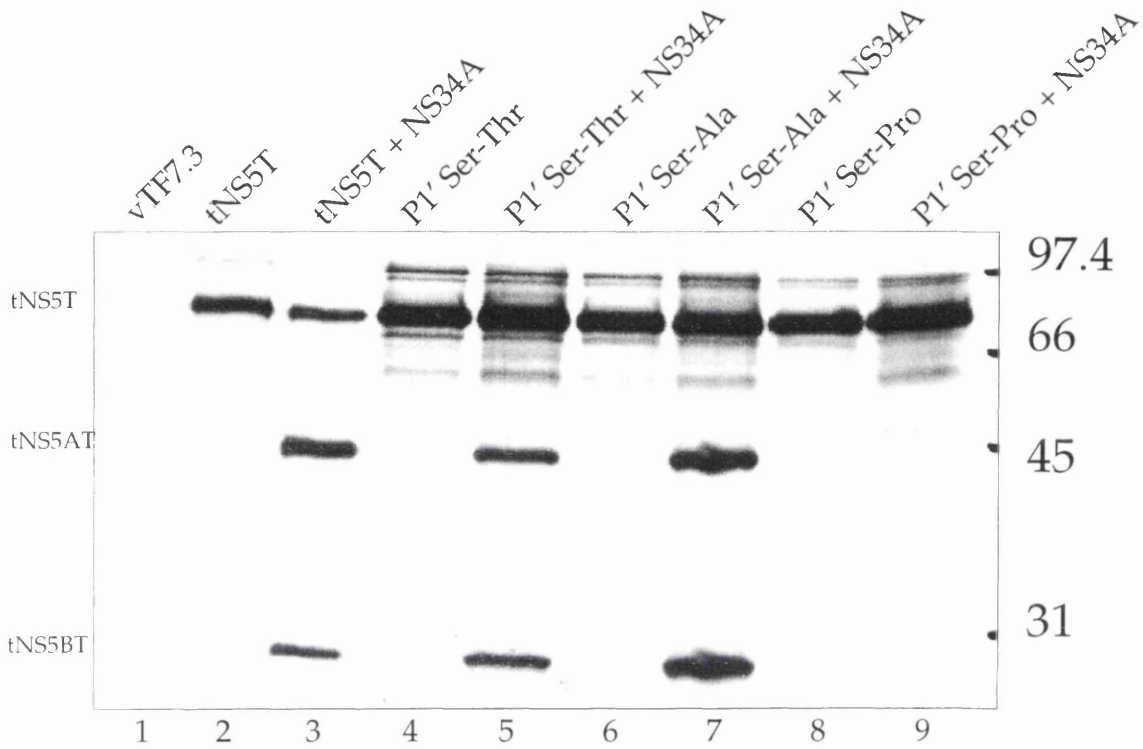


Fig. 5.10. Mutation of the P1′ amino acid in the NS5A/NS5B cleavage site within the tNS5T substrate. CV-1 cells were infected with vTF7.3 before transfection either with HCV plasmids (lanes 2 to 9) or without DNA as a control (lane 1). In lane 2 the wild type tNS5T substrate was expressed alone and in lane 3 the substrate was coexpressed with NS34A. A tNS5T substrate containing a threonine residue at P1′ was transfected alone (lane 4) or with NS34A (lane 5). A tNS5T substrate containing an alanine residue at P1′ was transfected with (lane 7) or without (lane 6) NS34A. A tNS5T substrate containing a proline at P1′ was transfected alone (lane 8) or with NS34A (lane 9). Fifteen hours postransfection the cells were labelled with ^{35}S-methionine for 2 hours , lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The effect of an amino acid substitution at P1′ depended on the amino acid introduced. A tNS5T substrate containing an alanine at P1′ was processed at almost wild type efficiency (lane 7). A threonine at P1′ resulted in a substrate processed less efficiently (lane 5) whereas a substrate containing proline at P1′ was not processed. The molecular weight markers are indicated on the right.

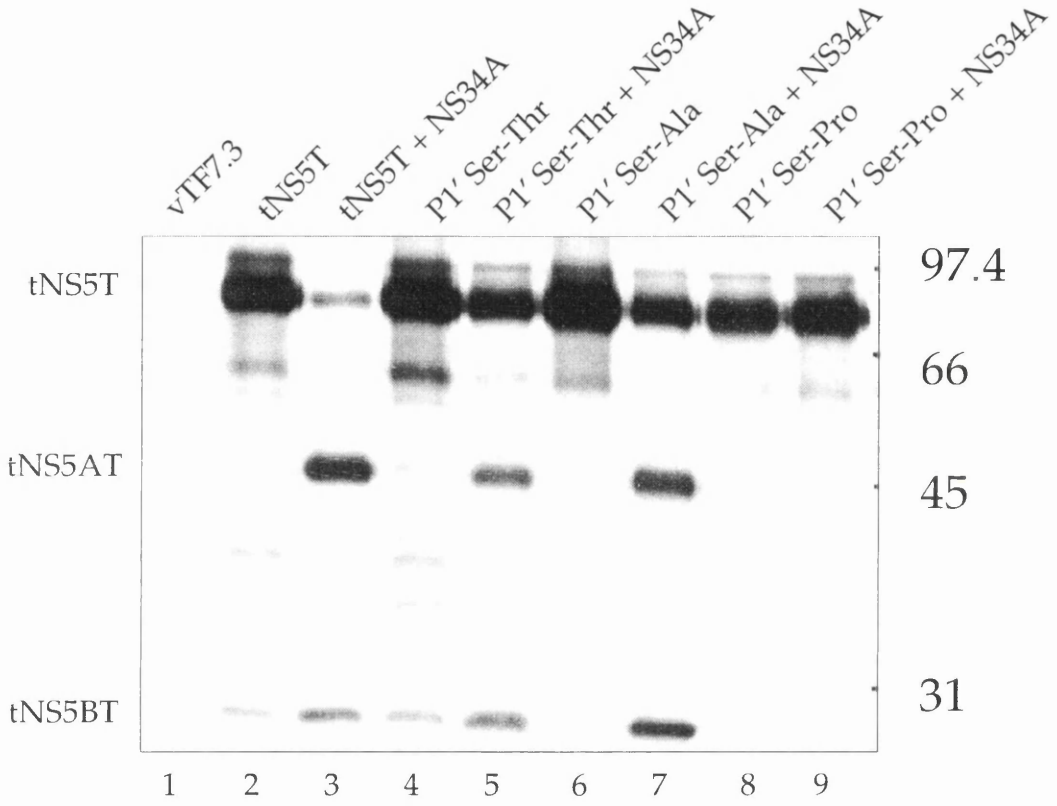


Fig. 5.11. Repeat analysis of mutation of the P1′ amino acid in the NS5A/NS5B cleavage site within the tNS5T substrate. CV-1 cells were infected with vTF7.3 before transfection either with HCV plasmids (lanes 2 to 9) or without DNA as a control (lane 1). In lane 2 the wild type tNS5T substrate was expressed alone and in lane 3 the substrate was coexpressed with NS34A. A tNS5T substrate containing a threonine residue at P1′ was transfected alone (lane 4) or with NS34A (lane 5). A tNS5T substrate containing an alanine residue at P1′ was transfected with (lane 7) or without (lane 6) NS34A. A tNS5T substrate containing a proline at P1′ was transfected alone (lane 8) or with NS34A (lane 9). Fifteen hours postransfection the cells were labelled with ^{35}S-methionine for 2 hours , lysed and then immunepreciptated with the SV5-P-k antibody before analysis by SDS-PAGE. The effect of an amino acid substitution at P1′ depended on the amino acid introduced. A tNS5T substrate containing an alanine at P1′ was processed at almost wild type efficiency (lane 7). A threonine at P1′ resulted in a substrate processed less efficiently (lane 5) whereas a substrate containing proline at P1′ was not processed. The molecular weight markers are indicated on the right.

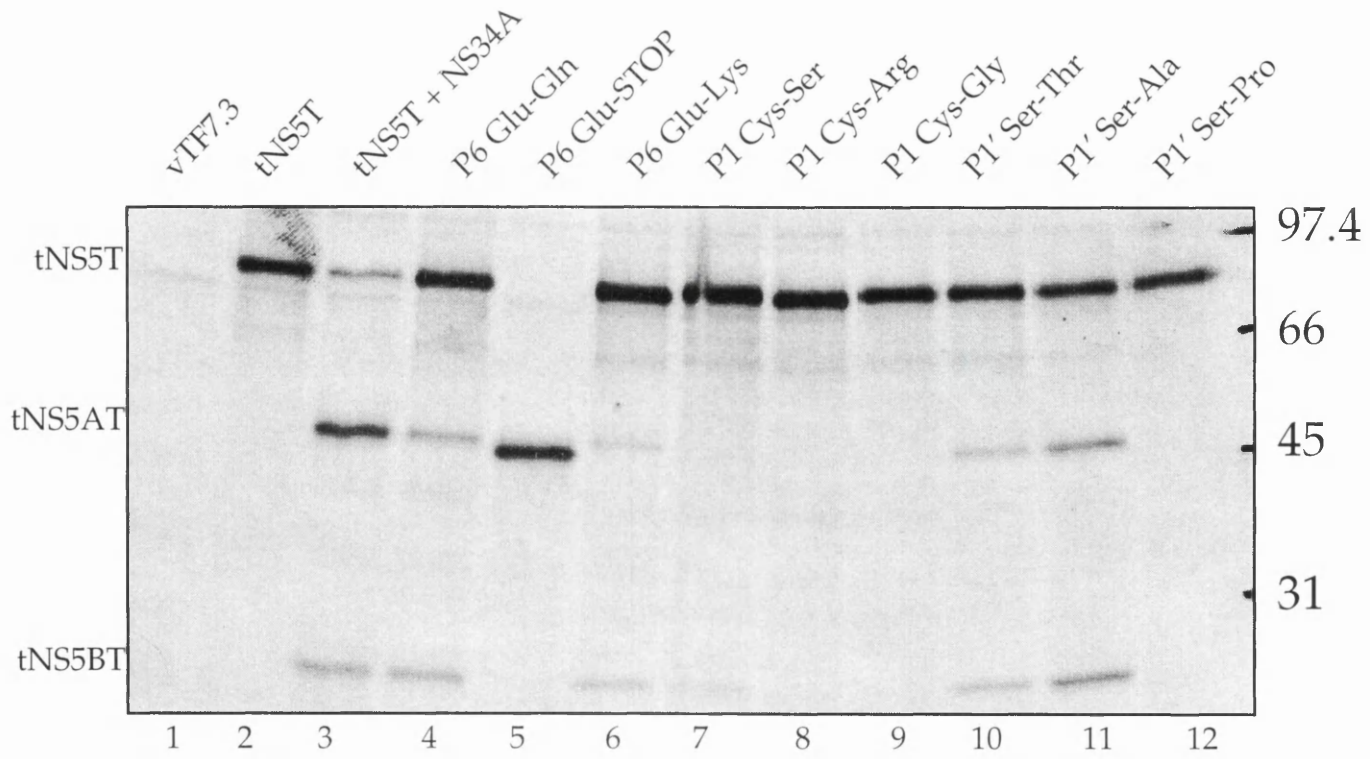


Fig. 5.12. Processing of all nine mutant cleavage sites (P6, P1, P1′). vTF7.3 infected CV-1 cells were either untransfected (lane 1) or transfected with 1µg tNS5T (lane 2) or 1µg substrate + 5µg NS34A. The substrates transfected with NS34A were tNS5T (lane 3), P6 Gln (lane 4), P6 STOP (lane 5), P6 Lys (lane 6), P1 Ser (lane 7), P1 Arg (lane 8), P1 Gly (lane 9), P1′ Thr (lane 10), P1′ Ala (lane 11), P1′ Pro (lane 12). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. Every amino acid substitution illustrated above was encoded in the NS5A/NS5B cleavage site within the tNS5T substrate. This analysis illustrates the importance of each residue (P6, P1 and P1′) in substrate specificity. Molecular weight markers are indicated on the right.

by SDS-PAGE after co-transfection with NS34A to illustrate the effects of all the mutations together.

The replacement of the acidic amino acid at P6 with either the basic lysine residue or the uncharged polar glutamine resulted in a significant fall in the efficiency of processing. The introduction of a stop codon upstream of the cleavage site resulted in a slightly truncated NS5A protein that did not contain the NS5A/NS5B cleavage site. This protein was included as a negative control in the analysis to demonstrate that the system was functioning correctly. Mutation of the P1 cysteine residue resulted in the abolition of processing except for a cysteine to serine substitution where a small amount of processing was observed at this site.

The effect of substitution at the P1′ amino acid depended on the amino acid change. Substitution with the nonpolar proline residue completely abolished processing whereas replacement with alanine still permitted a significant degree of processing. Slightly greater inhibition of processing resulted from replacement of the P1′ serine residue with the uncharged, nonpolar threonine residue.

The efficiency of processing at these mutant cleavage sites was estimated, with respect to wild type cleavage efficiency, in Table 5.1.

Amino Acid Substitution	Relative % Cleavage Efficiency
Wild Type	100
P6 Glu - Gln	30
P6 Glu - Lys	15-25
P1 Cys - Ser	<10
P1 Cys - Arg	0
P1 Cys - Gly	0
P1′ Ser - Thr	30-40
P1′ Ser - Ala	55-60
P1′ Ser - Pro	0

Table 5.1. Mutation of the amino acids comprising the cleavage site motif. The relative effect on cleavage with respect to the wild type is illustrated. Mutation of the P6 residue significantly inhibited the level of processing. An even stronger inhibitory effect was observed when the P1 residue was replaced. The effect of mutation at the P1′ residue depended on the amino acid introduced.

Chapter 6 - Investigation of potential cleavage sites within the HCV Gla polyprotein

6.1. Determination of potential cleavage sites within the HCV Gla polyprotein.

Only four sites within the HCV polyprotein are cleaved by the NS3 serine protease. Processing at these sites is essential for the generation of the HCV non-structural proteins downstream of NS3. If the NS3 serine protease cleaved at any other site within the HCV polyprotein it had yet to be reported. The amino acid sequence of the HCV Gla polyprotein was examined to determine whether there were any other cleavage sites that could potentially be processed by the NS3 serine protease. All the sites cleaved by the NS3 serine protease have an aspartic or glutamic acid residue followed by a cysteine or threonine residue six residues downstream and a serine residue seven residues downstream. The HCV Gla polyprotein was searched for this cleavage site motif using the Stringsearch program. The results of the search for D(E) X X X X C(T) S are illustrated in Table 6.1.

Position	Sequence	Location
1652 - 1658	DLEVVTS	NS3/NS4A
1706 - 1712	DEMEECS	NS4A/NS4B
2967 - 1973	DCTTPCS	NS4B/NS5A
2167 - 2173	DVAVLTS	In NS5A
2415 - 2421	EDVVCCS	NS5A/NS5B
2779 - 2785	DLELITS	In NS5B
2781 - 2787	ELITSCS	In NS5B

Table 6.1. Cleavage site motif search with HCV Gla. The HCV Gla polyprotein was searched for any amino acid sequence that corresponded to the NS3 cleavage site motif. As well as the cleavage sites processed to give the non-structural proteins, three additional, potential cleavage sites were identified within the NS5 region of the polyprotein. A potential cleavage site motif similar to the NS3/NS4A cleavage site, in that it contained a threonine residue at P1, was located within NS5A. The threonine residue at this position is unique to the NS3/NS4A cleavage site, a site that is processed *in cis*. Two potential cleavage sites within the NS5B region were found that overlap each other by five amino acids. One site contained a threonine residue at P1 and the other contained a cysteine residue at P1 (as is seen for the majority of cleavage sites, all of which are processed *in trans*).

6.2. Investigation of processing at the potential cleavage site within NS5A.

Cleavage at the first of these potential cleavage sites within the HCV Gla polyprotein would produce a truncated NS5A protein with its N-terminus at amino acid 2173. The N-terminus of the tNS5T substrate is at amino acid 2105, 68 amino acids upstream of this theoretical cleavage site. If the potential cleavage site was cleaved within tNS5T by NS3 then two NS5A cleavage products could be generated, depending on how efficiently this potential site was processed. These two NS5A products, although relatively similar in size, could be differentiated by immuneprecipitation using two different antibodies. The SV5-P-k antibody would only identify the full length tNS5AT product as the epitope tag it recognised was present at the N-terminus of the tNS5T substrate. If processing occurred at the potential cleavage site then the epitope tag would be attached to the small 68 amino acid protein generated by this cleavage event. The NS5(Baculo) antibody against the C-terminal region of the NS5A protein (supplied by GlaxoWellcome) would identify both the truncated NS5A proteins.

Duplicate cotransfections were carried out in vTF7.3 infected CV-1 cells. The tNS5T substrate cDNA and either the NS3 or the NS34A construct were cotransfected together. The potential cleavage site resembled the NS3/NS4A cleavage site as it contained a threonine at P1. As the NS3/NS4A cleavage site is processed *in cis* by NS3 alone the cotransfections were carried out using the HCV Gla NS3 construct as well as the NS34A construct. After radiolabelling with ^{35}S-methionine, the cells were lysed and each pair of extracts (tNS5T + NS3 and tNS5T + NS34A) were immunoprecipitated with either the SV5-P-k antibody or the NS5(Baculo) antibody. The immunoprecipitated lysates were analysed by SDS-PAGE (Fig. 6.1.). The substrate was only cleaved only when it was cotransfected with NS34A. The NS5A cleavage products identified by both the SV5-P-k antibody and the NS5A(Baculo) antibody were the same size. No additional NS5A band was detected. These results show that only the NS5A/NS5B cleavage site was processed by the NS3 serine protease within the tNS5T substrate. Cleavage at this site was observed to take place by Markland *et al*, 1996.

6.3. Investigation of the potential "*trans*" cleavage site in NS5B.

The other two potential cleavage sites, at amino acids 2779 - 2785 and 2781 - 2787, were situated downstream of the C-terminus of the tNS5T substrate (amino acid 2623). Whether these potential cleavage sites were processed within NS5B could not be determined using the tNS5T substrate. The full length NS5B protein, which contains these theoretical sites, was reported to be unstable when expressed in cells. The instability of the NS5B protein may be due to proteolysis by cellular-encoded enzymes (Lin *et al.*, 1994b). As no efficient antibody against the NS5B protein was available in the laboratory this protein could not be detected by immunoprecipitation even if it was stable, and so the full length NS5 substrate could not be used to investigate processing at these potential sites. As the assay for NS3 serine protease activity analysed processing *in trans* the cleavage site containing the cysteine residue at P1 (ELITS**C**S) was chosen for study.

An amino acid comparison was carried out between this potential cleavage site (including the flanking amino acids) and the cleavage sites processed by NS3 within the HCV Gla polyprotein (Fig. 6.2.). An amino acid sequence comparison between all HCV isolates showed that the potential cleavage site (ELITSCS) was present in every genotype at the same position.

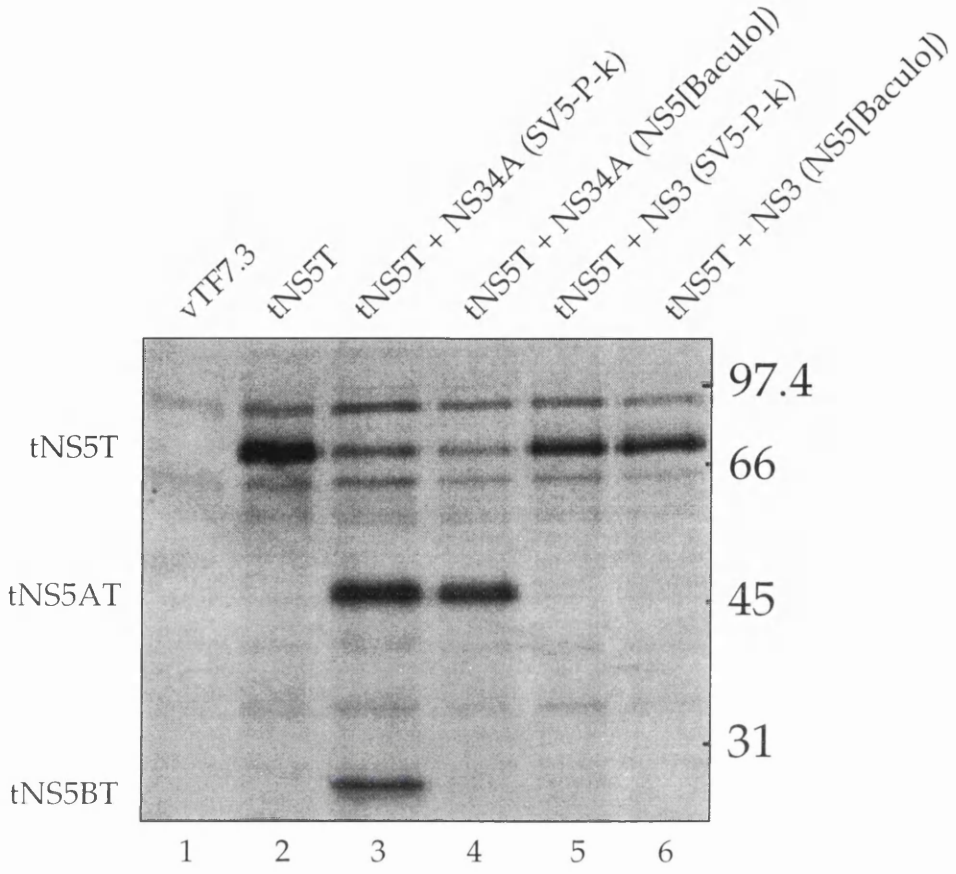


Fig. 6.1. Analysis of processing of the tNS5T substrate and determination whether processing occurs at a potential site. vTF7.3 infected CV-1 cells were transfected with HCV cDNAs (lanes 2 to 6) or remained untransfected as a control (lane 1). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immunepreciptated before analysis by SDS-PAGE. The tNS5T substrate was transfected alone (lane 2). In lanes 3 and 4 the tNS5T substrate was cotransfected with NS34A. For lane 3 the substrate was immuneprecpitated with the SV5-P-k antibody but for lane 4 the substrate was detected by the NS5 (Baculo) antiserum. In lanes 5 and 6 the tNS5T substrate was cotransfected with NS3 and immunepreciptated with either SV5-P-k (lane 5) or NS5 [Baculo] (lane 6). The theoretical cleavage site within NS5A is situated close to the C-terminus of the tNS5T substrate. If the potential cleavage site was cleaved a smaller tNS5AT product would be identified by the NS5A/NS5B antibody and not by the SV5-P-k antibody. No additional cleavage product was observed when the substrate was cotransfected with either NS3 or NS34A (lanes3, 4, 5, 6). The molecular weight markers are located at the right of the gel.

LYREF**D**EMEE**CS**	QHLPY-NS4A/NS4B
QWISS**D**CTTP**CS**	GSWLR-NS4B/NS5A
SEAGT**E**DVVC**CS**	MSYSW-NS5A/NS5B
PEYDL**E**LITS**CS**	SNVSV-In NS5B

Fig. 6.2. Sequence comparison between the observed sites cleaved *in trans* by the NS3 serine protease and the potential site in NS5B. The amino acids constituting the cleavage site motif are illustrated in bold.

As for the mutagenesis of the NS5A/NS5B cleavage site within tNS5T, two complementary oligonucleotides were designed. Each oligonucleotide encoded the NS5A/NS5B cleavage site with the wild type sequence from P5 - P2 (DVVC) replaced with LITS.

Following successful cloning into tNS5T and confirmation by sequencing, the tNS5T substrate containing the mutated cleavage site (tNS5T LITS) was cotransfected with NS34A into vTF7.3 infected CV-1 cells to determine whether this substrate was processed (Fig. 6.3.) using the standard *trans*-cleavage assay. When tNS5T LITS was coexpressed with NS34A no cleavage products were detected. When quantitation of processing was carried out a very small amount of processing was detected but this value was negligible. The lack of processing at the NS5A/NS5B site was seen despite the presence of the consensus amino acids at P6, P1 and P1'. These three amino acids at P6, P1 and P1' had been previously suggested to be sufficient to define a cleavage site suitable for processing by the NS3 serine protease.

When leucine, isoleucine, threonine and serine were present at positions P5 to P2 in the NS5A/NS5B cleavage site, this site could not be cleaved by the NS3 protease. The effect on the cleavage site could be due to an individual amino acid or a cumulative effect of more than one residue. The amino acid sequences of NS3 processed cleavage sites were examined to determine whether any of these four amino acids were present in different HCV subtypes or genotypes (Table 6.2.).

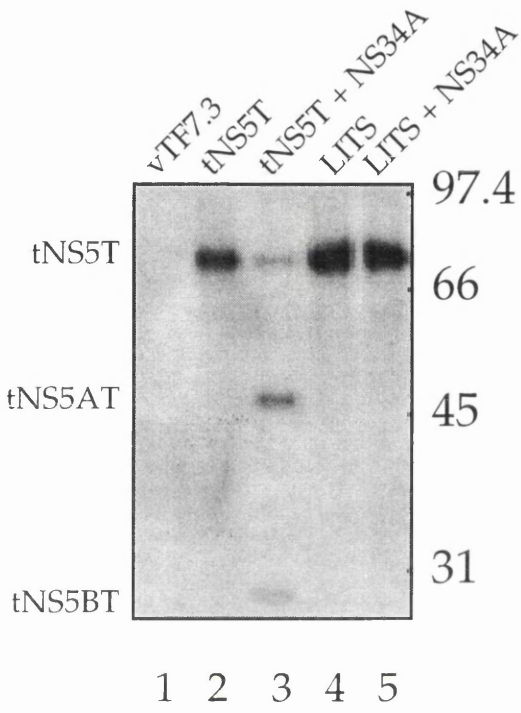


Fig. 6.3. Analysis of processing of the tNS5T substrate containing the P5 - P2 substitution from Asp, Val, Val, Cys to Leu, Iso, Thr, Ser (LITS). CV-1 cells were infected with vTF7.3 before transfection with tNS5T (lane 2), tNS5T + NS34A (lane 3), tNS5T LITS (lane 4) or tNS5T LITS + NS34A (lane 5). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours , lysed and immuneprecipitated with the SV5-P-k antibody , as were untransfected cells (lane 1). All these samples were analysed by SDS-PAGE. The tNS5T-LITS substrate was not cleaved by coexpressed NS34A (lane 5). The molecular weight markers are located at the right of the gel.

Position Within Cleavage Site	Amino Acids Present Within Site
5	E, C, S, Y, D, G, **L**
4	M, P, S, T, V, I, **E**
3	E, T, V, I, A, S, **V**
2	E, P, C, **V**

Table 6.2. Cleavage sites processed by the NS3 serine protease were compared from HCV genotypes 1-6. The amino acids found at positions 5, 4, 3 and 2 are shown above. The amino acids illustrated in bold correspond to those found at the highly conserved NS3/NS4A cleavage site (the only site processed *in cis* within the HCV polyprotein).

The presence of the LITS amino acid sequence from P5 to P2 in an NS3 protease cleavage site was not found in any HCV polyprotein for all six genotypes. A leucine at P5, an isoleucine at P4 and a threonine at P3 were all present individually. A serine residue at P2 was not seen in any other cleavage site. The presence of this residue could have a detrimental effect on the formation of a cleavage site, resulting in a site that could no longer be processed by the NS3 serine protease.

Another feature of the potential cleavage site was the presence of a serine residue at the P2′ position (Fig. 6.2.). Of the cleavage sites processed *in trans* by the NS3 protease only the NS4A/NS4B cleavage site in certain genotypes contained a serine residue at P2′. For these cleavage sites the P1′ residue is always an alanine residue. Some NS3/NS4A cleavage sites contain a serine residue at both P1′ and P2′, but as this site is processed *in cis* it is possible that the flanking residues are not as influential as they are for the *trans*-cleavage sites. Although the P2′ serine residue was not included in the initial potential cleavage site analysis the effect of this amino acid at P2′ was investigated. It was hypothesised that these serine residues could be phosphorylated, inhibiting access to the cleavage site to NS3, although no analyses were carried out to determine whether these residues were actually phosphorylated.

To analyse the effect of a serine at P2, two approaches were used. The first potential cleavage site was constructed in tNS5T with the wild type NS5A/NS5B amino acid sequence from P6 to P1′ except for a serine residue at P2. A second potential cleavage site was constructed with the LITS amino acid sequence in tNS5T LITS substituted at P2 with the wild type cysteine residue to give the amino acid sequence LITC from P5 to P2. These two constructs (tNS5T P2S and tNS5T LITC) were cloned and sequenced. A wild type NS5A/NS5B cleavage site with a serine residue at P2′ was constructed to give the substrate tNS5T P2′S. These cleavage sites are illustrated below.

P2 Cys - Ser	SEAGT EDVV**S**C/SMSY
LITC	SEAGT ELIT**C**C/SMSY
P2′ Met - Ser	SEAGT EDVVCC/S**S**SY

As well as these substrates an additional substrate was constructed with the LITCY amino acid sequence from P5 to P1 (tNS5T LITCY). This cleavage site was the result of a mistake during oligonucleotide synthesis. This substrate was also included in these experiments. All four of these substrate clones were cotransfected with NS34A into vTF7.3 infected CV-1 cells and the cells radiolabelled, lysed and immuneprecipitated before analysis by SDS-PAGE (Figs. 6.4., 6.5.).

The tNS5T substrate with the wild type P2 cysteine residue of the NS5A/NS5B cleavage site replaced with a serine residue was cleaved with a much lower efficiency in comparison to the wild type when it was coexpressed with NS34A. The tNS5T LITC substrate with the wild type cysteine residue substituted at P2 of the cleavage site was processed at an observable, but far from optimal, level. Little processing was observed for the substrate containing the ELITCYS cleavage site. The tNS5T P2'S substrate was cleaved by coexpressed NS34A but the efficiency of processing was only slightly higher than that observed for tNS5T P2S. The efficiency of processing estimated for these substrates is summarised below.

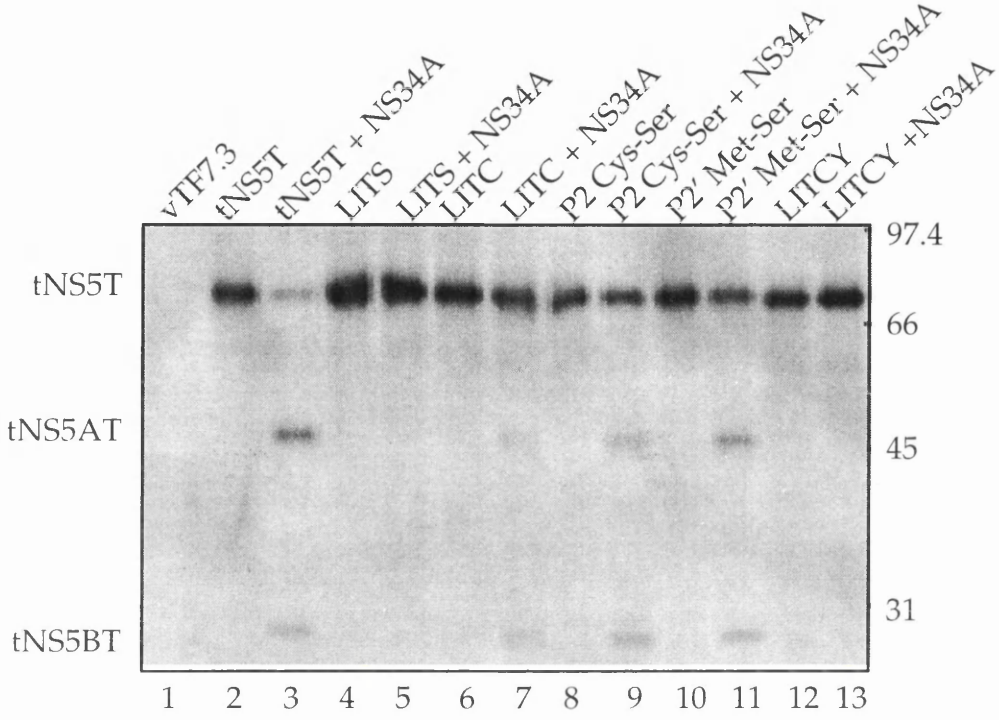


Fig. 6.4. Analysis of the effects on processing after further mutagenesis of the theoretical LITS cleavage site. CV-1 cells were infected with vTF7.3 before transfection with tNS5T (lane 2), tNS5T + NS34A (lane 3), tNS5T LITS (lane 4), tNS5T LITS + NS34A (lane 5), tNS5T LITC (lane 6), tNS5T LITC + NS34A (lane 7), tNS5T P2 Ser(lane 8), tNS5T P2 Ser + NS34A (lane 9), tNS5T P2' Ser (lane 10), tNS5T P2' Ser + NS34A (lane 11), tNS5T LITCY (lane 12) or tNS5T LITCY + NS34A (lane 13). Fifteen hours post-transfection the cells were labelled with ^{35}S methionine for 2 hours, lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE , as were untransfected cells (lane 1). The introduction of the wild type cysteine residue at P2 within the LITS cleavage site resulted in detectable processing of the substrate (lane 7), albeit at significantly lower levels compared to the wild type substrate. The introduction of a serine residue at either P2 or P2' had an inhibitory effect on the susceptibility to processing of the NS5A/NS5B cleavage site (lanes 9 and 11 respectively). The introduction of the tyrosine residue at P1 of the LITC cleavage site had an inhibitory effect on susceptibility to processing (lane 13). The molecular weight markers are located at the right of the gel.

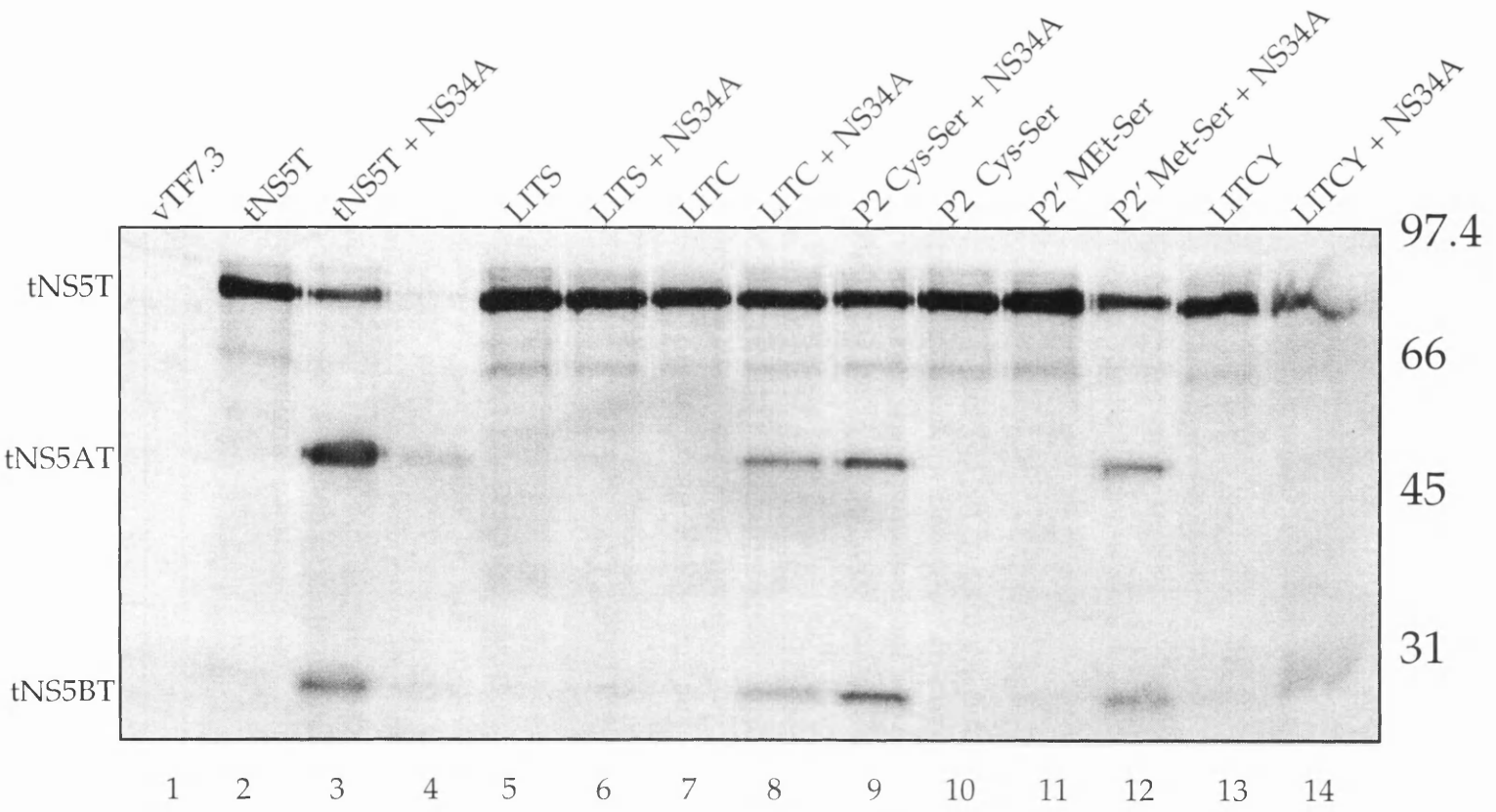


Fig. 6.5. Repeat analysis of of the theoretical LITS cleavage site. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with HCV cDNAs. Cells were transfected with tNS5T (lane 2), tNS5T + NS34A (lane 3), tNS5T LITS (lane 5), tNS5T LITS + NS34A (lane 6), tNS5T LITC (lane 7), tNS5T LITC + NS34A (lane 8), tNS5T P2 Ser (lane 10), tNS5T P2 Ser + NS34A (lane 9), tNS5T P2′ Ser (lane 11), tNS5T P2′ Ser + NS34A (lane 12), tNS5T LITCY (lane 13), tNS5T LITCY + NS34A (lane 14). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immunepreciptated with the SV5-P-k antibody before analysis by SDS-PAGE. Most of the substrates containing mutations derived from the LITS potential cleavage site were processed at low levels when these cleavage sites were incorporated into the NS5A/NS5B cleavage site (lanes 8, 9, 12, 14) except for the LITS cleavage site (lane 6) that was not processed. Molecular weight markers are indicated on the right.

Cleavage Site	Relative % Cleavage Efficiency
Wild Type	100
LITS	0
P2 Cys - Ser	45 - 50
LITC	20 - 25
LITCY	<10
P2' Met - Ser	60

Table 6.3. Cleavage of substrates containing cleavage sites derived from the potential LITS cleavage site. The tNS5T LITC substrate was processed at 25% the level of wild type efficiency. The tNS5T substrates that contained a serine residue at either P2 or P2' of the NS5A/NS5B cleavage site were processed with an efficiency approximately 50% of wild type for both substrates. The tNS5T LITCY substrate was cleaved at a low level.

As the tNS5T P2S and tNS5T LITC substrates were cleaved, the serine residue at P2 of the LITS cleavage site was not entirely reponsible for the abolition of cleavage. The remaining three amino acids within the LITS sequence are all present individually in other cleavage sites. As cleavage sites containing these amino acids alone are cleaved the presence of more than one of these amino acids (i.e. the presence of a pair or a combination of three or four amino acids) could influence the cleavability of a site. Three further tNS5T substrates were constructed that contained NS5A/NS5B cleavage sites derived from the potential LITS cleavage site. These three substrates were constructed by replacing first the P5 and P4 amino acids (tNS5T DVTS), then P4 and P3 (tNS5T LVVS) and finally P3 and P2 of the LITS sequence (tNS5T LIVC) as illustrated below.

Wild Type	SEAGT E**DVVC**C/SMSY
LITS	SEAGT ELITSC/SMSY
DVTS	SEAGT E**DV**TSC/SMSY
LVVS	SEAGT EL**VV**SC/SMSY
LIVC	SEAGT ELI**VC**C/SMSY

These new substrate constructs were cloned and sequenced before being studied in the standard *trans*-cleavage assay (Figs. 6.6., 6.7.). All three substrates that contained these potential cleavage sites were cleaved, albeit at low levels. When the tNS5T LITS substrate contained the wild type amino acids at P5 and P4 (tNS5T DVTS), this substrate was processed poorly. The tNS5T LVVS substrate was processed with a similar efficiency to the tNS5T DVTS substrate whereas the tNS5T LIVC substrate was processed with a

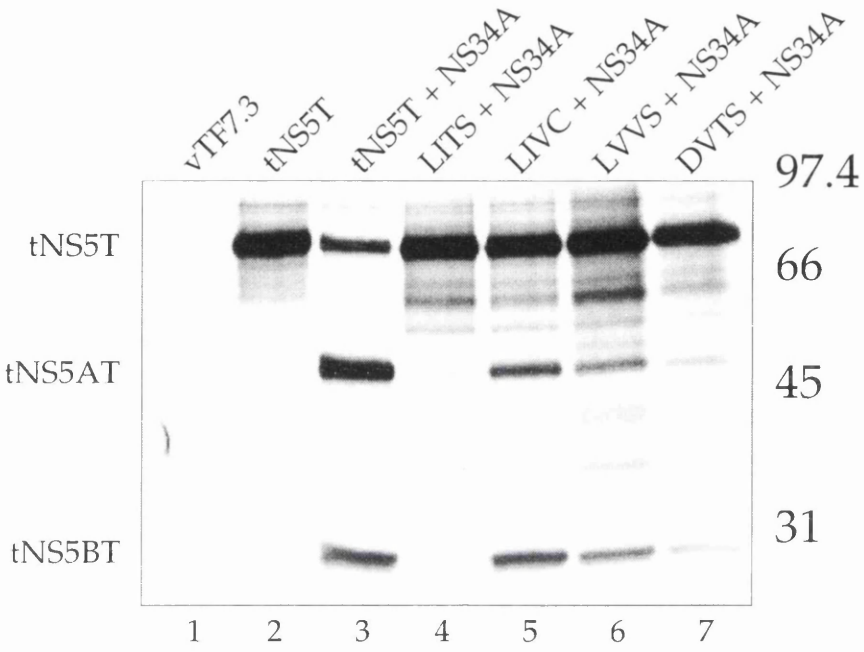


Fig. 6.6. Mutagenesis of the potential LITS cleavage site within the tNS5T substrate. vTF7.3 infected CV-1 cells were transfected with tNS5T (lane 2), tNS5T + NS34A (lane 3), tNS5T LITS + NS34A (lane 4), tNS5T LIVC + NS34A (lane 5), tNS5T LVVS + NS34A (lane 6), tNS5T DVTS + NS34A (lane 7) or were left untransfected (lane 1). Fifteen hours post-transfection the cells were labelled with ^{35}S methionine for 2 hours before lysis, immuneprecipitaion with the SV5-P-k antibody and then analysis by SDS-PAGE. The most efficiently processed of these mutated substrates contained LIVC from P5 to P2 (lane 5), followed by LVVS (lane 6) and the least efficiently processed substrate contained the sequence DVTS at P5-P2 (lane 7). The molecular weight markers are located at the right of the gel.

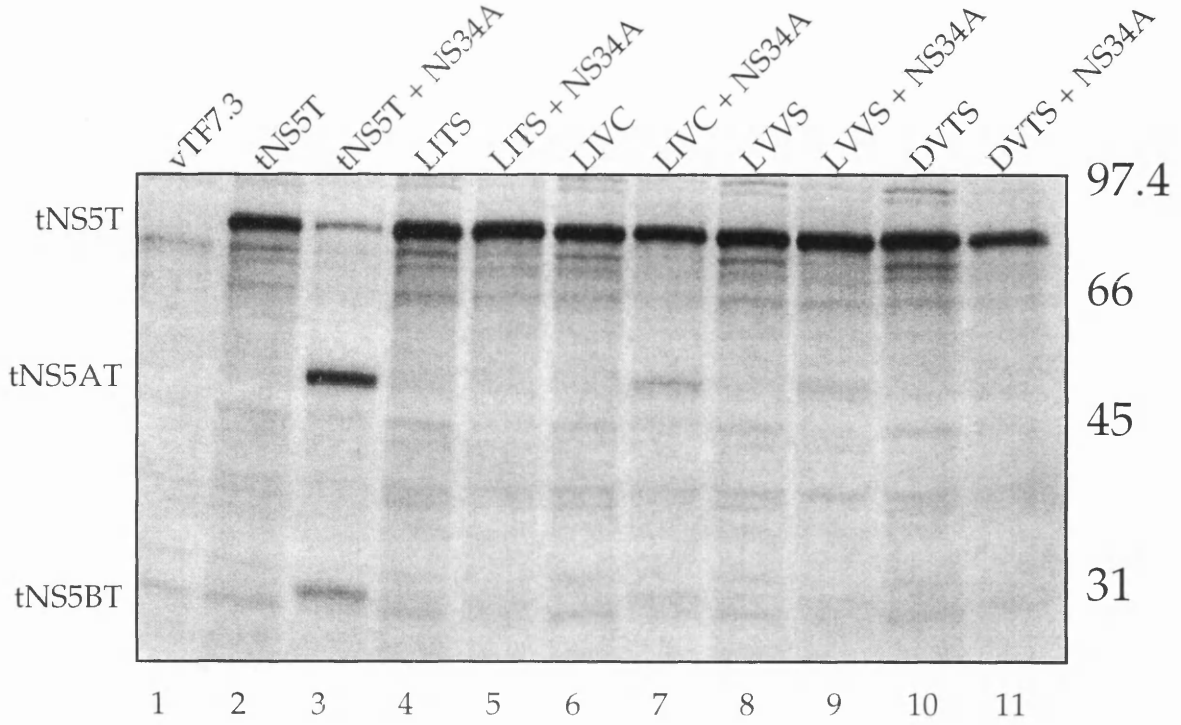


Fig. 6.7. Repeat analysis of further mutagenesis of the theoretical LITS cleavage site within the tNS5T substrate. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with tNS5T (lane 2), tNS5T + NS34A (lane 3), tNS5T LITS (lane 4), tNS5T LITS + NS34A (lane 5), tNS5T LIVC (lane 6), tNS5T LIVC + NS34A (lane 7), tNS5T LVVS (lane 8), tNS5T LVVS + NS34A (lane 9), tNS5T DVTS (lane 10), tNS5T DVTS + NS34A (lane 11). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immunoprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The mutations introduced into the LITS region within the altered NS5A/NS5B cleavage site all restored cleavage to a certain extent (lanes 7, 9, 11) although this processing took place at a reduced level in comparison with the wild type NS5A/NS5B cleavage site. Molecular weight markers are indicated on the right.

higher efficiency than the other two substrates (tNS5T DVTS and tNS5T LVVS). The efficiency with which these substrates were processed by the NS3 serine protease is estimated in Table 6.4.

Cleavage Site	Relative % Cleavage Efficiency
Wild Type	100
LITS	0
DVTS	~15
LVVS	10 - 20
LIVC	25 - 35

Table 6.4. The efficiencies with which tNS5T substrates containing derivatives of the potential LITS cleavage site were processed. The introduction of the wild type aspartic acid and valine residues at P5 and P4 to produce tNS5T DVTS created a substrate that was processed at 15% of the wild type tNS5T. The effect of replacing the P4 and P3 amino acids with the wild type valines produced a substrate (tNS5T LVVS) processed marginally higher than the previous substrate. When valine and cysteine were introduced at P3 and P2 to give tNS5T LIVC the largest increase in processing efficiency was observed. The substrate was processed with a slightly higher efficiency compared to the tNS5T LITC substrate. All three substrates were processed at a significantly lower level than the wild type substrate, tNS5T.

A cleavage site was constructed with only the mutant leucine residue at P5 of the cleavage site maintained from the original potential cleavage site (LITS). This cleavage site was cloned into the tNS5T cDNA and the new mutant substrate (tNS5T LVVC) introduced into the *trans*-cleavage assay. This substrate and all the previous substrates were cotransfected with NS34A into vTF7.3 infected CV-1 cells and the efficiencies with which these substrates were processed compared. After radiolabelling with ^{35}S-methionine the cells were lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. All the substrates constructed to investigate the potential LITS cleavage site were included to give a full picture of the analysis (Fig. 6.8.). The tNS5T LVVC substrate containing the wild type amino acids from P4 to P2 was processed at a significantly lower level of efficiency in comparison to tNS5T. The relative percentage of cleavage efficiency by which these substrates were cleaved is summarised in Table 6.5.

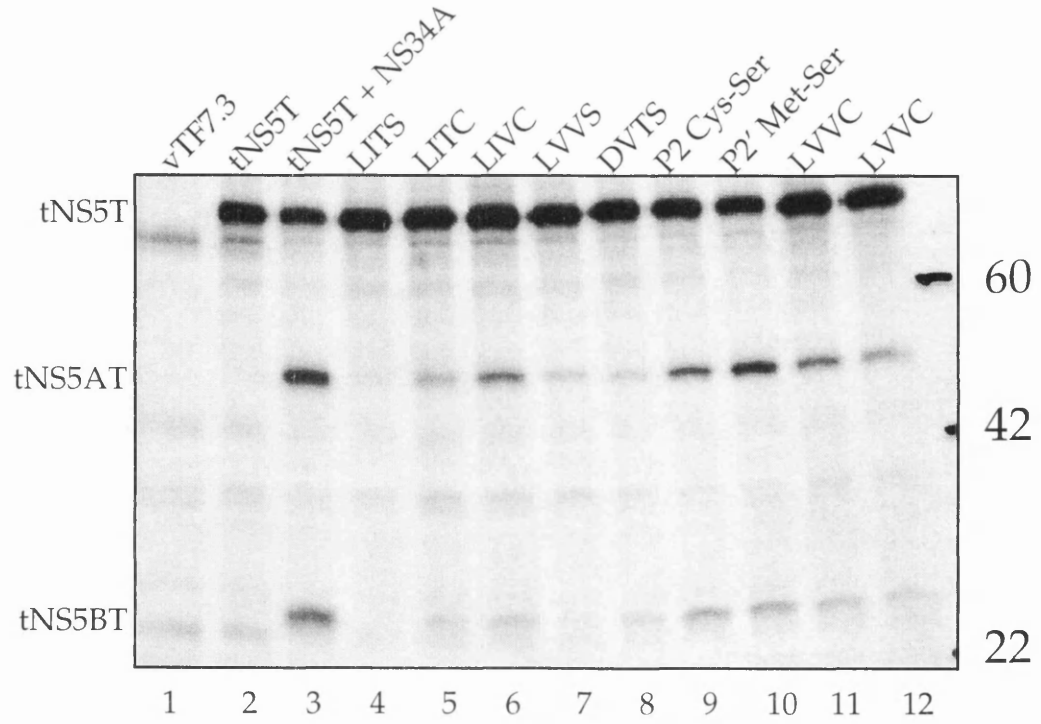


Fig. 6.8. Mutational analysis with respect to the potential LITS cleavage site. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with tNS5T alone (lane 2) or substrate + NS34A. The substrates cotransfected with NS34A were tNS5T (lane3), tNS5T LITS (lane 4), tNS5T LITC (lane 5), tNS5T LIVC (lane 6), tNS5T LVVS (lane 7), tNS5T DVTS (lane 8), tNS5T P2 Ser (lane 9), tNS5T P2' Ser (lane 10), tNS5T LVVC (lanes 11 and 12). Fifteen hours post-transfection the cells were labelled with ^{35}S methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. All the LITS derived mutations were introduced into the NS5A/NS5B cleavage site in tNS5T. The results illustrated above demonstrate the full analysis carried out on the potential LITS cleavage site. With the exception of the two serine substitutions in the wild type substrate the most efficiently processed cleavage site derivative was the LIVC cleavage site. Molecular weight markers are indicated on the right.

Cleavage Site	Relative % Cleavage Efficiency
Wild Type	100
LITS	0
LITC	20 - 25
LIVC	25 - 35
LVVS	10 - 20
DVTS	~15
LVVC	15 - 25
P2 Cys - Ser	45 - 50
P2' Met - Ser	60

Table 6.5. Relative % cleavage efficiencies of tNS5T substrates with their NS5A/NS5B cleavage sites altered to investigate the potential LITS cleavage site. No significant increase in efficiency was observed between the LIVC substrate and the LVVC mutant substrate as expected with no significant structural change between isoleucine and valine.

6.4. Analysis of different cleavage sites within the assay system.

The NS3 serine protease is responsible for cleavage at three other sites (NS3/NS4A, NS4A/NS4B, NS4B/NS5A) as well as cleavage at NS5A/NS5B. Analysis of processing at these sites would provide a more complete insight into the substrate specificity for cleavage by NS3. Antisera against the NS4 region of the HCV Gla polyprotein was not available in the laboratory so development of suitable NS4A/NS4B and NS4B/NS5A substrates would require the addition of the SV5 epitope tag to the termini of these proteins. The NS4A protein is also small and difficult to detect by SDS-PAGE. The NS34A protein undergoes self-processing *in vitro*. It was decided to introduce the three unstudied cleavage sites (NS3/NS4A, NS4A/NS4B, NS4B/NS5A) from P6-P1' into the tNS5T substrate in place of the NS5A/NS5B cleavage site.

Oligonucleotides were designed, as for the previous mutagenesis of the NS5A/NS5B cleavage site, encoding the NS3/NS4A, NS4A/NS4B and NS4B/NS5A cleavage sites so that these sequences could be introduced into the tNS5T substrate. These substrates would allow a comparison between the different cleavage sites and the relative efficiency of processing by the NS3 serine protease. All three cleavage sites were cloned and sequenced to create three new substrates: tNS5T 34A, tNS5T 4A4B, tNS5T 4B5A. Due to an error during oligonucleotide synthesis a substrate containing an additional cleavage site was obtained encoding the NS4B/NS5A cleavage site with a glutamic acid residue at P6 instead of an aspartic acid (tNS5T 4B5A P6E).

This additional substrate was included in the experiment and all four substrates were cotransfected with NS34A into vTF7.3 infected CV-1 cells. After radiolabelling with ^{35}S-methionine, the cell lysates were immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The experiments were repeated to ensure there were no discrepancies in the results obtained (Figs. 6.9., 6.10.).

When the tNS5T 34A substrate was coexpressed with NS34A no cleavage products were observed. The tNS5T 4A4B substrate was processed less efficiently than tNS5T and tNS5T 4B5A was cleaved at a lower efficiency than tNS5T 4A4B. The tNS5T 4B5A substrate with its P6 residue substituted with a glutamic acid was cleaved with a similar efficiency to the tNS5T 4B5A substrate.

The relative processing efficiencies obtained are illustrated in Table 6.6.

Cleavage Site	Relative % Cleavage Efficiency
NS5A/NS5B	100
NS3/NS4A	0
NS4A/NS4B	45 - 50
NS4B/NS5A	20 - 25
NS4B/NS5A (P6 E)	20 - 30

Table 6.6. Relative cleavage efficiencies of tNS5T substrates that contained the four cleavage sites processed within the HCV polyprotein. The altered NS4B/NS5A cleavage site was also encoded within the tNS5T substrate. The NS3/NS4A cleavage site was not processed *in trans* when it was substituted for the NS5A/NS5B cleavage site within tNS5T. The cleavage sites that were processed *in trans* within the HCV polyprotein were all cleaved when present within the tNS5T substrate, albeit with varying efficiencies. The NS4A/NS4B cleavage site was processed at an efficiency 50% of that seen for the NS5A/NS5B cleavage site. Both the NS4B/NS5A and the NS4B/NS5A (P6 E) were cleaved by the NS3 serine protease at approximately 25% of the efficiency with which the NS5A/NS5B cleavage site was processed.

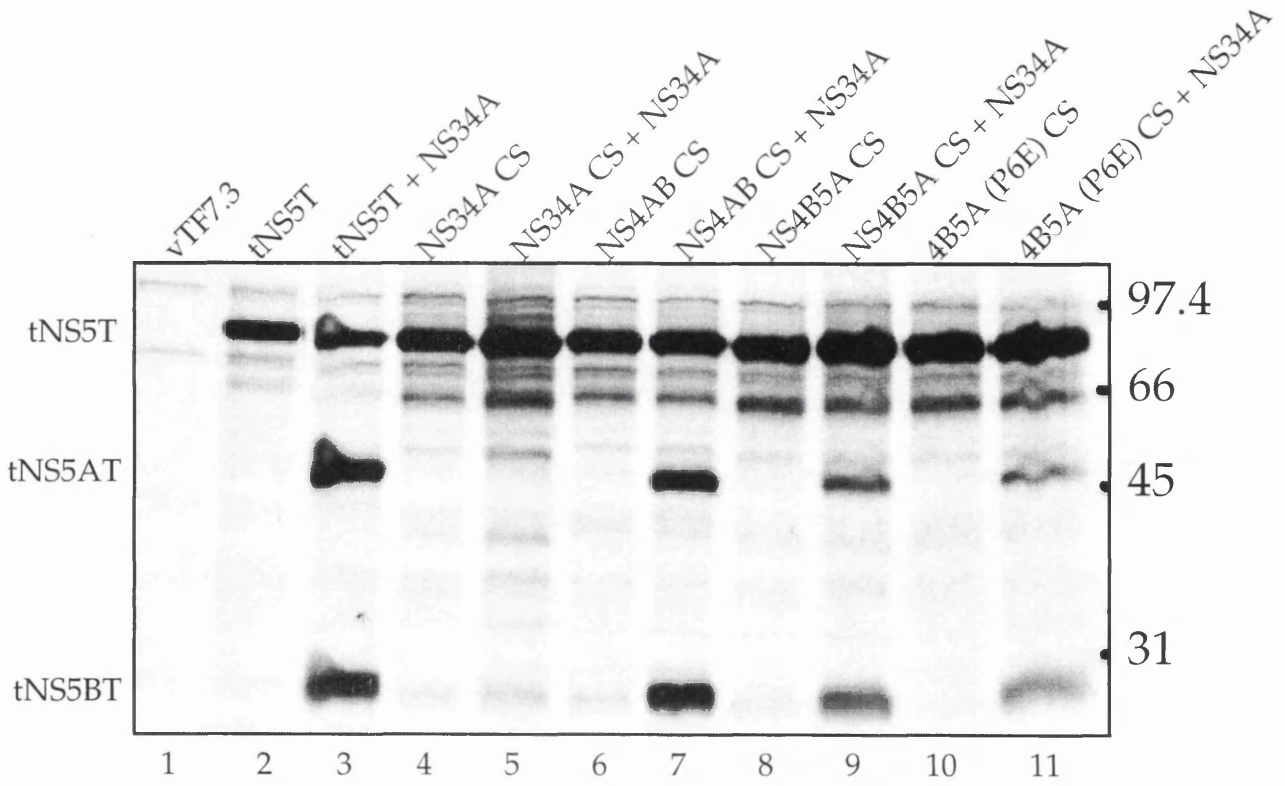


Fig. 6.9. Comparison of different HCV cleavage sites within the tNS5T substrate. All three cleavage sites (CS) cleaved within the HCV polyprotein were encoded within the tNS5T substrate in place of the NS5A/NS5B cleavage site. In addition to these sites the NS4B/NS5A cleavage site containing a glutamic acid at P6 was also encoded in place of the NS5A/NS5B cleavage site within tNS5T. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with tNS5T (lane 2), tNS5T + NS34A (lane 3), NS34A CS (lane 4), NS34A CS + NS34A (lane 5), NS4A4B CS (lane 6), NS4A4B CS + NS34A (lane 7), NS4B5S CS (lane 8), NS4B5A CS + NS34A (lane 9), NS4B5A (P6E) CS (lane 10), NS4B5A (P6E) CS + NS34A (lane 11). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. All the cleavage sites normally processed in the polyprotein by the NS3 serine protease *in trans* were cleaved (lanes 7, 9, 11), albeit at different efficiencies, when they replaced the NS5A/NS5B cleaveage site. The NS3/NS4A cleavage site, normally processed *in cis*, was not cleaved (lane 5). Molecular weight markers are illustrated on the right.

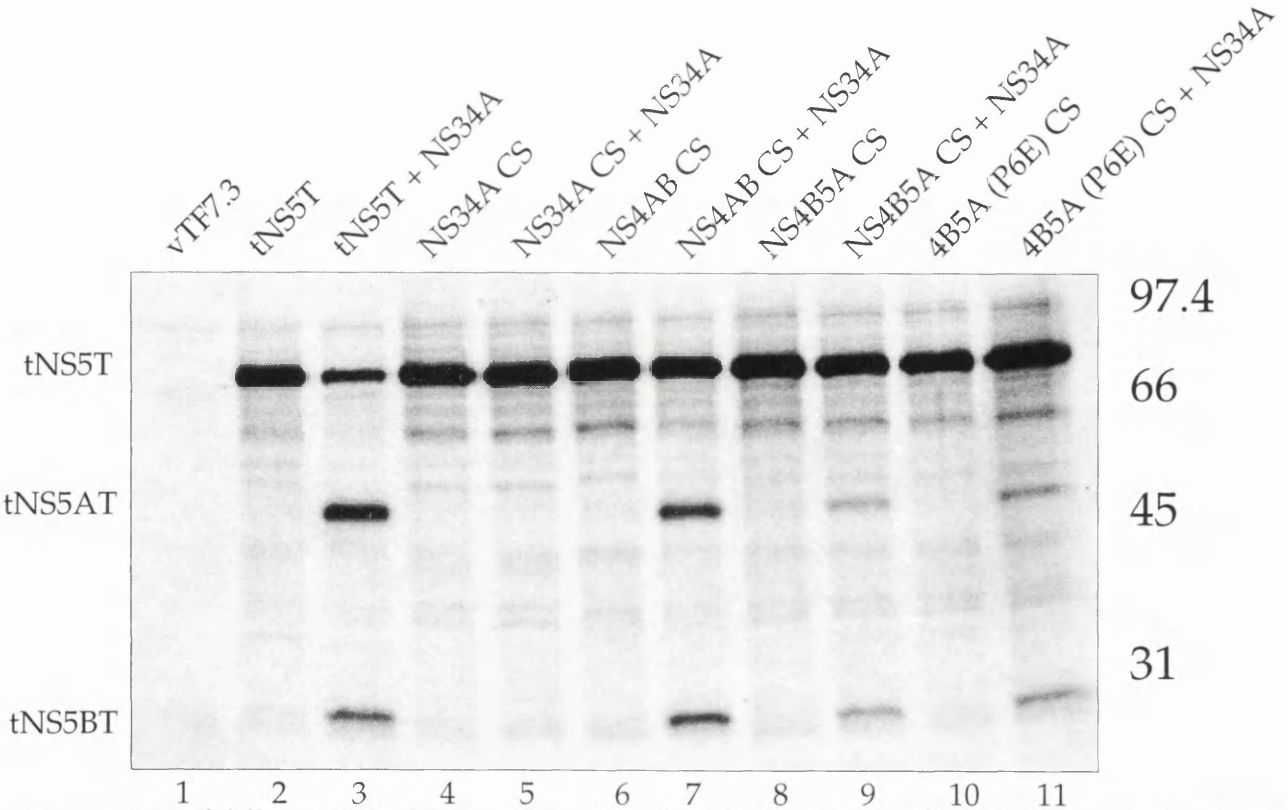


Fig. 6.10. Repeat comparison of different HCV cleavage sites within the tNS5T substrate. All three cleavage sites (CS) cleaved within the HCV polyprotein were encoded within the tNS5T substrate in place of the NS5A/NS5B cleavage site. In addition to these sites the NS4B/NS5A cleavage site containing a glutamic acid at P6 was also encoded in place of the NS5A/NS5B cleavage site within tNS5T. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with tNS5T (lane 2), tNS5T + NS34A (lane 3), NS34A CS (lane 4), NS34A CS + NS34A (lane 5), NS4A4B CS (lane 6), NS4A4B CS + NS34A (lane 7), NS4B5S CS (lane 8), NS4B5A CS + NS34A (lane 9), NS4B5A (P6E) CS (lane 10), NS4B5A (P6E) CS + NS34A (lane 11). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immunoprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. All the cleavage sites normally processed in the polyprotein by the NS3 serine protease *in trans* were cleaved (lanes 7, 9, 11), albeit at different efficiencies, when they replaced the NS5A/NS5B cleaveage site. The NS3/NS4A cleavage site, normally processed *in cis*, was not cleaved (lane 5). Molecular weight markers are illustrated on the right.

Chapter 7 - Analysis of the NS3 serine protease

7.1. Influence of the N-terminus of NS3 on protease activity.

The standard protease assay required the cotransfection of a suitable substrate construct with the NS34A protease construct. To further determine the protease requirement for this assay alternative protease constructs were tested. Dr M.J. McElwee provided a construct encoding the NS234A region of the HCV Gla polyprotein (amino acid 737 to 1709) within pTZ18. The polyprotein encoded begins 73 amino acids upstream of the NS2 N-terminus and encompasses the p7 protein, NS2, NS3 and all but the last 2 amino acids of NS4A. This construct contained a point mutation within the NS2 region that inactivated the NS2/3 protease. When the NS234Mut polyprotein was expressed by coupled *in vitro* transcription/translation it produced an NS3 protein with the NS2 protein still attached at the N-terminus (Dr M. J. McElwee, unpublished data). The presence of the p7 protein had no effect on NS2/3 protease activity as a similar construct was cloned, containing a single amino acid change in NS2, that produced an active NS2/3 protease. The NS234Mut protein expressed *in vitro* was cleaved to release the NS4A protein whereas the NS2/NS3 cleavage event did not take place. The effect of having the NS2 protein still attached to the N-terminus of NS3 on the *trans*-cleavage activity of NS3 was not known.

NS5 and tNS5T were cotransfected with NS34A (as a positive control) or NS234Mut into vTF7.3 infected CV-1 cells. After radiolabelling with ^{35}S-methionine the cells were lysed, immuneprecipitated with the relevant antisera and analysed by SDS-PAGE (Fig. 7.1.). Both substrates were processed when coexpressed with NS234Mut showing that additional sequence at the NS3 N-terminus did not inactivate the serine protease activity of NS3. Coexpression of the substrate with NS234Mut did not result in the same level of processing as coexpression with NS34A. As 5μg of each protease construct was used in the cotransfections and the NS234Mut construct has a larger insert, the relative amount of NS3 produced would be lower for the NS234Mut construct. However no titration was carried out to determine whether NS3 with NS2 still attached was as proteolytically active as NS3 alone.

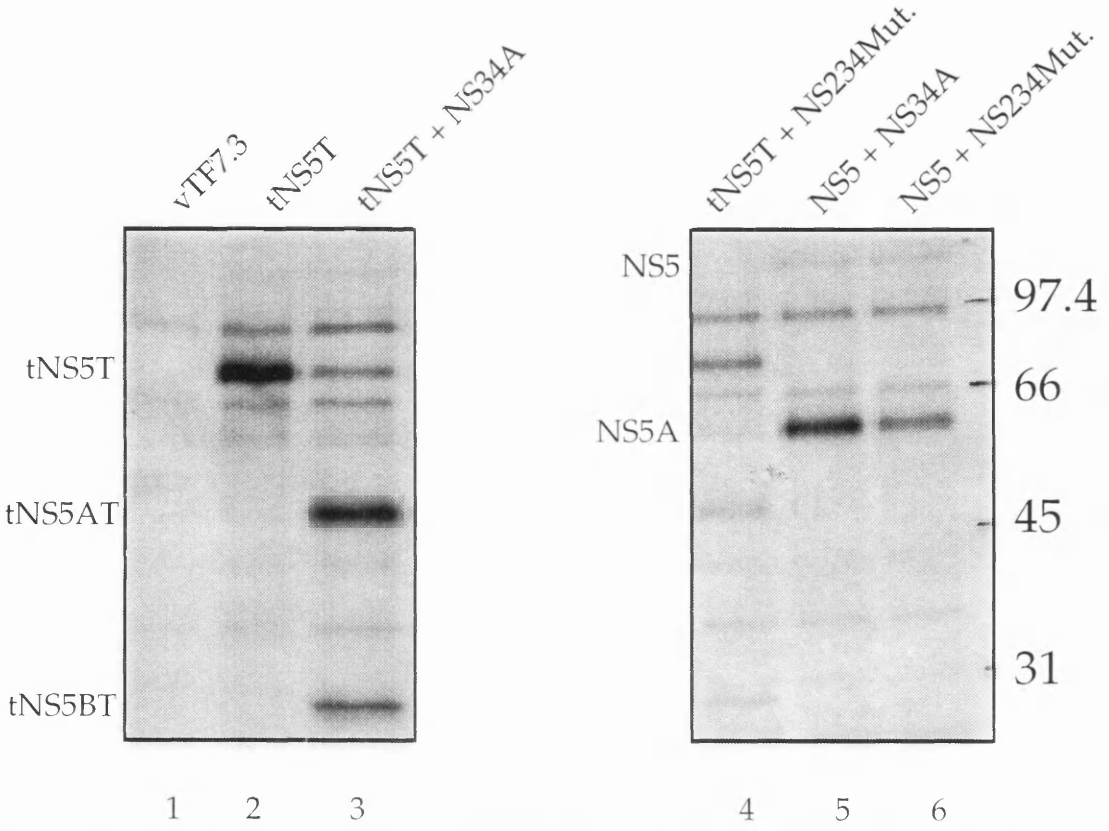


Fig. 7.1. Cotransfection of both NS5 substrates with different protease constructs. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or transfected with HCV cDNAs. The tNS5T substrate was transfected alone (lane 2) and both substrates (tNS5T and NS5) were cotransfected with either the NS34A construct (lanes 3 and 5) or the NS234Mut construct (lanes 4 and 6 respectively). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the relevant antisera before analysis by SDS-PAGE. Both the NS34A construct and the NS234Mut construct produced an active serine protease that processed both the NS5 and the tNS5T substrate. The NS234Mut construct produced an NS23 protein that retained its NS3 serine protease activity despite the presence of uncleaved NS2 at the N-terminus of NS3. Molecular weight markers are indicated on the right.

7.2. Ability of different protease constructs to process *in trans.*

The full length NS3 and the protease domain (NS3P) cDNAs were cloned within pTF7.5 and pTM1 respectively. These protease constructs, as well as the NS4A cofactor gene cloned within pTM1, could be used in the *trans*-cleavage assay to determine the protease requirements for cleavage *in trans*. Both NS3 and NS3P were cotransfected with tNS5T in the presence or absence of NS4A into vTF7.3 infected CV-1 cells. After radiolabelling with ^{35}S-methionine the immuneprecipitated cell lysates were analysed by SDS-PAGE (Fig. 7.2.). Only cotransfection of tNS5T with NS34A resulted in the production of cleavage products. Even when the full length NS3 construct was cotransfected with NS4A and tNS5T, no cleavage products were detected.

To determine whether the use of different cell lines could influence the requirements for processing the cotransfections were carried out using tNS5T as a substrate and the different combinations of protease constructs in a liver-derived cell line, HuH7. In addition to using an alternative cell line both transfection methods (cationic liposomes and PEI) were used (Fig. 7.3.). As for the analysis carried out in CV-1 cells, only cotransfection of the substrate with NS34A resulted in the *trans*-cleavage. The result of the cotransfections were the same whether cationic liposomes or PEI were used to introduce the DNA into the cells.

An NS3 construct was supplied by Dr B. E. Clarke (GlaxoWellcome) which had been shown to process an NS5 substrate in the absence of the NS4A cofactor. This construct encoded the full length NS3 gene (derived from the type 1b HCV-A genome) in the vector pTM3. This plasmid, like pTM1, contained the EMCV IRES element that directs cap-independent translation. Both the HCV Gla NS34A construct and the GlaxoWellcome NS3 construct were cotransfected with tNS5T into vTF7.3 infected CV-1 cells to determine their catalytic activity *in trans*. After the cells were labelled with ^{35}S-methionine the cells were lysed, immuneprecipitated with the SV5-P-k antisera and analysed by SDS-PAGE (Fig. 7.4.).

Both proteases processed the tNS5T substrate. A comparison of processing efficiencies could have been carried out between the HCV Gla NS34A and the HCV-A NS3 but this comparison would not have been informative as the HCV Gla NS34A was expressed from pTF7.5, a vector containing the T7 promoter only. The HCV-A NS3 construct was expressed from pTM3, a vector that contains the EMCV IRES element that boosts protein expression. The pTM3 HCV-A NS3 construct would produce significantly higher levels of protease than the pTF7.5 HCV Gla NS34A construct. The difference in the level of processing between protease constructs would be influenced by the increased ratio of protease to substrate molecules for the HCV-A NS3. For a direct comparison of the two proteases the genes would have to be cloned in the same plasmid.

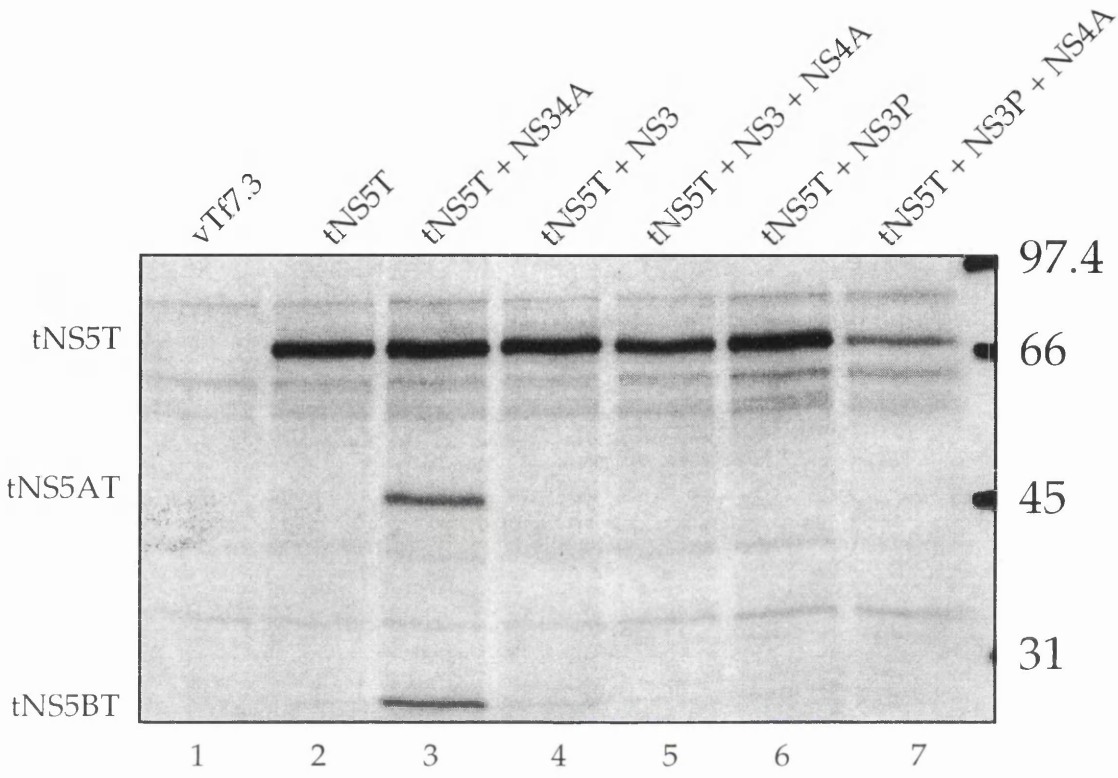


Fig. 7.2. Analysis of cleavage of tNS5T when coexpressed with different combinations of proteases. CV-1 cells infected with vTF7.3 were not transfected with DNA (lane 1) or were transfected with tNS5T alone (lane 1), tNS5T + NS34A (lane 3), tNS5T + NS3 (lane 4), tNS5T + NS3 + NS4A (lane 5), tNS5T + NS3P (lane 6), tNS5T + NS3P + NS4A (lane 7). Fifteen hours post-transfection the cells were radiolabelled with ^{35}S-methionine for 2 hours, lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. One microgram of the tNS5T substrate was cotransfected with 5µg each of the different proteaseand cofactor constructs. Processing of the substrate was only observed when the substrate was cotransfected with NS34A (lane 3). The molecular weight markers are located at the right of the gel.

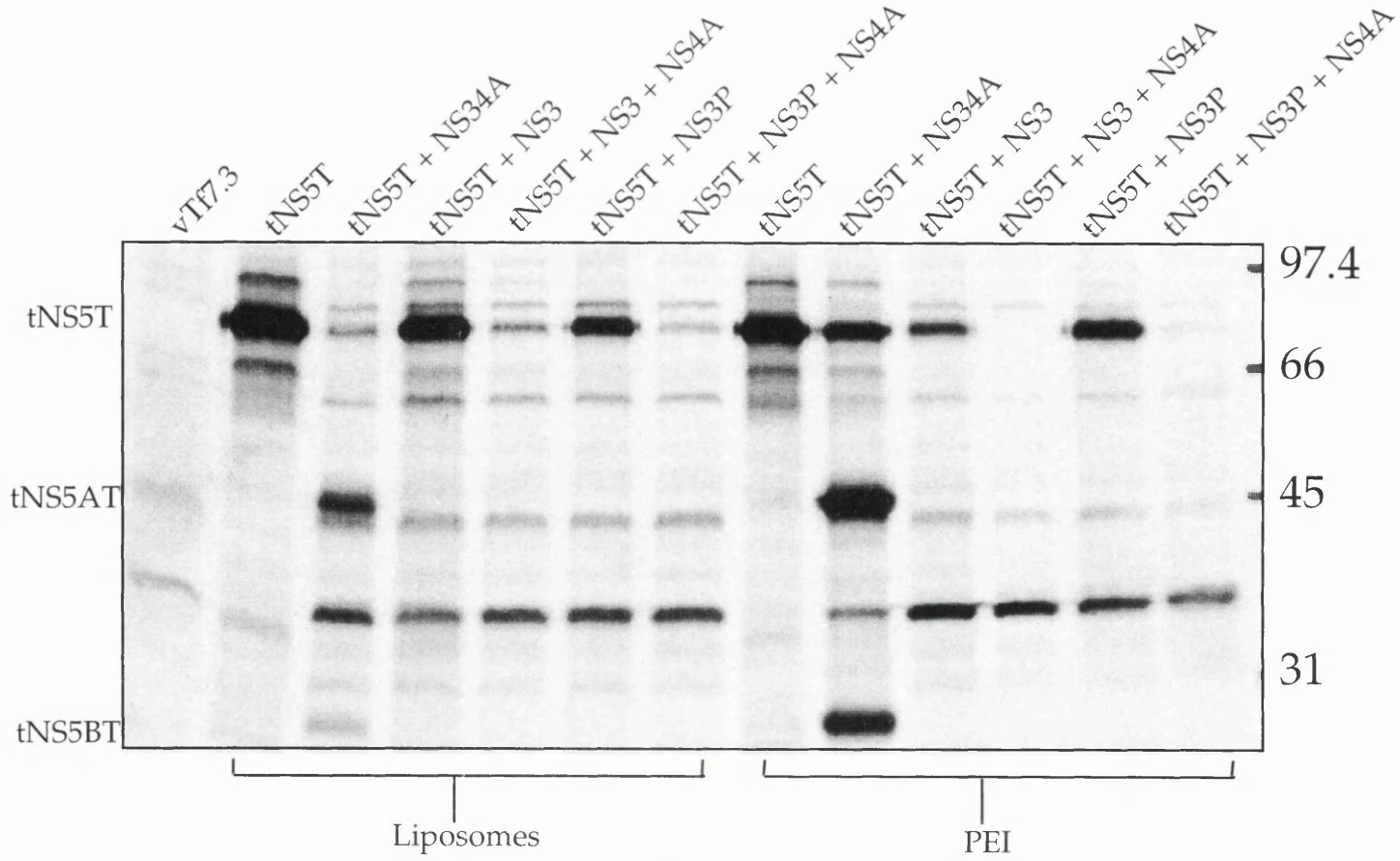


Fig. 7.3. Cotransfection of substrate and different protease constructs in HuH 7 cells. The tNS5T substrate was co-transfected with different protease constructs into vTF7.3 infected HuH 7 cells. The tNS5T substrate was cotransfected with NS34A, NS3, NS3 + NS4A, NS3P and NS3P + NS4A. Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The tNS5T substrate was only processed when it was coexpressed with NS34A. The transfections were carried out using both liposomes and PEI to compare transfection efficiencies in a different cell line. As for cotransfection in CV-1 cells there was no difference in processing efficiency when either liposomes or PEI were used as a transfection medium. Molecular weight markers are illustrated on the right.

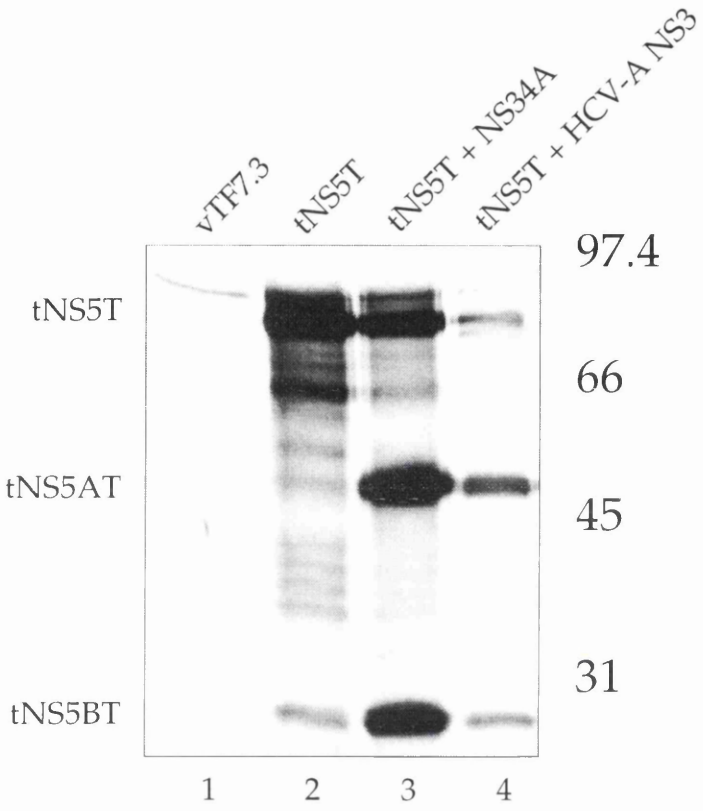


Fig. 7.4. Comparison of processing of the tNS5T substrate by HCV Gla NS34A and HCV-A NS3 (supplied by GlaxoWellcome). CV-1 cells infected with vTF7.3 were left untransfected (lane 1) or were transfected with tNS5T alone (lane 2), tNS5T + HCV Gla NS34A (lane 3) or tNS5T + HCV-A NS3 (lane 4). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine, lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. Both proteases processed the tNS5T substrate (lanes 3 and 4). The ability of the HCV-A NS3 protease to process the substrate in the absence of NS4A is in contrast to the results obtained with the HCV Gla NS3. The molecular weight markers are located at the right of the gel.

The amino acid sequence of the two NS3 genes were compared. The ability of the GlaxoWellcome NS3 alone to process the substrate and not the HCV Gla NS3 could be due to a difference in the amino acid sequence between the two proteases. An amino acid sequence comparison between the protease domains of the two NS3 constructs was carried out (Fig. 7.5.).

Many of the amino acid changes were either conservative (ie one small, non-polar residue for another) and/or observed in other HCV isolates, i.e. for a number of the amino acid differences observed, the amino acid was present in different HCV isolates. In the HCV Gla amino acid sequence there was a positively charged lysine residue next to the aspartic acid of the catalytic triad. In the HCV-A sequence this residue was a glutamine. In other HCV isolates this amino acid was predominantly a glutamine residue although glycine residues and leucine residues were found in each of two isolates. The lysine residue could have an inhibitory effect on the NS3 serine protease activity that was only overcome in the presence of NS4A. For this reason an NS3 construct containing this lysine residue substituted with a glutamine residue was constructed.

Dr. M.J. McElwee provided an NS34A construct with this lysine residue (at amino acid 1106) substituted with the glutamine residue observed for the majority of other HCV isolates . This construct had been designated NS34A[K-Q]. As for the wild type NS34A this protease, when expressed in rabbit reticulocyte lysates, underwent cleavage *in cis* to produce NS3 from the NS34A complex. The NS3 gene contained a *Sal* I site approximately 130bp upstream of the codon encoding the relevant lysine residue and a *Sac* I site approximately 1270bp downstream. Digestion of the NS34A[K-Q] DNA with these restriction enzymes produced a 1400bp insert that contained the lysine to glutamine substitution encoded within the nucleotide sequence. This insert was ligated into *Sac* I/*Sal* I digested pTF7.5 NS3 to create pTF7.5 NS3[K-Q]. The ability of this new protease to process the substrate was determined as well as the wild type protease construct and the NS34A[K-Q] construct. All these constructs were cotransfected with tNS5T into vTF7.3 infected CV-1 cells. The cells were radiolabelled with ^{35}S-methionine, lysed, immuneprecipitated with the SV5-P-k antibody and analysed by SDS-PAGE (Figs. 7.6. and 7.7.).

The single mutation from the lysine residue to the glutamine residue (adjacent to the aspartic acid of the catalytic triad) enabled the HCV Gla NS3 to process tNS5T in the absence of NS4A. NS3[K-Q] processed tNS5T with an efficiency approximately 50% of that observed for wild type NS34A.

The HCV-A NS3 contained a glycine residue 11 amino acids downstream of the aspartic acid of the catalytic triad (Fig. 7.5.). The arginine residue at this position within the HCV Gla NS3 is observed for a number of other HCV genotypes. The HCV-A NS3 retained the ability to process despite the loss of a highly conserved, positively charged residue at this position (twelve isolates have an arginine at this position, three have a lysine). The effect of losing this positive charge was investigated to determine if there was an effect on the processing activity of the NS3 serine protease.

```
1026 APITAYAQQTRGLLGCIITSLTGRDKNQVEGEVQIVSTA
1026 APITAYSQQTRGLLGCIITSLTGRDKNQVEGEVQVVSTA

     AQTFLATCINGVCWTVYHGAGSRTISGPKGPVIQMYTNV
     TQSFLATCVNGVCWTVYHGAGSKTLAGPKGPVIQMYTNV

     DKDLVGWPAPQGARSLTPCTCGSSDLYLVTRHADVSIPV
     DQDLVGWPAPPGAGSLTPCTCGSSDLYLVTRHADVSIPV

     RRRGDSRGSLLSPRPISYLKGSSGGPLLCPAGHAVGIFR
     RRRGDSRGSLLSPRPISYLKGSSGGPLLCPSGHAVGIFR

     AAVCTRGVAKAVDFIPVENLETTMRSPVFTDNSSP 1216
     AAVCTRGVAKAVDFIPVESMETTVRSPVFTDNSSP 1216
```

Fig. 7.5. Sequence comparison between the protease domains of the HCV-A and HCV Gla amino acid sequences. The amino acids of the catalytic triad are illustrated in red. The amino acids differing between the two sequences are shown in green with the two "significant" differences (K-Q, R-G) shown in blue.

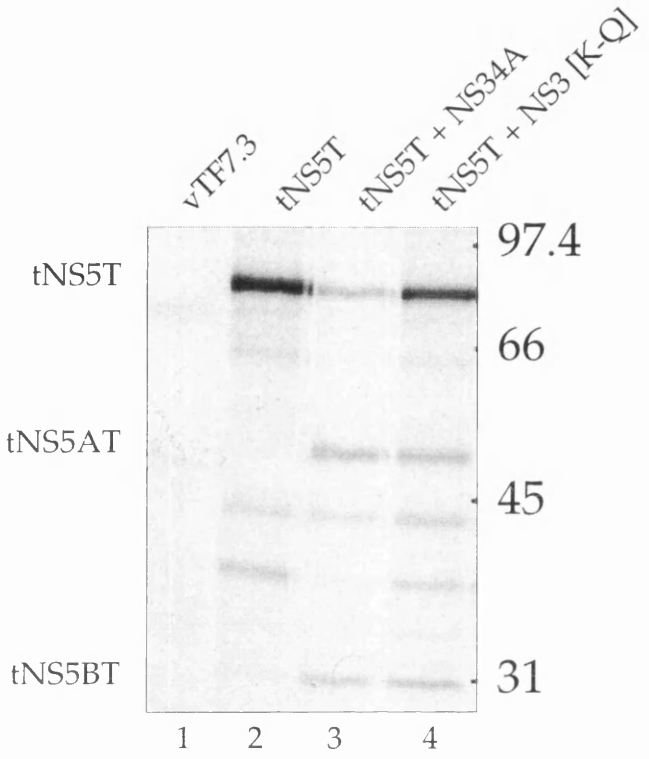


Fig. 7.6. Analysis of processing by HCV Gla NS3[K-Q]. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with tNS5T (lane 2), tNS5T + NS34A (lane 3) or tNS5T + NS3[K-Q] (lane 4). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine, lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. Cotransfection of the tNS5T substrate with the NS34A HCV Gla construct resulted in cleavage of the substrate (lane 3). When NS3 alone was cotransfected with the tNS5T substrate no cleavage took place. The substitution of a lysine residue by a glutamine residue adjacent to the aspartic acid of the catalytic triad enabled NS3 alone to process the tNS5T substrate (lane 4). The molecular weight markers are located at the right of the gel.

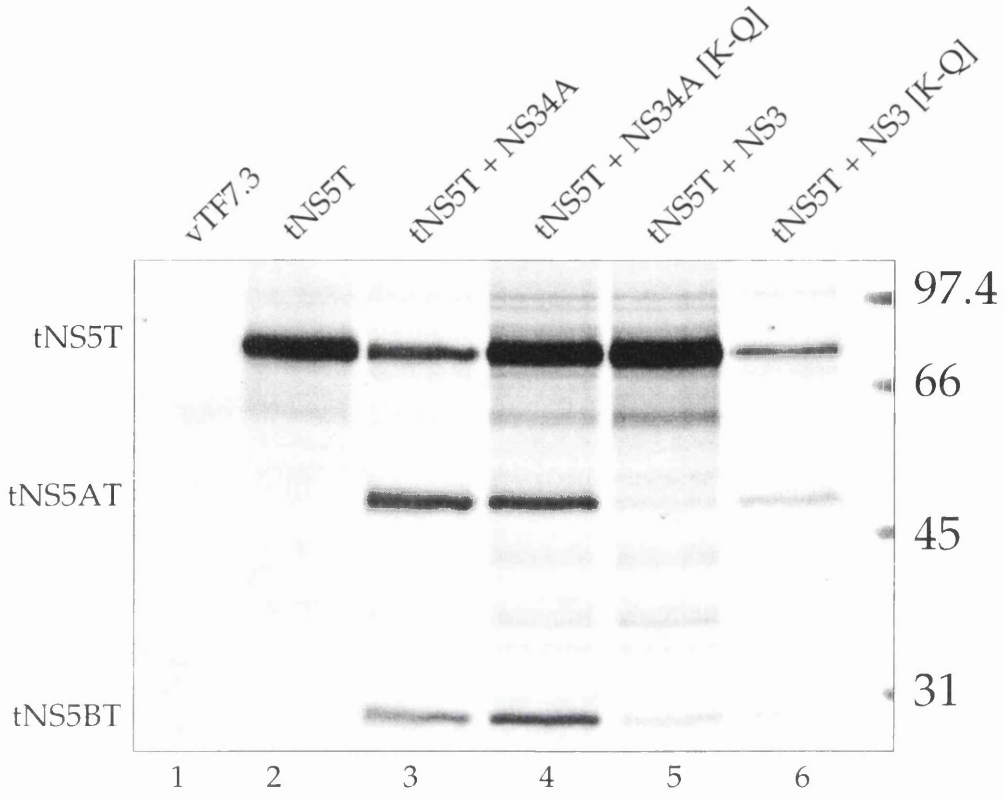


Fig. 7.7. Comparison of the processing efficiencies of proteases containing single amino acid changes. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with tNS5T alone (lane 2), tNS5T + NS34A (lane 3), tNS5T + NS34A[K-Q] (lane 4), tNS5T + NS3 (lane 5), tNS5T + NS3[K-Q] (lane 6). Fifteen hours post-transfection the cells were labelled with ^{35}S methionine, lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The only protease that did not process the tNS5T substrate was the wild type NS3 (lane 5). Although the K-Q substitution enabled NS3 to process alone (lane 6) the wild type NS34A (lane 3) was more efficient at processing the tNS5T substrate than NS34A[K-Q] (lane 4). The molecular weight markers are located at the right of the gel.

Four protease constructs were constructed with the arginine to glycine mutation introduced into both NS3 and both NS34A constructs. Two 27 nucleotide complementary oligonucleotides were designed that encompassed the region encoding the target amino acid (arginine). Both oligonucleotides contained the relevant nucleotide changes from the wild type sequence to give a glycine codon 11 amino acids downstream of the aspartic acid of the catalytic triad. Two additional oligonucleotide primers were supplied by Dr M. J. McElwee. One oligonucleotide (NS3N) corresponded to the 5′ terminus of the NS3 gene and contained a *Bam* HI site followed by an initiation codon and 21 nucleotides corresponding to the 5′ terminus of the NS3 gene. The second oligonucleotide (3′Hel) was in the complementary orientation to the first and corresponded to the 3′ terminus of the NS4A gene. This oligonucleotide also contained a *Bam* HI site and a termination codon preceding 18 nucleotides of the NS4A gene. All four primers are illustrated in Fig. 7.8.

Two sets of PCRs using *Pfu* polymerase were carried out using both the NS34A and NS34A[K-Q] plasmids as templates. The first set of PCRs were carried out using the NS3N primer and the 3′-5′ oligonucleotide encoding the R-G substitution. The second set of PCRs used the 3′Hel primer and the 5′-3′ mutant oligonucleotide (Fig. 7.9.). For both templates a 1:20 and 1:200 dilution of the plasmid stock DNA was used. The results of the first round PCR reactions are illustrated in Fig. 7.10.

Each PCR fragment was purified through a silica gel matrix and eluted into 20µl dH_2O. A second round of PCRs were carried out using the NS3N and 3′Hel primers to produce a full length NS34A gene. The purified, first round PCR products were diluted in dH_2O 1:50 and one microlitre of this was used as a template. To produce the NS34A[R-G] PCR product the PCR fragments amplified from the original wild type NS34A construct were combined and used as templates to produce the final mutant NS34A[R-G] gene (Fig. 7.9.). To produce the NS34A[K-Q, R-G] insert the first round PCR reactions amplified from the NS34[K-Q] template were combined and used as templates.

The second round PCR products were digested with *Bam* HI before ligation into *Bam* HI cut pTF7.5 to produce pTF7.5 NS34A[R-G] and pTF7.5 NS34A[K-Q, R-G]. After the nucleotide sequence of the two protease constructs was confirmed, the region around the aspartic acid of the catalytic triad was excised from both constructs with *Sal* I and *Sac* I (as for the creation of pTF7.5 NS3[K-Q]). These inserts were ligated into the NS3 constructs (digested with *Bam* HI) to produce pTF7.5 NS3[R-G] and pTF7.5 NS3[K-Q, R-G].

These new proteases were analysed for their processing ability as were the wild type NS34A construct and both K-Q protease constructs in the standard *trans*-cleavage assay (Figs 7.11. and 7.12.). Both the wild type NS34A and NS34A[R-G] protease complexes cleaved tNS5T with the highest efficiency, with little difference in their processing efficiencies. Both the NS34A complexes containing the lysine to glutamine substitution were far less efficient in comparison to the other NS34A complexes. With the exception of the wild type HCV Gla NS3, which did not cleave the substrate,

Fig. 7.8. PCR primers designed for the introduction of the arginine to glycine mutation within the NS3 gene of the HCV Gla construct. Restriction enzyme sites (*Bam* HI) are illustrated in red, HCV sequence is illustrated in blue, initiation and termination codons are illustrated in green and the nucleotide mutations are illustrated in yellow. The nucleotide positions are shown above the nucleotide sequence of the primers with the amino acids encoded shown below.

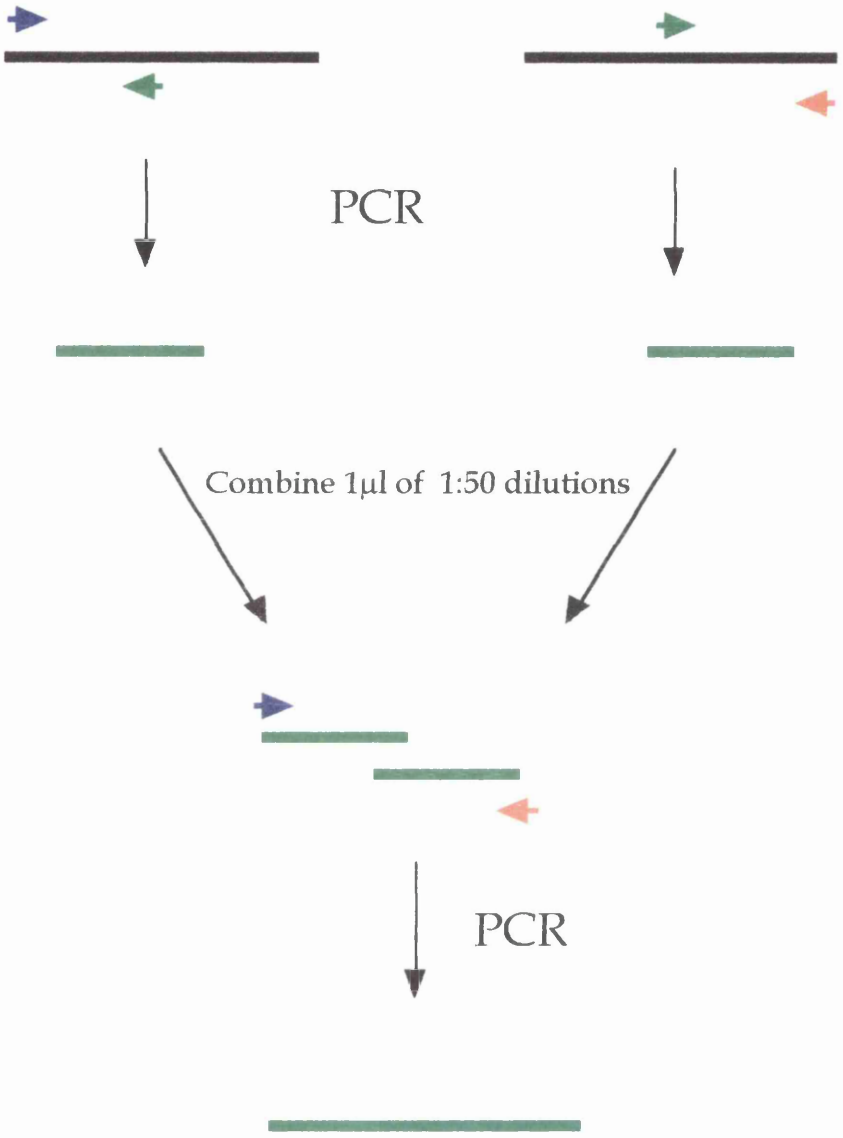


Fig. 7.9. Amino acid substitution by PCR. The central oligonucleotide primers (◄) encode the substituting amino acid codon for the NS34A gene. After the first round of PCR the products were purified, diluted and combined to act as templates for the second round of PCR. The external primers used in the first PCR reactions (►, ◄) were used to produce the full length NS34A PCR product.

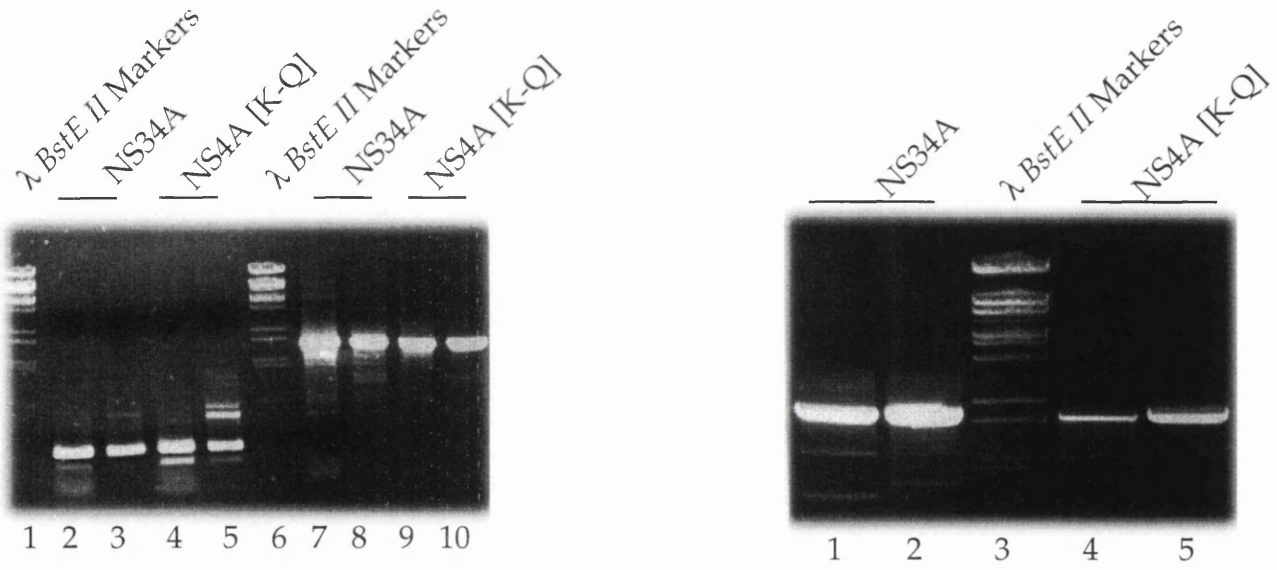


Fig. 7.10. PCR for the production of NS34A[R-G] and NS34A[K-Q, R-G] inserts. The first round PCR products (left) produced the 5′(lanes 2-5) and 3′ (lanes 7-10) cDNA fragments of the NS34A gene. After purification by silica gel matrix adsorption the PCR products were diluted 1:50 and 1µl of each used as a template for the second round of PCR (right). The second round of PCR produced the full length NS34A[R-G] gene (lanes 1 and2) and the NS34A[K-Q, R-G] gene (lanes 4 and 5) with suitable restriction enzyme sites at both termini for cloning into pTF7.5. The left-hand lane of each pair used a 1:20 dilution of the PCR primers and the right-hand lane of each pair used a 1:100 dilution.

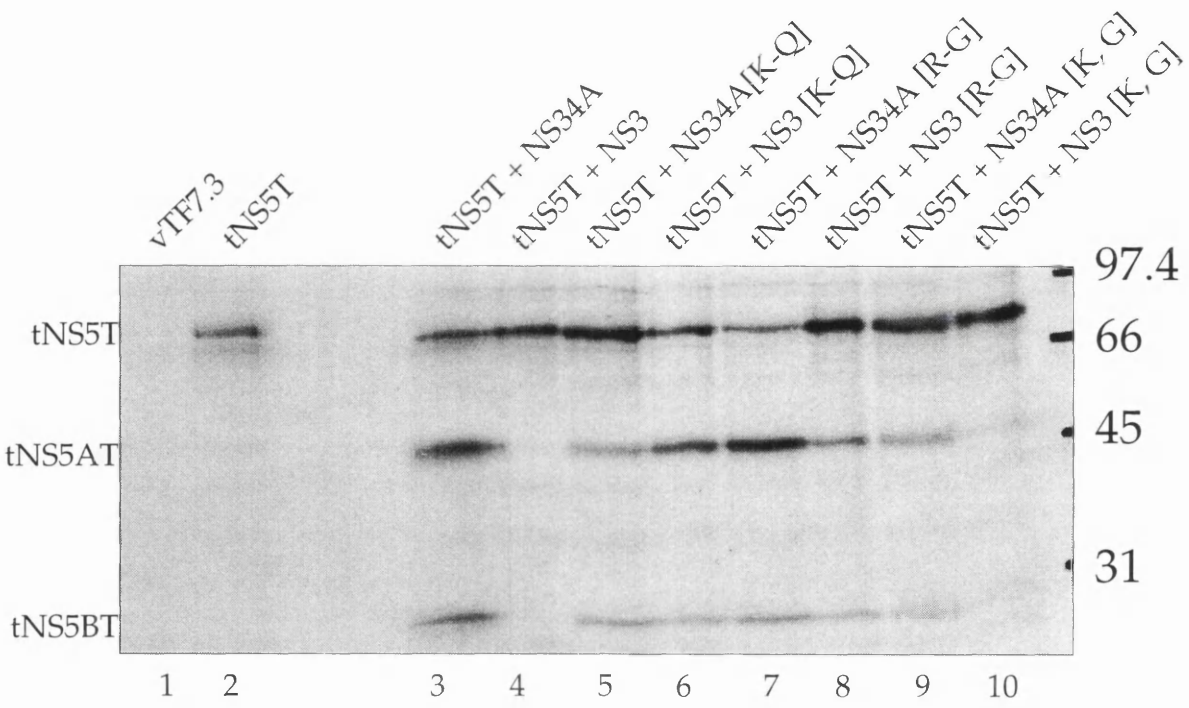


Fig. 7.11. Comparison of processing efficiencies of different NS3 proteases containing aingle and double amino acid changes. vTF7.3 infected CV-1 ceils were left untransfected (lane 1) or were transfected with tNS5T (lane2), tNS5T + NS34A (lane 3), tNS5T + NS3 (lane 4), tNS5T + NS34A[K-Q] (lane 5), tNS5T + NS3[K-Q] (lane 6), tNS5T + NS34A[R-G] (lane 7), tNS5T + NS3[R-G] (lane 8), tNS5T + NS34A[K-Q, R-G] (lane 9), tNS5T + NS3[K-Q, R-G] (lane 10). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The tNS5T substrate was cotransfected with different NS3 and NS34A constructs. All the proteases processed the tNS5T substrate except the wild type HCV Gla NS3 (lane 4). The proteases processed the substrate with varying efficiencies with the most efficient protease the NS34A[R-G] complex (lane 7) and the most efficient NS3 being the NS3[K-Q] (lane 6). Molecular weight markers are on the right.

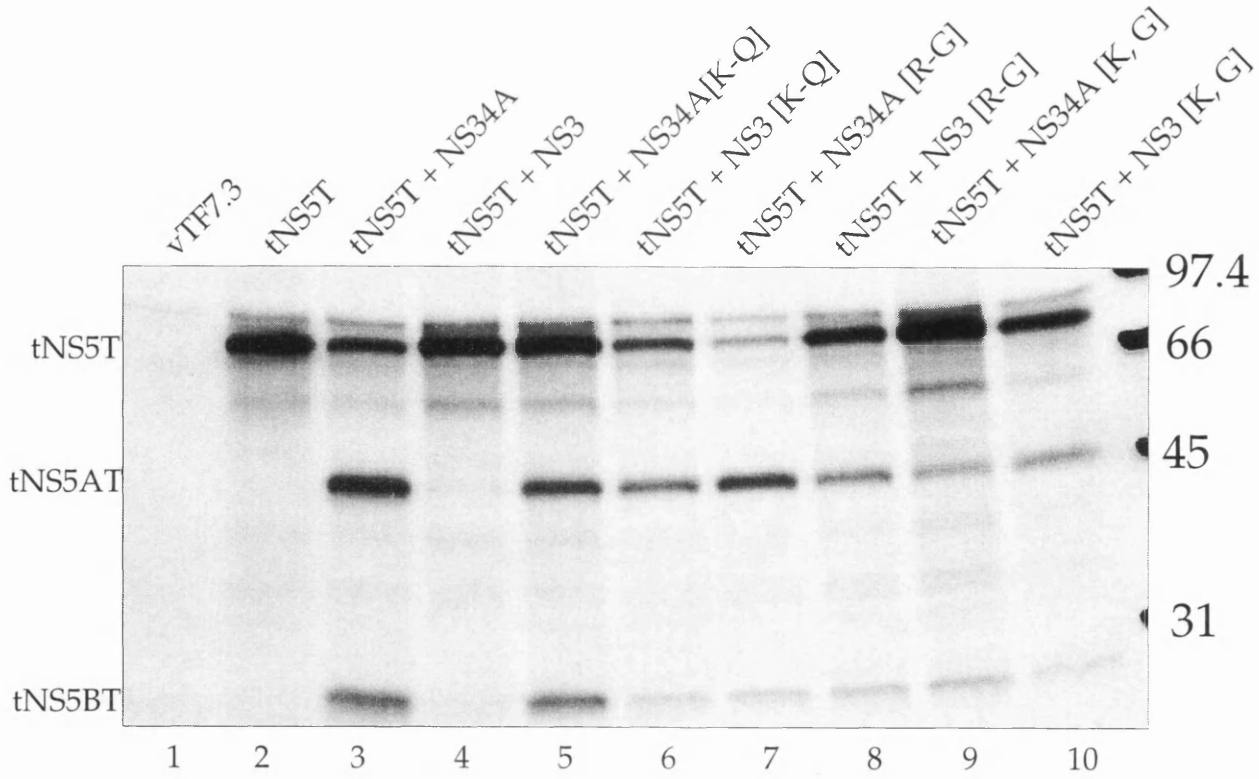


Fig. 7.12.. Repeat analysis of the ability of different proteases to process the tNS5T substrate. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with tNS5T (lane2), tNS5T + NS34A (lane 3), tNS5T + NS3 (lane 4), tNS5T + NS34A[K-Q] (lane 5), tNS5T + NS3[K-Q] (lane 6), tNS5T + NS34A[R-G] (lane 7), tNS5T + NS3[R-G] (lane 8), tNS5T + NS34A[K-Q, R-G] (lane 9), tNS5T + NS3[K-Q, R-G] (lane 10). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. The tNS5T substrate was cotransfected with different NS3 and NS34A constructs. All the proteases processed the tNS5T substrate except the wild type NS3 (lane 4). The proteases processed the substrate with varying efficiencies with the most efficient protease the NS34A[R-G] complex (lane 7) and the most efficient NS3 being the NS3[K-Q] (lane 6). Molecular weight markers are on the right.

all the altered NS3 proteases processed at a similar efficiency (approximately 50% the efficiency of the most efficient NS34A complex).

The efficiency of processing of these proteases constructs is summarised in Table 7.1.

Protease Construct	Relative % Cleavage Efficiency
Wild Type NS34A	100
NS34A[K-Q]	39 - 44
NS34A[R-G]	110
NS34A[K-Q, R-G]	28 - 44
Wild Type NS3	0
NS3[K-Q]	44 - 50
NS3[R-G]	33 - 39
NS3[K-Q, R-G]	28 - 44

Table 7.1. Relative % cleavage efficiency of HCV Gla proteases containing various point mutations. NS34A[R-G] was the most efficient protease at processing tNS5T. The most efficient NS3 contained the K-Q substitution, although by only a very small margin.

The R-G substitution within NS3 exerted a positive influence on processing that was more significant when the NS3 protease was expressed as NS34A. The K-Q substitution appeared to stimulate processing when NS3 was expressed alone but when expressed as NS34A there was an inhibitory effect. When both substitutions were present in the same protease NS34A[K-Q, R-G] processed at a similar level to the NS34A[K-Q] construct. The effect of the R-G mutation does not seem to have an effect when present with the K-Q mutation. The NS3[K-Q, R-G] processing efficiency was essentially the same as the other two NS3 proteins. These observations were unexpected and further investigation was carried out to determine if any other factors were influencing these results.

It was speculated that contamination within the DNA preparations could be interfering with the assay and therefore account for these observed differences in processing. Four protease constructs (NS34A[R-G], NS3[R-G], NS34A[K-Q, R-G], NS3[K-Q, R-G]) were digested with *Bam* HI to produce the entire insert. The digested vectors were analysed through a 1% TBE gel to determine a) the purity of the DNA preparations and b) that the correct digest patterns were obtained. The digests revealed that although all four DNA preparations were free of contamination the NS34A[K-Q, R-G] preparation produced an insert that was marginally smaller than the NS34A[R-G] and larger than both NS3 constructs (Fig. 7.13.). The NS34A[K-Q, R-G] construct was produced by PCR from the NS34A[K-Q] construct.

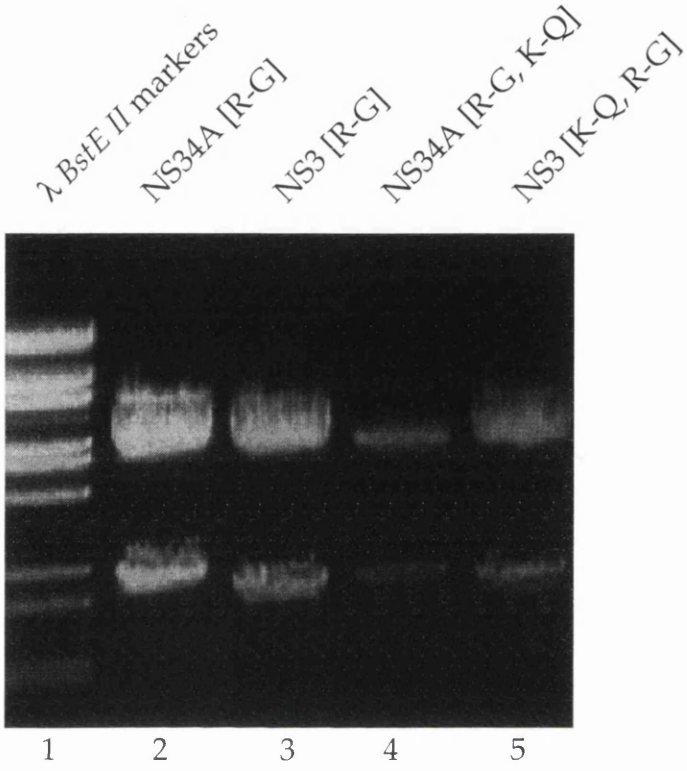


Fig. 7.13. *Bam* HI digestion of four protease cDNAs encoded within pTF7.5. The λ *BstE II* markers (lane 1) illustrate the sizes of the digested DNA fragments. The digests carried out were loaded: NS34A[R-G] (lane 2), NS3[R-G] (lane 3), NS34A[K-Q, R-G] (lane 4) and NS3[K-Q, R-G] (lane 5). Digestion with *Bam* HI excises the entireHCV insert. Both NS3 inserts are the same size (lanes 3 and 5). The NS34A insert derived from the original wild type construct (NS34A[R-G]) is larger than the insert derived from the NS34A[K-Q] construct. (lanes 2 and 4).

In case the DNA digests were misleading all the proteases were expressed by coupled *in vitro* transcription/translation in rabbit reticulocyte lysates and the reactions analysed through a 12% SDS polyacrylamide gel (Fig. 7.14.). The TNT analysis shows that although all the NS34A complexes processed to give the NS3 protein the NS34A[K-Q] complex produced a smaller NS34A band although the NS3[K-Q] band appeared to be the same size as the wild type NS3. The NS34A[K-Q. R-G] complex was not observed and looked to be fully processed to NS3.

Although these constructs expressed an active NS3 serine protease the presence of a small deletion within the cDNA may effect the catalytic activity of the protease produced. Sequencing analysis was not carried out but as the NS3 proteins appeared to be the same size the deletion was presumed to be within the NS4A region. The NS34A[K-Q] and NS34A[K-Q, R-G] construct were digested with *Sal* I and *Sph* I. The *Sal* I site was 50bp upstream of the catalytic histidine codon and the *Sph* I site was 195bp downstream of the catalytic serine codon. Digestion with these enzymes produced a 500bp insert that contained the region encoding the catalytic triad from these cDNAs. These inserts were ligated into *Sal* I/*Sph* I digested pTF7.5 NS34A constructs to give full length NS34A cDNAs that contained the desired amino acid substitutions. After cloning the inserts, the constructs were digested with *Bam* HI to check the size of the new insert (Fig. 7.15.). All the new clones of NS34A[K-Q] and NS34A[K-Q, R-G] produced inserts that were the same size as the wild type NS34A. The new constructs were tested for expression by coupled *in vitro* transcription/translation in rabbit reticulocyte lysates and analysed by SDS-PAGE (Fig. 7.16.). The NS34A constructs produced NS34A bands the same size as the wild type NS34A and were subsequently tested for their *trans*-cleavage ability in comparison with the previous clones (Figs. 7.17. and 7.18.). All the correctly sized NS34A complexes processed the tNS5T substrate with high efficiencies. The NS34A constructs containing the lysine to glutamine substitution produced protease complexes that were more efficient than the other two NS34A complexes (wt NS34A and NS34A[R-G]).

The efficiencies with which the NS34A complexes processed tNS5T are summarised in Table 7.2.

NS34A Constructs	Relative % Cleavage Efficiency
Wild Type NS34A	100
NS34A[K-Q]	122 - 125
NS34A[R-G]	100
NS34A[K-Q. R-G]	120 - 125

Table 7.2. Efficiency of processing by the full length NS34A complexes. Both the wild type NS34A and the NS34A[R-G] process at similar levels. The inclusion of the K-Q substitution gives optimal processing both with and without the R-G substitution.

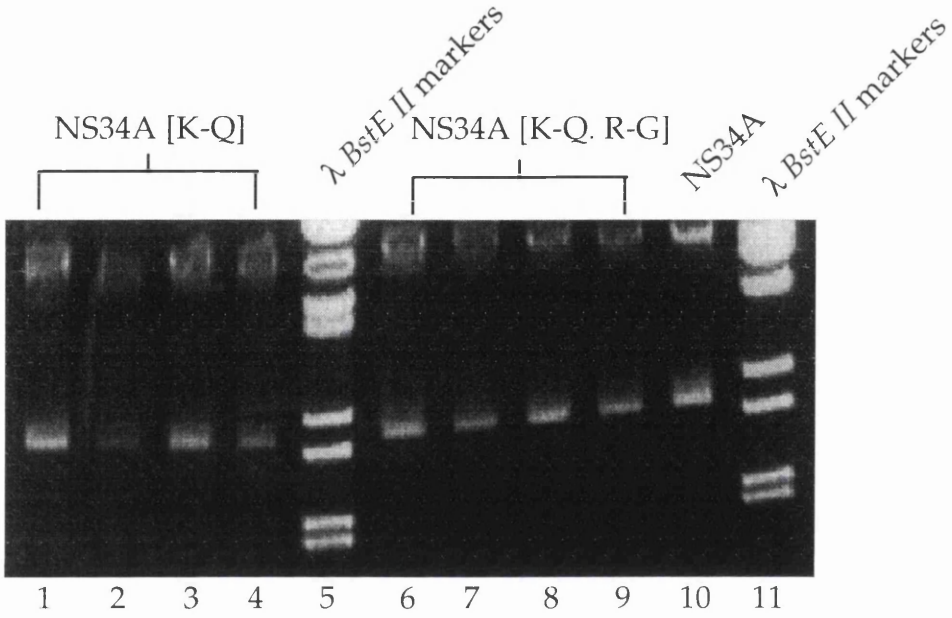


Fig. 7.15. *BamH I* restriction enzyme digestion of recloned mutant NS34A constructs. Digestion with *Bam* HI excises the entire insert. Four NS34A[K-Q] constructs (lanes 1-4), four NS34A[K-Q, R-G] constructs (lanes6-9) and the wild type NS34A construct (lane 10) were digested with *BamH I* before analysis through a 1% TBE agarose gel. All eight mutant NS34A plasmid preparations contained correctly sized inserts (in comparison to the wild type NS34A control digest).

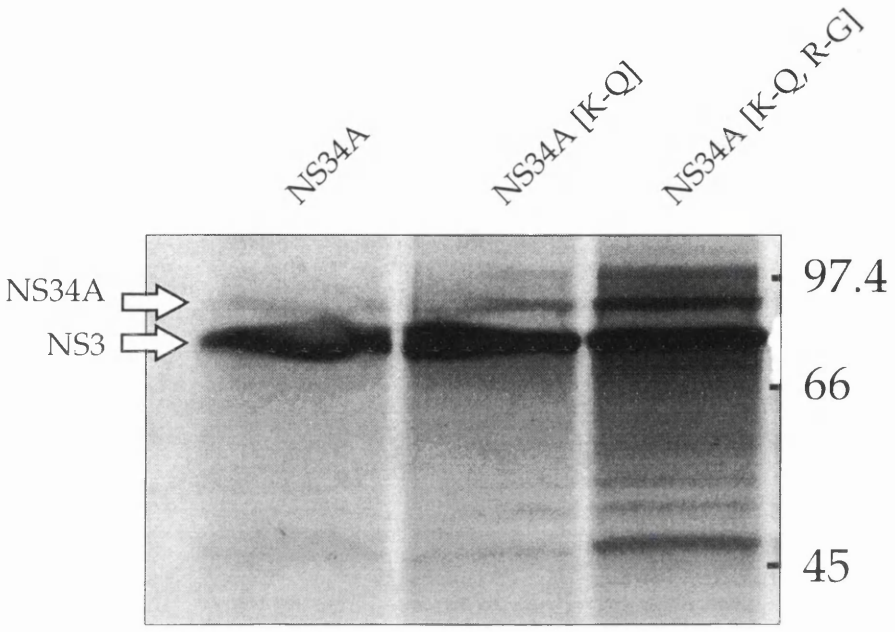


Fig. 7.16. Expression of recloned mutant NS34A constructs by *in vitro* transcription/translation. The wild type NS34A (lane 1), the NS34A[K-Q] (lane 2) and the NS34A[K-Q, R-G] were all expressed within a 25 µl TNT reaction and then analysed by SDS-PAGE. This analysis corroborated the observation from the restriction enzyme digestion analysis of the plasmid DNAs (Fig. 7.14.). The recloned NS34A[K-Q] and NS34A[K-Q, R-G] constructs produce NS34A complexes of the correct size when expressed *in vitro*. These NS34A complexes self-processed to produce NS3 protein bands of the correct size. The molecular weight markers are located at the right of the gel.

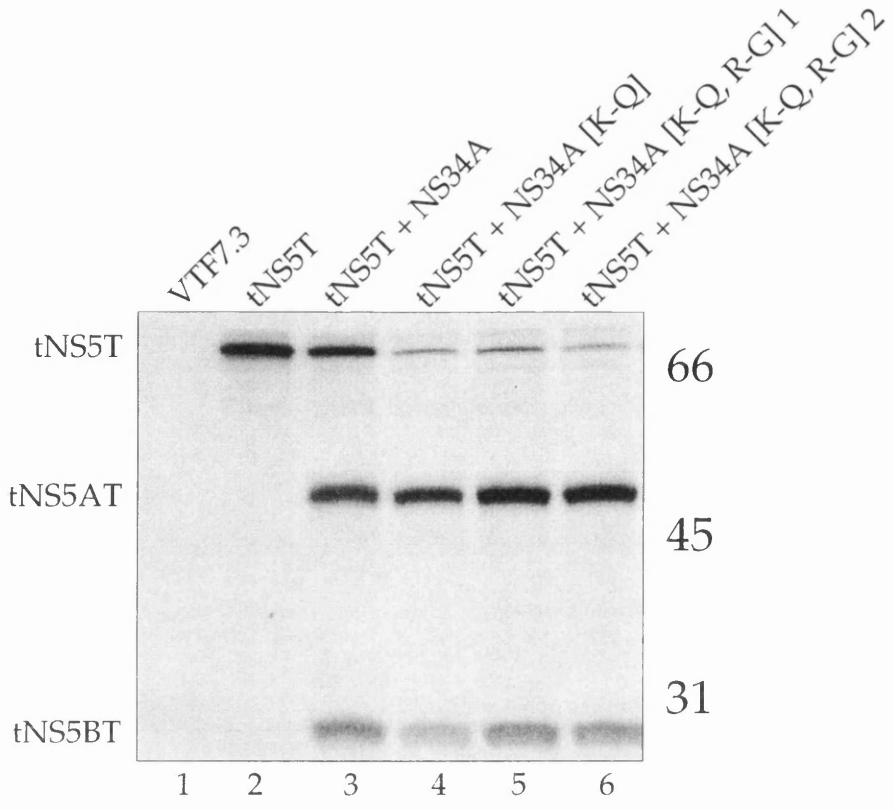


Fig. 7.17. *Trans*-cleavage analysis with recloned mutant NS34A constructs. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or transfected with tNS5T alone (lane 2) or tNS5T + NS34A (lane 3), tNS5T + NS34A[K-Q] (lane 4), tNS5T + NS34A[K-Q, R-G]1 (lane 5), tNS5T + NS34A[K-Q, R-G]2 (lane 6). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine, lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. All three, correctly-sized mutant NS34A complexes were cotransfected with the tNS5T substrate and their cleavage efficiencies compared to the wild type NS34A construct (lanes 3 to 6). The introduction of the amino acid substitution lysine to glutamine increased the efficiency of processing with respect to the wild type (lane 4). The difference in cleavage efficiency when the arginine to glycine mutation was introduced to the NS34A[K-Q] construct appeared to be negligible (lanes 5 and 6). The molecular weight markers are located at the right of the gel.

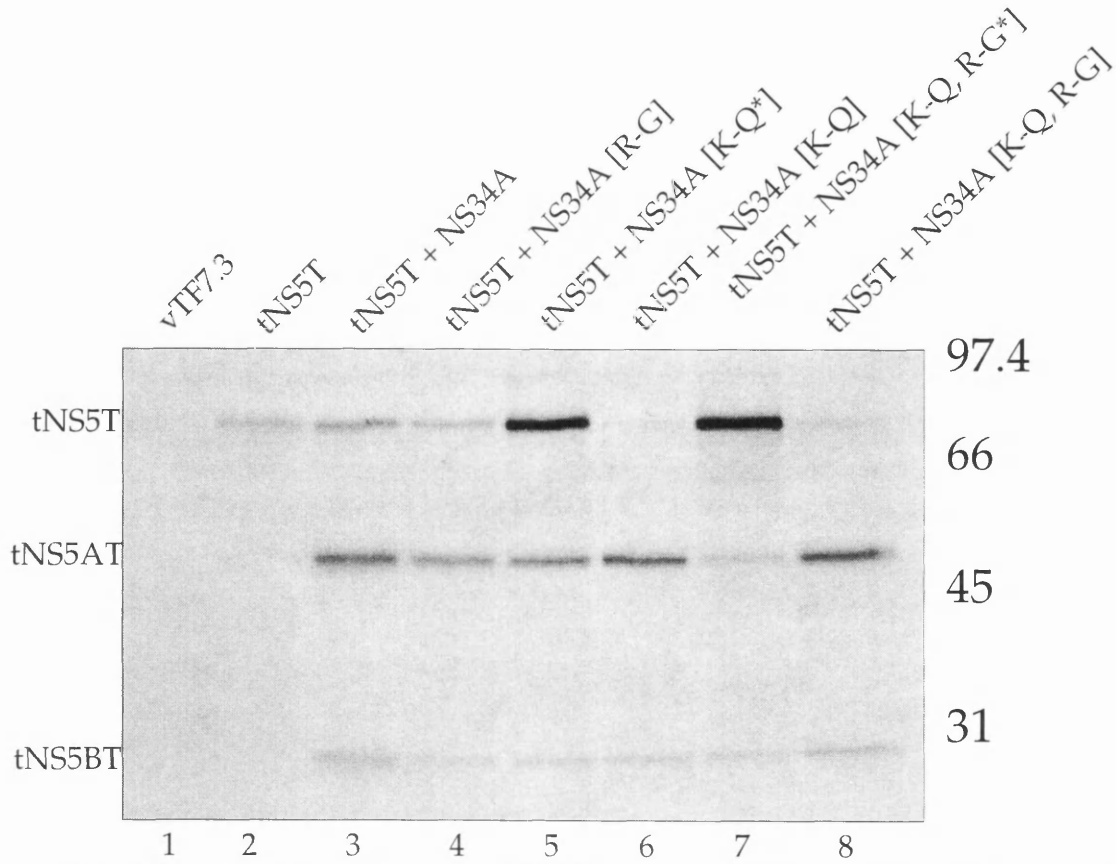


Fig. 7.18. Comparison of processing efficiencies between truncated and full length mutant NS34As. vTF7.3 infected CV-1 cells were left untransfected (lane 1) or were transfected with tNS5T (lane 2), tNS5T + NS34A (lane 3), tNS5T + NS34A[R-G] (lane 4), tNS5T + NS34A[K-Q*] (lane5), tNS5T + NS34A[K-Q] (lane 6), tNS5T + NS34A[K-Q, R-G*] (lane 7), tNS5T + NS34A[K-Q, R-G] (lane 8). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. All the NS34A complexes process the tNS5T substrate. The correct NS34A[K-Q] complex (lane 6) processed the tNS5T with an improved efficiency in comparison to the original NS34A[K-Q] complex (lane 5). This observation was also seen for the original and correct NS34A[K-Q, R-G] complexes (lanes 7 and 8 respectively). Both correct complexes (lanes 6 and 8) processed at a similar level of efficiency indicating that the lysine to glutamine substitution had a significant effect. Molecular weight markers are indicated on the right

Both the wild type NS34A construct and the NS34A[K-Q, R-G] construct were cotransfected into vTF7.3 infected CV-1 cells with the P6, P1 and P1′ tNS5T substrate mutants described in section **3.3.2.** to see if there was any difference in their ability to process the mutant cleavage sites. After labelling with ^{35}S-methionine the cell lysates were immuneprecipitated with the SV5-P-k antibody and analysed by SDS-PAGE (Figs. 7.19. and 7.20.). Despite the NS34A[K-Q, R-G] processing more of the substrate than the wild type NS34A there was no difference in the relative amounts of processing for the mutant cleavage sites. The cleavage efficiencies with which the proteases cleaved the mutant cleavage sites are illustrated below.

Substrate	Relative % Cleavage Efficiency - Wild Type	Relative % Cleavage Efficiency - [K-Q, R-G]
Wild Type	100	100
P6 Glu - Gln	30	35
P6 Glu - Lys	15-25	20
P1 Cys - Ser	<10	<10
P1 Cys - Arg	0	0
P1 Cys - Gly	0	0
P1′ Ser - Thr	30-40	33
P1′ Ser - Ala	55-60	55
P1′ Ser - Pro	0	0

Table 7.3. Comparison of processing between wild type NS34A and NS34A[K-Q. R-G]. Even though there was a difference in the amount of substrate processed by two NS34As the relative processing of the mutant cleavage sites was the same.

7.3. Investigation into processing of the HCV Gla substrate by an NS3 serine protease from a different genotype.

There are six major genotypes of HCV and the polyprotein produced by each genotype is processed by the NS3 serine protease. Each genotype contains amino acid differences compared to the polyprotein of a different genotype. The ability of an NS3 serine protease from one genotype to process a substrate from another has not been significantly investigated. It has been shown that a type 1a protease can process a type 1b substrate and vice versa (Lin *et al.*, 1994b). A plasmid encoding the protease domain of one genotype was cotransfected with a polyprotein substrate from the C-terminal region of NS3 to the end of NS5B. The NS3 protease added *in trans* was able to process all four cleavage sites of the polyprotein of a different

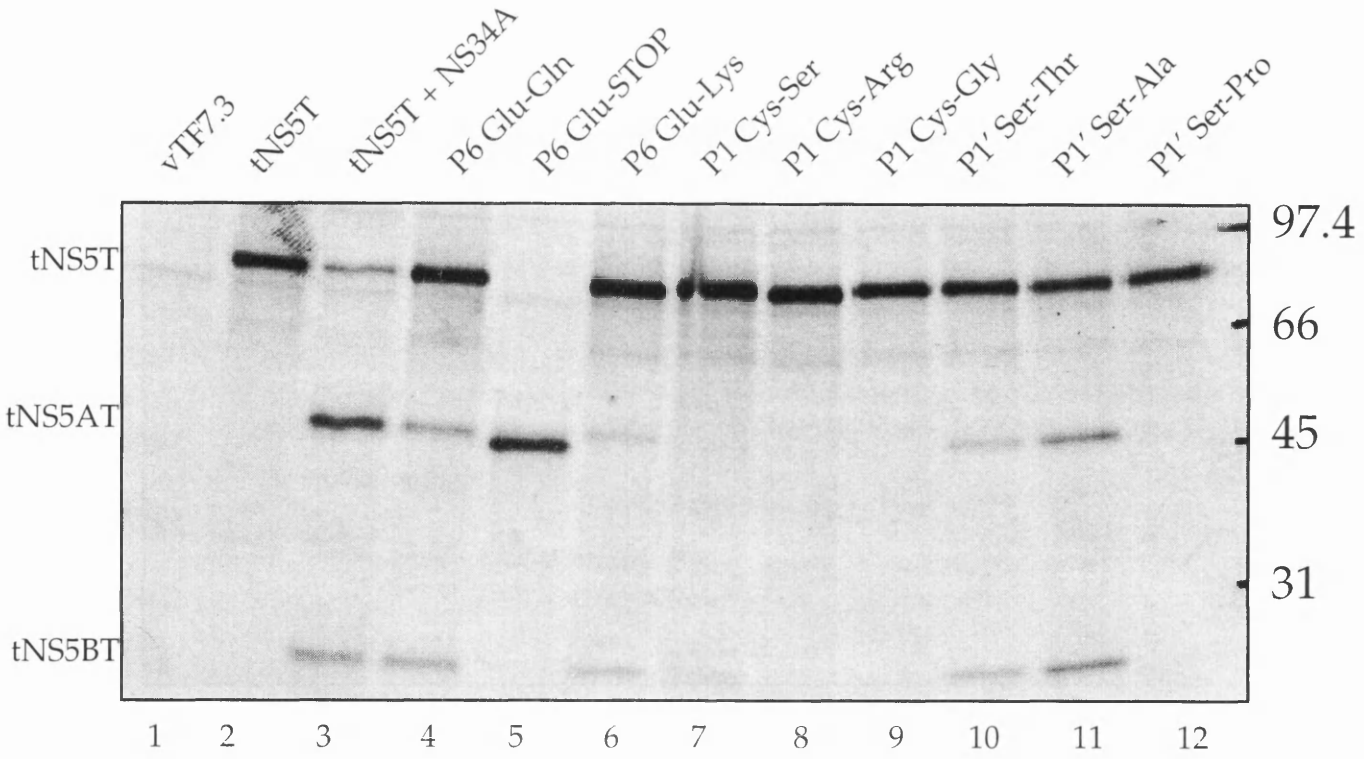


Fig. 7.19. Processing of all nine mutant cleavage sites (P6, P1, P1′) by the wild type NS34A. vTF7.3 infected CV-1 cells were either untransfected (lane 1) or transfected with 1μg tNS5T (lane 2) or1μg substrate + 5μg NS34A. The substrates transfected with NS34A were tNS5T (lane 3), P6 Gln (lane 4), P6 STOP (lane 5), P6 Lys (lane 6), P1 Ser (lane 7), P1 Arg (lane 8), P1 Gly (lane 9), P1′ Thr (lane 10), P1′ Ala (lane 11), P1′ Pro (lane 12). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. Every amino acid substitution illustrated above was encoded at the NS5A/NS5B cleavage site within the tNS5T substrate. This analysis illustrated the effect of mutating P6, P1 and P1′ of the NS5A/NS5B cleavage site and the importance of each residue in substrate specificity. Molecular weight markers are indicated on the right.

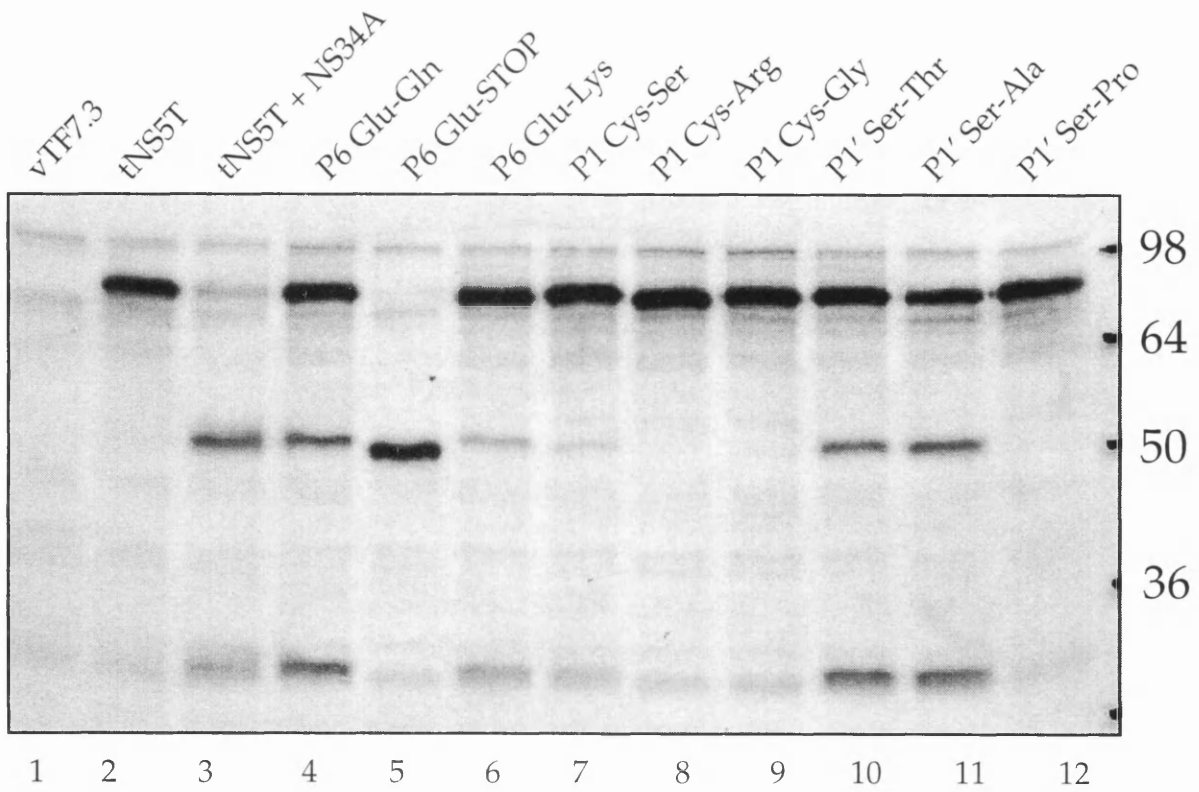


Fig. 7.20. Processing of all nine mutant cleavage sites (P6, P1, P1′) with the optimal protease NS34A[K-Q, R-G]. vTF7.3 infected CV-1 cells were either untransfected (lane 1) or transfected with 1μg tNS5T (lane 2) or1μg substrate + 5μg NS34A[K-Q, R-G]. The substrates transfected with NS34A[K-Q, R-G] were tNS5T (lane 3), P6 Gln (lane 4), P6 STOP (lane 5), P6 Lys (lane 6), P1 Ser (lane 7), P1 Arg (lane 8), P1 Gly (lane 9), P1′ Thr (lane 10), P1′ Ala (lane 11), P1′ Pro (lane 12). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine for 2 hours, lysed and then immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. Every amino acid substitution indicated above was encoded at the NS5A/NS5B cleavage site within the tNS5T substrate and cotransfected with the NS34A [K-Q, R-G] construct. The cleavage pattern obtained for each substrate was the same as that observed for cleavage by the wild type NS34A. Molecular weight markers are indicated on the right.

subtype. This result has been corroborated by the work carried out in **7.2.** as the HCV-A is a type 1b isolate and its NS3 protease was able to process the type 1a HCV Gla NS5A/NS5B substrate.

A type 4 HCV isolate had been sequenced within the laboratory (Chamberlain *et al.*, 1997) and so genotype 4 HCV clones were available in the laboratory to attempt to produce a type 4 HCV protease construct. Three independent plasmid clones (clones 9, 14, 15) corresponding to the 5′ half of the type 4 open reading frame (encompassing the NS34A region) were supplied by Dr. R. Chamberlain. PCR primers were designed for the amplification of the type 4 HCV NS34A gene by PCR and ligation of the PCR product into pTF7.5. These primers were similar to those used for the production of the HCV Gla NS34A construct but were based on the type 4 sequence. The 5′ primer contains a *Bam* HI site followed by a start codon and nucleotide sequence corresponding to the 5′ end of the NS3 gene. The 3′ primer contains a *Bam* HI site, a termination codon and nucleotide sequence complementary to the 3′ end of the NS4A gene.

PCRs using the proof-reading *Pfu* DNA polymerase were carried out using the three templates (clones 9, 14, 15) and these reactions produced inserts of the expected size (Fig. 7.21.). The PCR products were digested with *Bam* HI and ligated into *Bam* HI cut pTF7.5 the orientation of the insert was checked by digestion with *Nco* I/*Xba* I. There is an *Xba* I site 668 nucleotides upstream of the multiple cloning site in pTF7.5 and an *Nco* I site 430 nucleotides downstream of the initiation codon of the NS3 gene. Digestion of the type 4 NS34A gene with *Nco* I/*Xba* I would give an insert of approximately 1100bp if the insert was ligated in the correct orientation. A number of type 4 NS34A clones were obtained but analysis by coupled *in vitro* transcription/translation (TNT) showed that none of the constructs produced an NS34A complex that processed *in cis* to give the cleaved NS3 protein (Fig. 7.22.). The nucleotide sequence of a few of the clones was determined using an automated sequencer to determine whether any mutations had been introduced into the catalytic triad. Such a mutation may have explained the lack of processing observed. The sequence obtained showed that the nucleotide sequence encoding the amino acids of the catalytic triad was correct.

The inability of the type 4 NS34A constructs to self-process may be due to a mutation in the NS3/NS4A cleavage site. However as it has been demonstrated that the NS3 protease is still active even when the NS4A protein is uncleaved from the C-terminus of the NS3 protein, these constructs were tested for *trans*-cleavage activity in case the construct was still active despite the presence of the uncleaved NS4A. No processing was observed for any of these type 4 NS34A constructs (Fig. 7.23.). There was insufficient time to completely sequence all the clones to determine the reason for their inability to process.

An active type 4 NS34A construct was subsequently cloned by Dr R. Chamberlain and this construct, when expressed *in vitro*, could self-process. This construct has been shown to process the tNS5T substrate using the *trans*-cleavage assay I developed.

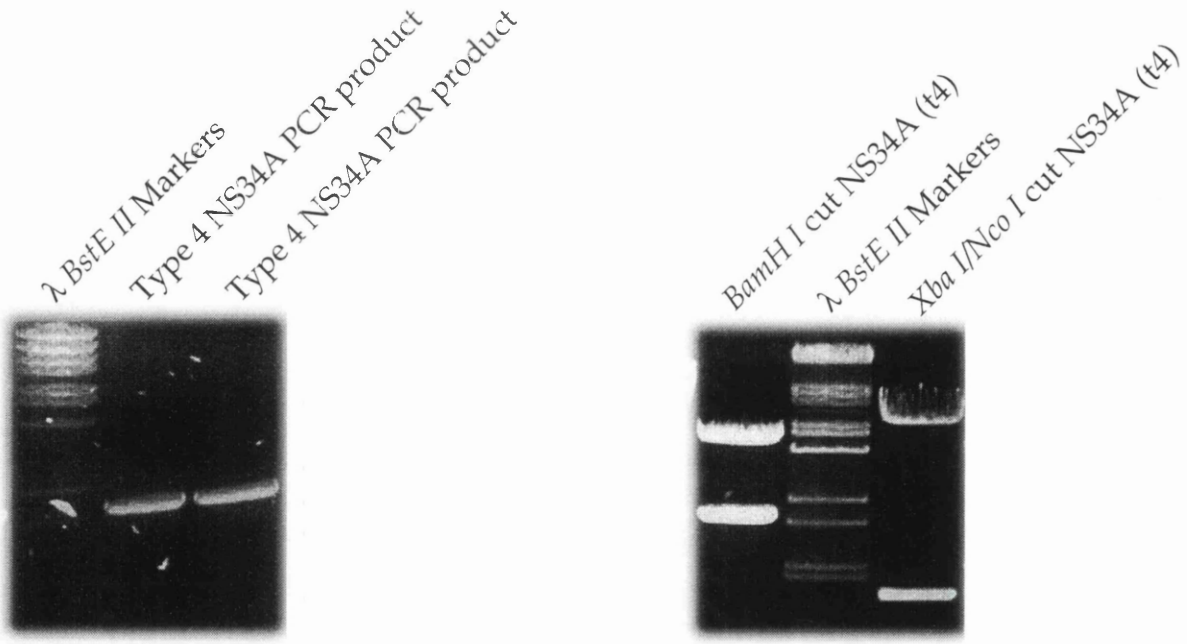


Fig. 7.21. Cloning of HCV genotype 4 NS34A. The type 4 NS34A cDNA was amplified by PCR using *Pfu* DNA polymerase (left) and cloned into pTF7.5. The insert was introduced using *Bam* H I cut pTF7.5 and checked for orientation by digestion with *Xba* I/*Nco* I (right). The digest with *Bam* HI confirms that the insert is the correct size.

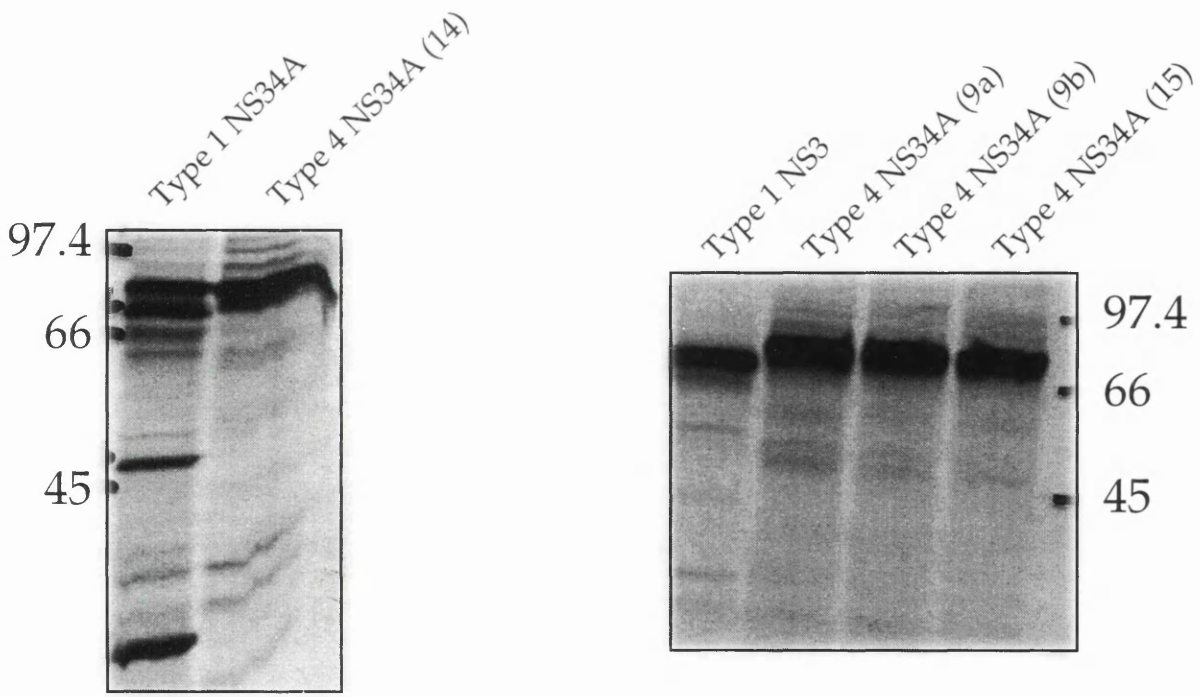


Fig. 7.22. ***In vitro*** expression of type 4 NS34As in rabbit reticulocyte lysates. All four type 4 NS34A constructs (derived from different long range PCR products) produce NS34A bands of the correct size but that do not process to NS3.

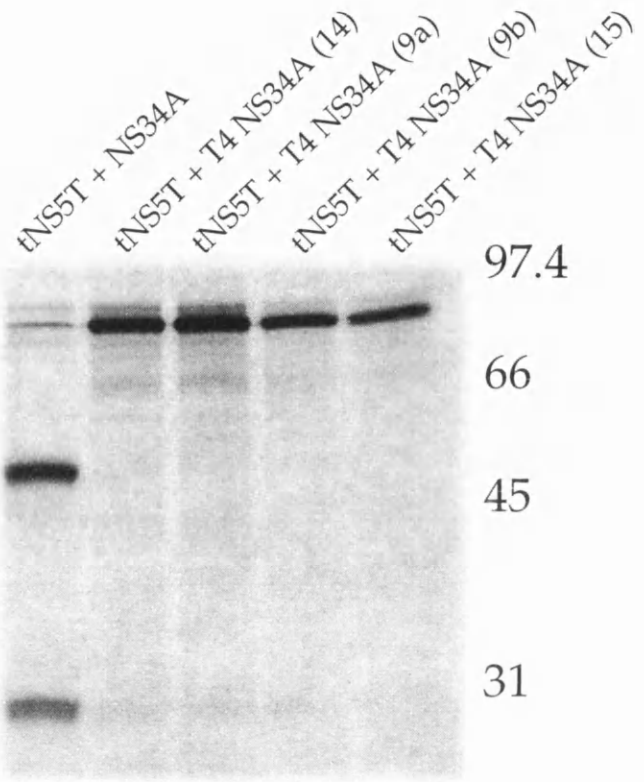


Fig. 7.23. Analysis of trans-cleavage activity of the type 4 NS34A clones. vTF7.3 infected CV-1 cells were transfected with tNS5T + type 1 NS34A (lane 1) or tNS5T + type 4 NS34A (lanes 2-5). The type 4 NS34A constructs were derived from three separate long range PCR products (9, 14, 15). Fifteen hours post-transfection the cells were labelled with ^{35}S-methionine, lysed and immuneprecipitated with the SV5-P-k antibody before analysis by SDS-PAGE. None of the type 4 NS34A clones processed the tNS5T substrate when co-transfected into mammalian cells. The molecular weight markers are located at the right of the gel.

Summary of results

An assay for *trans*-cleavage of HCV substrates by the NS3 serine protease was developed. This assay involved the co-transfection of the NS3 serine protease and relevant substrate cDNAs into mammalian cells infected with vTF7.3. After labelling with ^{35}S-methionine, cell lysis and immuneprecipitation the samples were analysed by SDS-PAGE to determine the success of the *trans*-cleavage.

Characterisation of the substrate requirements for processing was carried out. The effects of mutagenesis of the P6, P1 and P1' residues within the NS5A/NS5B cleavage site were examined as was the ability of the NS3 protease to process potential cleavage sites. No (potential) cleavage sites, apart from those reported, were processed by NS3. A specific potential site (LITS) was examined in detail to determine if any specific residue was responsible for the absence of processing. No single residue was responsible for this effect.

An investigation of the protease requirements for processing at the NS5A/NS5B cleavage site was carried out. The HCV Gla NS3 serine protease was only able to process the NS5A/NS5B cleavage site in the presence of the NS4A protein, in contrast to reported data. A *trans*-cleavage assay was carried out using an NS3 protease from a different HCV isolate and this NS3 alone was able to cleave the NS5A/NS5B cleavage site within tNS5T. An amino acid sequence comparison was carried out between the HCV Gla NS3 and the HCV-A NS3 and any differences examined. A number of potentially important amino acid changes were discovered and the effect of these on serine protease activity investigated. Replacement of two amino acids with the HCV-A NS3 amino acids enabled the HCV Gla NS3 protease to process the NS5A/NS5B cleavage site in the absence of NS4A.

The effect of the NS2 protein uncleaved from the N-terminus of NS3 on the serine protease activity was investigated. The presence of this uncleaved protein did not inhibit NS3 protease activity.

Chapter 8 - Discussion

8.1. Development of a *trans*-cleavage assay for the NS3 serine protease.

The initial object of this study was to develop a convenient assay for the analysis of processing by the NS3 serine protease. With such an assay for the *trans*-cleavage by the NS3 serine protease the substrate and protease requirements for processing could be examined. A *trans*-cleavage assay using *in vitro* transcription/translation in rabbit reticulocyte lysates proved to be of no use. Although the protease and substrate proteins were expressed well *in vitro* no evidence for processing was observed. The co-expression of substrate and protease within mammalian cells resulted in cleavage of both a full length and a truncated (at both termini) NS5 substrate. The truncation of the NS5B region at the C-terminus produced a smaller NS5B protein that was stable when expressed in cells. This truncated NS5B protein was easily identifiable using a monoclonal antibody against an epitope tag introduced at its C-terminus.

In **5.1.** a *trans*-cleavage assay was established and optimised with cotransfection of 1μg substrate cDNA and 5μg of protease cDNA into T7 polymerase expressing mammalian cells. The cDNAs were cotransfected using either cationic liposomes or polyethylene imine (PEI). Fifteen hours post transfection the cells were labelled with ^{35}S-methionine, lysed and immuneprecipitated with the relevant antisera before analysis by SDS-PAGE.

Having established an assay for cleavage *in trans* by the NS3 serine protease, analysis of the substrate requirement for processing was examined.

It was not known how authentic the tNS5T substrate was in comparison to the full length NS5. The tNS5T substrate may have folded in a different manner due to the truncations at both termini giving a different secondary structure. The "Plotstructure" program was used to plot the measures of protein secondary structure of the NS5 substrates. The secondary structure was represented by a two-dimensional "squiggly" diagram that demonstrated the Chou-Fasman prediction for the protein. The secondary structural features were initially calculated using the "Peptidestructure" program, from which the two-dimensional representation was plotted. The secondary structures of the full length NS5 and truncated NS5 substrate were calculated and their representations compared (Fig. 8.1.). The peptide sequence corresponding to the tNS5T region is predicted to have the same secondary structure when expressed as part of the full length NS5 protein as it does when expressed alone.

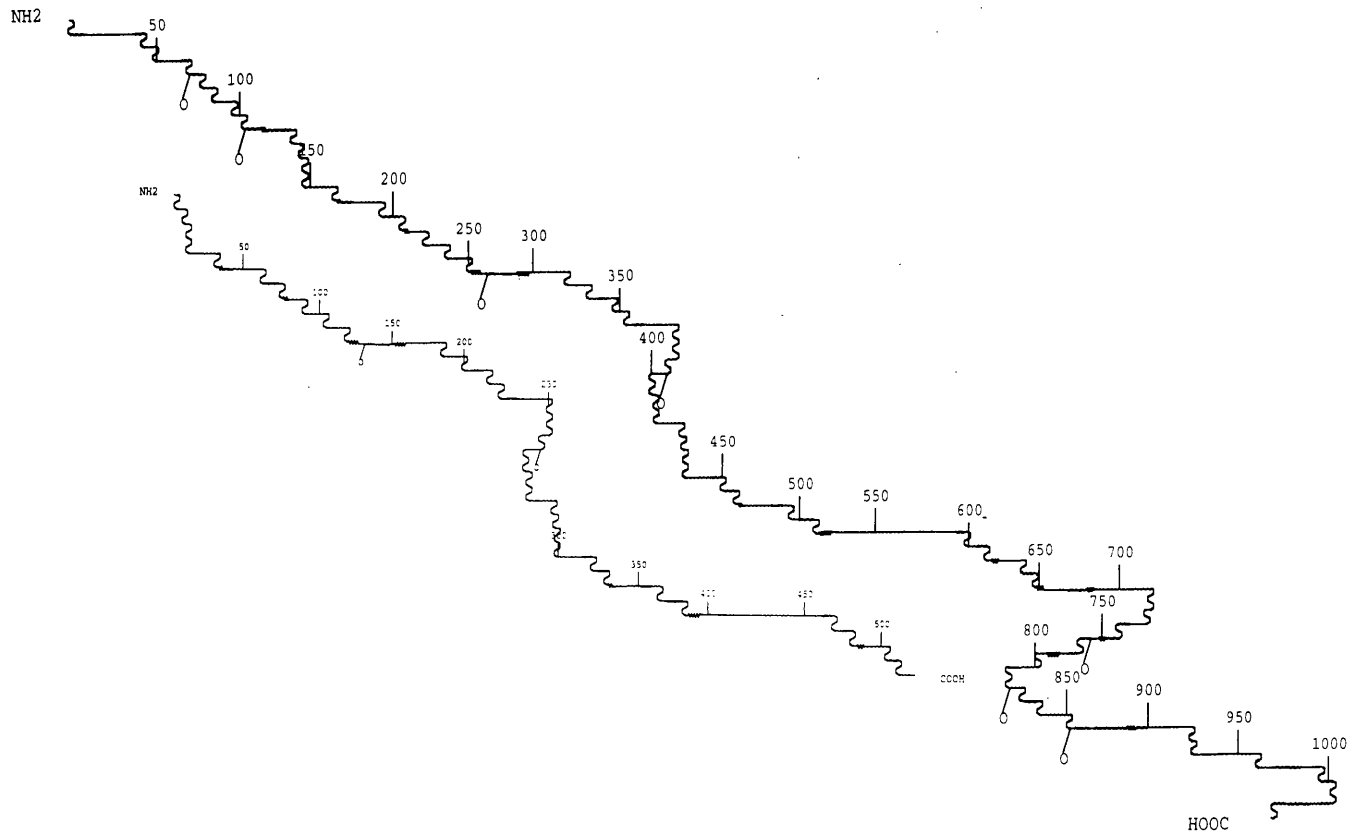


Fig. 8.1. The secondary structure of the full length NS5 and the truncated tNS5T substrates were predicted using the Plotstructure program. This program represents the secondary structure as a two-dimensional "squiggly" diagram and plots the Chou-Fasman prediction. The tNS5T substrate predicted secondary structure (below) is identical to the same region within the full length NS5 substrate.

8.2. Mutagenesis of the cleavage site motif of the NS5A/NS5B cleavage site.

Since the beginning of my project there have been a number of mutagenesis studies carried out on the substrate specificity of NS3 mediated cleavage. These analyses were carried out on several different cleavage sites in an attempt to fully determine the importance of any conserved amino acids. Within the HCV polyprotein there are four sites processed by the NS3 serine protease (NS3/NS4A, NS4A/NS4B, NS4B/NS5A, NS5A/NS5B). All these sites are processed in *trans* by NS3 except the NS3/NS4A cleavage site which is processed in *cis*. By comparing the amino acid sequences of these cleavage sites the consensus amino acid motif for an NS3 mediated cleavage site was determined. This motif consisted of an acidic residue at P6 (aspartic acid or glutamic acid), a cysteine or threonine at P1 and a serine residue at P1'. The NS3-mediated cleavage occurred between P1 and P1'. As only three amino acid positions exhibited any conservation it was these three positions that were extensively studied.

Leinbach *et al.* (1994) carried out a mutational analysis of the HCV BK (type 1b) NS3/NS4A site *in vitro* and showed that when the amino acids at P6 and P1' were mutated there was little or no effect on the efficiency with which this site was processed. Substitutions to polar or aliphatic side chain containing amino acids were permitted at P1, with the cleavage site still processed. There was no tolerance for basic or aromatic residues at P1 and the degree of processing was drastically reduced. NS3/NS4A is processed in *cis* and the restrictions for processing may be different for processing in *trans*. The NS3 protein folds around the substrate in a conformation suitable for processing. As the cleavage site is part of the protein the amino acids defining may not be as important in protease recognition of the cleavage site.

Substitution at the HCV-H (type 1a) NS4A/NS4B site of the P1 L-Cys to D-Cys was not tolerated indicating a stereospecific S1 binding pocket for L-Cys (Zhang *et al.*, 1997). At the HCV BK NS4A/NS4B site, substitutions at P1 abolished processing although threonine was tolerated (Urbani *et al.*, 1997). The replacement of the sulphydryl group of the cysteine residue to an amino or a hydroxyl group abolished cleavage at this site. The P1 residue was determined to make a relatively minor contribution to ground state binding (Urbani *et al.*, 1997). The effect of mutation at the P6 residue at the NS4A/NS4B cleavage site was dependent on the amino acid substitution. Neutralisation of the negative charge decreased the efficiency of processing by a factor of five, whereas introduction of a positive charge resulted in a more pronounced decrease. This decrease was attributable to an impairment in ground state binding of protease to the substrate (Urbani *et al.*, 1997). Mutations at the P1' residue had little effect although the introduction of a proline residue at this site dramatically reduced cleavage (Kolykhalov *et al.*, 1994). Alanine scanning mutagenesis of the cleavage site showed mutation of P3 had a significant effect on the efficiency with which

this site was cleaved and that P4' favoured hydrophobic residues (Urbani *et al.*, 1997).

Kolykhalov *et al.* (1994) demonstrated that, for the HCV H strain, the NS3/NS4A site was the least sensitive to mutation, followed by the NS4A/NS4B site, the NS4B/NS5A and most sensitive to mutation, the NS5A/NS5B site. In this study the effect of P1 mutations at different sites showed that selected substitutions partially inhibited processing at the NS3/NS4A site. The same mutations at the NS4A/NS4B site showed significantly greater inhibition of processing. For the NS4B/NS5A site only one of these mutations permitted cleavage whereas for the NS5A/NS5B site all the substitutions blocked cleavage.

8.2.1. Mutagenesis of the P6 residue.

In all the cleavage site analyses there has been little study of the NS5A/NS5B cleavage site. The P6, P1 and P1' amino acid substitutions illustrated in **5.2.** demonstrated that mutation of the P6 residue within the NS5A/NS5B cleavage site had a significant negative effect on the efficiency of processing at this site. The amino acid substitutions at P6 are illustrated in Fig. 8.2. The neutralisation of the negative charge by substitution of the wild type glutamic acid with glutamine resulted in a fall in the efficiency by which this site was processed by approximately 70%. The glutamine side chain is the same relative size and organisation as the glutamic acid side chain with the negatively charged oxygen replaced by the NH_2 group. The replacement of the negatively charged glutamic acid with the positively charged lysine residue had an even more deleterious effect on cleavage at this site, with a fall in efficiency by 80-85%. The lysine side chain is longer and thinner than the wild type glutamic acid as well as containing the positively charged NH_3 group.

At the NS3/NS4A cleavage site, mutation of the P6 aspartic acid to either asparagine, lysine or serine had no effect on the efficiency of processing, suggesting that P6 was not important in cleavage site specificity (Leinbach *et al.*, 1994). This site is processed *in cis* and so the P6 residue may be important in specificity of cleavage when processing takes place *in trans*. At the NS4A/NS4B cleavage site, Kolykhalov *et al.* (1994) demonstrated that mutation of the P6 residue to threonine, alanine, asparagine, glutamic acid or proline had no effect on the efficiency with which the cleavage site was processed. These observations also imply that an acidic residue at P6 is not important. In contrast to this it was shown that, at the same site within a different HCV isolate, the neutralisation of the negative charge at P6 (by the substitution of the glutamic acid with asparagine) led to a decrease in the efficiency of processing by a factor of 5 (Urbani *et al.*, 1997). The introduction of a positively charged lysine residue for the glutamic acid led to an even more pronounced decrease in cleavage efficiency (by a factor of 12). These effects were observed in the absence of the NS4A cofactor (Urbani *et al.*, 1997). In the presence of NS4A the effects of these mutations were less

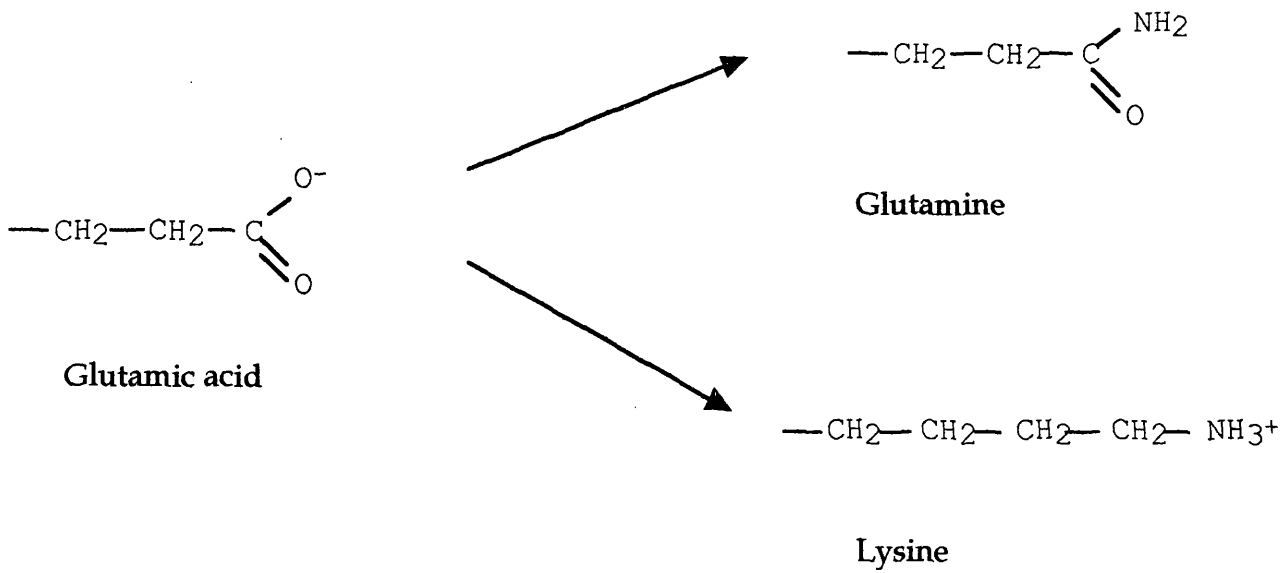


Fig. 8.2. Mutagenesis of the P6 residue. The wild type amino acid at P6 (Glutamic acid) of the NS5A/NS5B cleavage site is illustrated on the right. Substitution of this amino acid with glutamine has very little effect on the size of the side chain but there is a loss of the negative charge. Substitution with lysine introduces a longer side chain with a positive charge instead of the negative charge.

pronounced (a decrease in the efficiency of cleavage by a factor of 2 and 5 respectively). These results illustrate similar effects as those observed in **5.2.**

The importance of the P6 residue on the specificity of cleavage by the NS3 serine protease appears to depend on the nature by which the cleavage site is processed. The P6 residue is essentially unimportant for processing *in cis* at the NS3/NS4A site. Although an acidic P6 residue is not essential in the cleavage sites processed *in trans* its presence is required for optimal cleavage. The P6 residue may only be important for the cleavage sites processed *in trans* because of the nature in which these sites are processed. A possible way to determine if cleavage *in cis* lessens the importance of the P6 residue could be to replace the NS3/NS4A cleavage site with a *trans*-cleavage site (NS4A/NS4B, NS4B/NS5A, NS5A/NS5B). This would produce an NS34A substrate containing a cleavage site normally processed *in trans*. If this altered NS34A substrate was processed then mutation of the P6 residue would determine whether it was the residue within the cleavage site that alone had the effect or if the nature by which the cleavage site was processed was also important.

Determination of the crystal structure of the NS3 showed that the P6 residue of the cleavage site interacted with Arg-161 and Lys-165 of the NS3 protein. These two positively charged side chains would interact with the negatively charged P6 acidic side chain. This substrate/protease interaction explains the results obtained for cleavage *in trans* with loss of the negative charge at P6. The introduction of a positive charge especially would be expected to have a deleterious effect on processing as a positive charge at P6 would repel the side chains extending from the protease that normally stabilise the substrate/protease binding. The negligible effect observed for mutagenesis of P6 within the NS3/NS4A cleavage site (Leinbach *et al.*, 1994) may be because the protease and cleavage site are part of the same protein. Despite the antagonistic nature of the mutated P6, the protease still folds around the cleavage site to form the uncleaved NS34A precursor complex before processing takes place.

8.2.2. Mutagenesis of the P1 residue.

Mutations of the P1 residue in **5.2.** all essentially abolished cleavage at the NS5A/NS5B site with only replacement of the wild type cysteine with serine permitting a small level of processing (approximately 10% of wild type efficiency). The amino acids introduced at P1 are illustrated in Fig. 8.3.

For the NS3/NS4A cleavage site, substitution of the wild type threonine residue with cysteine, serine, glycine, alanine or leucine still resulted in cleavage of the substrate (Leinbach *et al.*, 1994). When arginine or tyrosine were introduced in this experiment there was abolition of cleavage at this site. These results imply that amino acids with polar or aliphatic side chains of different lengths are tolerated at P1 whereas basic or aromatic side chains are not. These results may only be relevant to cleavage *in cis*.

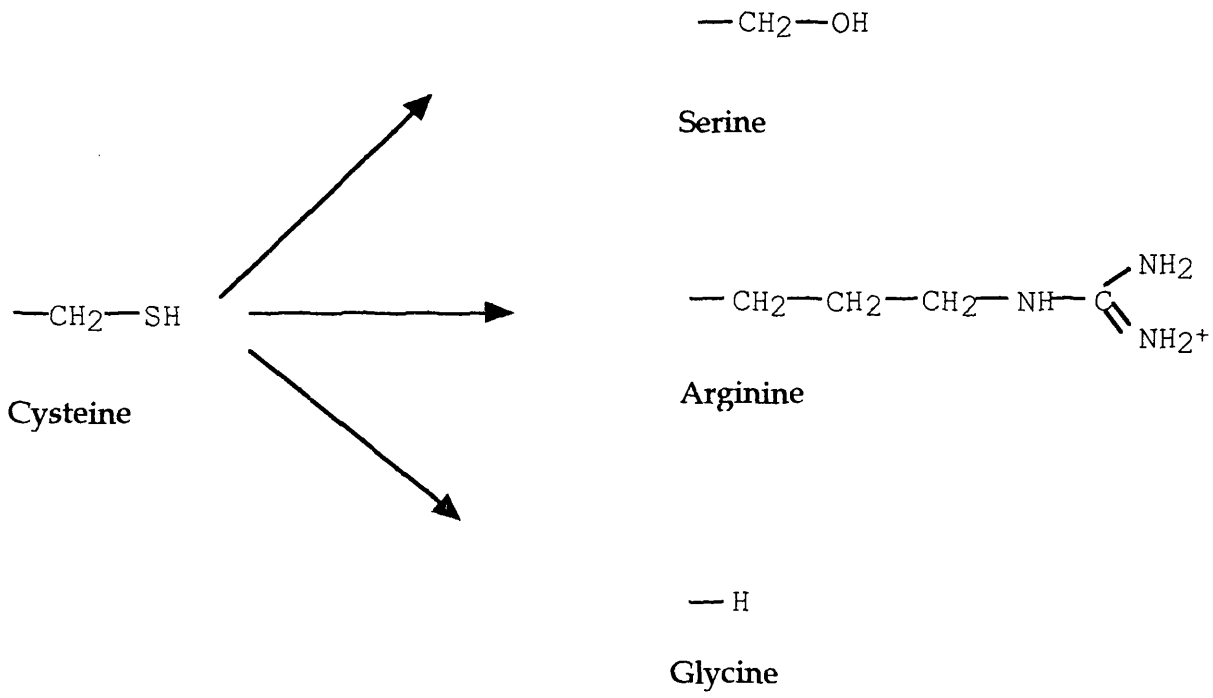


Fig. 8.3. Mutagenesis of P1. The wild type P1 residue (Cysteine) is illustrated on the right. When the P1 residue was replaced with serine a similar side chain is introduced with the sulphydryl group replaced by the hydroxyl group. Substitution of the P1 residue with arginine introduces a significantly larger, positively charged residue. Replacement with glycine introduces a single hydrogen atom as a side chain.

Kolykhalov *et al.* (1994) demonstrated that, at the NS4A/NS4B cleavage site, mutation of the P1 residue (cysteine) to threonine or leucine was tolerated whereas substitution with arginine rendered the site almost completely resistant to processing. Of the sites cleaved, the site containing an aspartic acid at P1 was cleaved with the lowest efficiency while still permitting some cleavage to occur. A gradual decrease in the inhibitory effect of P1 residue substitution was observed with asparagine followed by glycine and finally serine. At the NS3/NS4A cleavage site only mutation of the P1 threonine to arginine resulted in significant inhibition of processing with the other mutations introduced (glycine, asparagine and aspartic acid) having little effect. These observations are in agreement with the results obtained by Leinbach *et al.* (1994).

At the NS4B/NS5A cleavage site mutation of the P1 cysteine residue to glycine, arginine or aspartic acid resulted in abolition of cleavage whereas replacement with asparagine dramatically reduced cleavage. At the NS5A/NS5B cleavage site none of the mutations introduced at P1 (glycine, arginine, asparagine and aspartic acid) permitted cleavage of the substrate (Kolykhalov *et al.*, 1994). A number of P1 mutations were introduced into an NS4A/NS4B cleavage site (threonine, alanine, proline, phenylalanine, serine) with only the presence of a threonine residue at P1 resulting in detectable cleavage. Although the threonine residue at P1 permitted cleavage at this site the efficiency of processing was reduced by a factor of 140. A cysteine to serine substitution at P1 of the NS5A/NS5B cleavage site formed a substrate that was very inefficiently processed (Zhang *et al.*, 1997). The documented mutagenesis of P1 agrees with the data reported in **5.2.** as only substitution with a serine residue at P1 produced a mutated NS5A/NS5B cleavage site that was processed by NS3.

All these results illustrate the importance of the P1 residue for all the cleavage sites processed by the HCV NS3 with the NS3/NS4A cleavage site the most tolerant to mutation. This feature is probably due to the in *cis* nature by which this site is processed. For all the cleavage sites processed in *trans*, mutation of P1 is not well tolerated. The P1 residue was demonstrated to interact with a specific binding pocket (the S1 specificity pocket) within the NS3 protease. This specificity pocket, which is shallow and nonpolar, is formed primarily by the side chains of Phe-154, Ala-157 and Leu-135 (Love *et al.*, 1996). The amino acid Phe-154 determines the shallow nature of the specificity pocket. The side chain of the P1 residue in the substrate (SH although substitution with an OH group is tolerated) will interact with the electron-rich π clouds on the aromatic ring of Phe-154. This observation explains the specific nature of the residue found at P1 and the tolerance for a serine residue only at this position.

8.2.3. Mutagenesis of the P1′ residue.

The effect of mutation of the P1′ residue depended largely on the amino acid introduced (**5.2.**). Replacement of the wild type P1′ serine

residue with threonine resulted in a fall in processing by 60-70%. Substitution with alanine gave a fall in processing of approximately 40% and introduction of a proline residue completely abolished any detectable processing. These amino acid substitutions are illustrated in Fig. 8.4.

At the NS3/NS4A cleavage site the basic amino acid arginine, the polar amino acids serine, cysteine, and threonine, and the aliphatic amino acids glycine, alanine, and leucine were all tolerated at P1′ (Leinbach *et al.*, 1994) producing cleavage sites that were processed by NS3. Kolykhalov *et al.* (1994) demonstrated that at the NS3/NS4A cleavage site only the presence of proline at P1′ produced a cleavage site that was processed at a significantly lower level compared to the wild type NS3/NS4A cleavage site. Substitution of the P1′ residue with an arginine residue had little effect on the efficiency with which this cleavage site was processed.

At the NS4A/NS4B cleavage site P1′ substitutions with isoleucine, threonine, arginine, alanine, asparagine and histidine all gave cleavage sites that were efficiently processed whereas the presence of proline dramatically reduced the level of processing (Kolykhalov *et al.*, 1994). At the NS4B/NS5A and the NS5A/NS5B cleavage sites the presence of proline at P1′ abolished cleavage and a cleavage site containing arginine at exhibited a slightly lower level of processing (Kolykhalov *et al.*, 1994). Urbani *et al.* (1997) analysed an NS4A/NS4B cleavage site that contained a naturally occuring alanine at P1′. When this residue was substituted with a serine residue, as observed at this position in other sites, there was a fall in processing of approximately 2-5 fold. This result implies that the optimal favoured amino acid at P1′ is an alanine. Replacement of the P1′ serine residue at the NS5A/NS5B cleavage site with alanine had little effect on the efficiency of processing (Zhang *et al.*, 1997). All these results imply that the effect of an amino acid substitution at the P1′ residue is dependent on the amino acid introduced. That the introduction of an alanine residue at P1′ has little effect on the efficiency by which the cleavage site is processed is understandable as some HCV NS4A/NS4B cleavage sites naturally contain an alanine at P1′.

The results of these mutational analyses show that the most important residue in cleavage site determination is the P1 amino acid. The acidic residue at P6 appears to play an important role in the cleavage sites processed *in trans* within the HCV polyprotein. Mutation of the P6 residue, although not resulting in the abolition of cleavage, leads to a significant increase in the amount of uncleaved substrate. The only exception to this observation is seen for the NS3/NS4A cleavage site, which is processed *in cis*. The P1′ serine residue is not essential for processing of the cleavage sites with alanine present at other sites and in other genotypes. Only the introduction of a proline residue at this position abolishes cleavage by the NS3 serine protease. The P1′ residue was tolerant to the introduction of other amino acids which all permitted cleavage, albeit with different processing efficiencies.

The structural prediction program "Plotstructure" was used to carry out comparisons between the wild type tNS5T substrate and the mutated tNS5T substrates, to determine whether any amino acid substitution could be predicted to have a negative effect on processing of the substrate. In Fig.

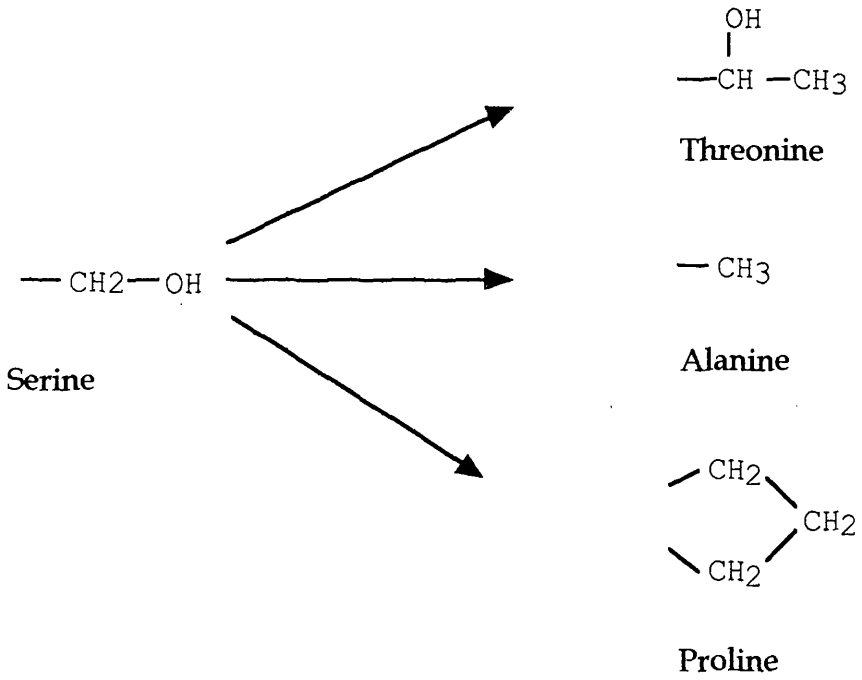


Fig. 8.4. Mutagenesis of P1′. The wild type P1′ residue (serine) is illustrated on the right. Substitution with threonine introduces a side chain that retains the hydroxyl group but that has a slightly different organisation. The alanine substitution introduces a shorter, non-polar side chain. The introduction of the imino acid proline can have a significant effect on the peptide backbone.

8.5. it can be seen that there is no difference in the predicted cleavage site structure for either the wild type substrate or the substrates containing mutations of P6. As none of these mutations abolished processing the predictions from the structural analyses seem to corroborate observed results. There were significant differences in the predicted secondary structure for the P1 serine and glycine mutations in comparison with the wild type substrate and the substrate containing arginine at P1. The "Plotstructures" of the P1′ substrates agreed with the observed cleavage patterns, with the uncleaved P1′ Pro substrate having a significantly different predicted structure or the wild type and the two processed substrates exhibiting no discernible difference. It can not be determined for certain whether a substrate would be processed by NS3 using this "Plotstructure" program. When compared to the wild type substrate, the majority of substrates that have the same predicted structure as the wild type were processed. The majority of substrates whose predicted secondary structures differed from the wild type were not cleaved by NS3.

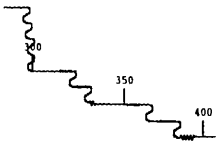
8.3. Investigation of cleavage site specificity.

There have been many investigations into the importance of the P6, P1 and P1′ residues of the HCV cleavage site motif, but little analysis of the additional residues present within the cleavage site sequence has been undertaken. It was demonstrated that when an alanine residue was introduced at P3 of the NS4A/NS4B cleavage site the level of processing fell by 7-fold (Urbani *et al.*, 1997). This effect on processing was observed in the absence of NS4A. In the presence of NS4A this cleavage site was processed at a similar level to substrates containing other amino acid substitutions (i.e. P5-P2, P2′ and P3′) shown to be unimportant in substrate specificity. It was also demonstrated that there was a preference for a hydrophobic residue at P4′ although the effect of amino acid substitution at this position was not investigated. Sequence comparison between different genotypes confirms that there is a hydrophobic residue at P4′ for all HCV isolates.

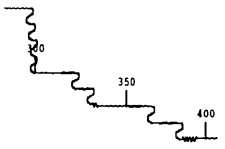
Kolykhalov *et al.* (1994) mutated P7 (Phe to His or Arg), P5 (Glu to Lys), P3 (Glu to Asn or Lys), P2 (Glu to Asn or Lys) and P2′ (Gln to Lys) of the HCV-H NS4A/NS4B cleavage site with no observable effect on cleavage. No other investigation into cleavage site specificity has been done.

8.3.1. Investigation of potential cleavage sites within the HCV Gla polyprotein.

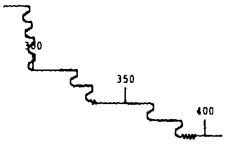
If the amino acid sequence required to define an NS3-mediated cleavage site involved only the P6, P1 and P1′ residues, it would be unlikely that any other similar sequence existed within the HCV polyprotein except the documented cleavage sites. The amino acid sequence of the HCV Gla



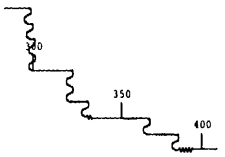
Wild Type



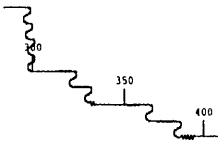
P6 Gln



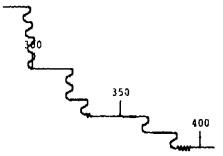
P6 Lys



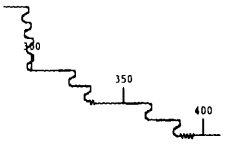
P1 Ser



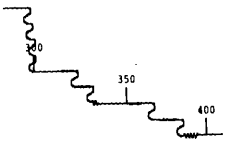
P1 Arg



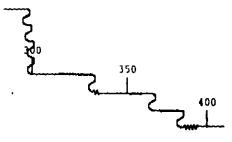
P1Gly



P1′ Thr



P1′ Ala



P1′ Pro

Fig. 8.5. Predicted secondary structures of the NS5A/NS5B cleavage sites containing point mutations at specific positions within the cleavage site motif. For all the substrates the rest of the predicted secondary structure was unchanged. The NS5A/NS5B cleavage site is processed at amino acid 315 of the tNS5T substrate. For both the P6 substitutions the predicted secondary structure is unchanged. Both the P1 Ser and P1 Gly substrates have different predicted secondary structures to that of wild type, whereas the P1 Arg substrate appears to be identical. The only P1′ mutant substrate whose predicted secondary structure is different to the wild type is that containing′ Pro at P1′.

polyprotein was searched for any D/E X X X X C/T S motif and this search yielded three additional, potential cleavage sites. These potential sites all contained the "motif" shown to give a site suitable for processing by the NS3 protease. The first of these cleavage sites, present in NS5A, was not processed by the NS3 serine protease when either the full length or truncated NS5A proteins were cotransfected with NS34A (**6.2.**). The lack of processing observed at this potential site may be due to the folding of the NS5A protein, making the potential site inaccessible to the NS3 protease. Further analysis would be required to determine whether this site is either unsuitable for cleavage due to the amino acids surrounding P6, P1 and P1' or that it is merely inaccessible to the NS3 protease due to the folding of the substrate used.

Investigations into this site could be carried out in a number of ways. As this site resembles the NS3/NS4A cleavage site, the seven amino acid sequence (DVAVLTS) could be substituted for the wild type NS3/NS4A cleavage site (DLEVVTS) within the HCV Gla NS34A protein. This method of mutagenesis would be similar to that used for the analysis of the LITS potential cleavage site (**6.3.**) within the tNS5T substrate. Sequence comparison between the two sites at the primary amino acid level reveals little dissimilarity. Except for the acidic residue aspartic acid, at P4 in the NS3/NS4A cleavage site and the alanine at P4 within the potential site there is little difference in the amino acids present. There is, however, little sequence variation between the NS3/NS4A cleavage sites of different isolates. The aspartic acid at P4 is completely conserved as are the amino acid residues at P6, P5, P3, P1 and P1'. The P2 residue is primarily a valine although threonine and methionine have been reported at this position. The conservation of the acidic residue at P5 may be important for efficient cleavage *in trans*. As the NS3/NS4A cleavage site was extremely tolerant to mutations at P6, P1 and P1' it is unlikely that mutation of the P4 residue would have any effect on processing.

An alternative analysis could be carried out on the cleavage site as it is found naturally. To determine whether this site could be processed *in trans* a substrate based around the NS5A protein could be used. It has been demonstrated that this site was not processed as part of the full length NS5A or the NS5A substrate truncated by 132 amino acids (**6.2.**) at its N-terminus. In the context of the truncated NS5A substrate the potential cleavage site was situated 68 amino acids from the N-terminus and 248 amino acids from the C-terminus. A series of further truncations from either terminus may produce an NS5A substrate that could be cleaved at this potential site. This method may not provide any further information if the substrate is not processed. A lack of cleavage at this site could still be due to the structural conformation of the NS5A protein. To determine the reason for any lack of processing, a site cleavable by NS3 (e.g. NS5A/NS5B) could be introduced in place of the theoretical cleavage site. Truncations of the substrate termini could be carried out in tandem for the wild type and "cleavable" NS5A substrates. Any lack of processing for the NS5A substrate containing a cleavable site would be due to the structural conformation of the protein

and not the amino acid sequence of the cleavage site. This potential site within the NS5A protein was also present in many other HCV isolates.

The two potential cleavage sites in NS5B are present in all the HCV isolates and are highly conserved. The ELITSCS potential cleavage site is completely conserved between different HCV isolates. The DLELITS potential site is highly conserved between isolates with only one isolate differing at P5 with the leucine residue replaced by a glutamine residue.

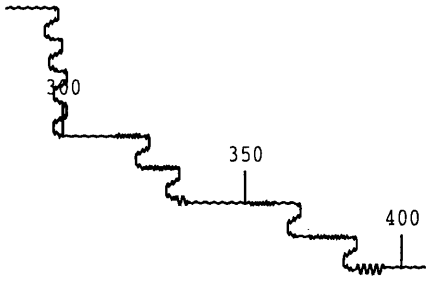
No analyses were carried out on the DLELITS cleavage site in this study. To investigate whether this potential site could be processed it could be analysed in a similar manner to the potential site in NS5A. The cleavage site could replace the NS3/NS4A cleavage site within the NS34A protein. The DLELITS sequence is even more similar to the NS3/NS4A site than the previous potential site. The only difference between the potential site and the wild type NS3/NS4A site (DLEVVTS) is the leucine and isoleucine at P3 and P2 instead of the pair of valines. All three of these amino acids are similar and it would be expected that this cleavage site would be processed if it was expressed in this manner.

Data were not available on whether this site was processed within the wild type NS5B. The NS5B protein could not be identified within cells as no antisera against this protein were available. The tNS5T substrate could not be used as the C-terminal truncation includes the potential site. Production of truncated NS5B substrates (with an epitope tag for identification) may provide an insight into whether this site could be processed *in trans* within the NS5B protein.

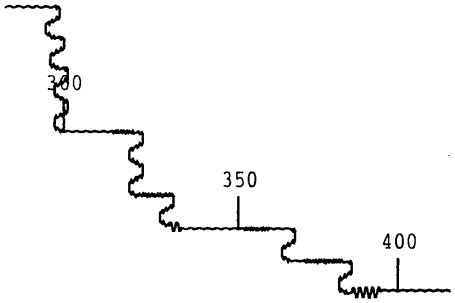
These two potential cleavage sites were not analysed for their cleavability but potential tNS5T substrates containing these two cleavage sites were analysed using the "Plotstructure" program (Fig. 8.6.). The predicted structure of the tNS5T substrate containing the potential site from NS5A was different to the wild type tNS5T prediction. The tNS5T substrates containing the potential site from NS5B had a predicted structure identical to the wild type tNS5T. Whether these substrates are actually processed is not yet known but from these predictions one would be processed (NS5B potential site) and the other would not (NS5A potential site).

8.3.2. Investigation of a potential cleavage site within the NS5B protein.

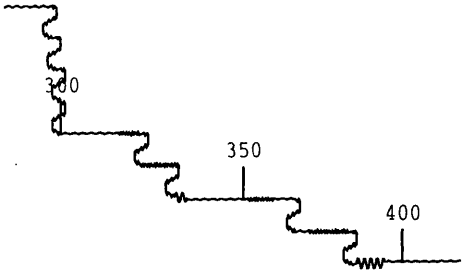
The potential cleavage site (ELITSCS) within NS5B was not studied at its normal position within the NS5B protein as this was downstream of the C-terminus of the tNS5T substrate. The use of a full length NS5 substrate would be of no use as identification of the full length NS5B was not possible. Even a full length NS5B protein with an epitope tag attached may not be of any use as the NS5B is reported to be unstable in mammalian cells (Lin *et al.*, 1994b). As this potential site resembles the cleavage sites processed *in trans* by the NS3 protease it was decided to replace the NS5A/NS5B cleavage site within the tNS5T substrate with the potential cleavage site (ELITSCS). This new tNS5T substrate, containing the amino



Wild Type



Potential NS5A site



Potential NS5B site

Fig. 8.6. Comparison between the predicted secondary structures of the wild type tNS5T substrate and tNS5T substrates containing potential cleavage sites from within the HCV Gla. polyprotein. The potential cleavage site within NS5A (DVAVLTS) has a predicted secondary structure different to the wild type tNS5T whereas the potential site within NS5B (DLELITS) has a predicted secondary structure identical to the wild type substrate. From these secondary structure predictions it could be hypothesised that the potential NS5A cleavage site would not be processed within tNS5T whereas the potential NS5B cleavage site would be processed.

acid sequence LITS from P5-P2 at the NS5A/NS5B cleavage site, was cotransfected with NS34A. Coexpression of this substrate with the NS34A protease complex did not yield any observable cleavage products. Quantitation using the phosphorimager demonstrated only a small degree of processing but this value was so small as to be negligible. This absence of processing is in contrast to the reported data that only P6, P1 and P1′ are required to define a cleavage site that can be processed by the NS3 protease.

This potential cleavage site contains the required P6, P1 and P1′ residues for correct interaction with NS3 protease and cleavage *in trans*. The lack of processing observed must be due to influence exerted by some or all of the amino acids from P5-P2. Sequence comparison of cleavage sites processed by the NS3 protease was carried out for all six genotypes to determine if any of these amino acids were present in cleavage sites processed by the NS3 protease. A leucine at P5 is found in cleavage sites of other HCV isolates although only in the highly conserved NS3/NS4A cleavage site. There does not seem to be any conservation of the type of amino acid present at this position. Non-polar side chains (Cys, Gly, Leu), uncharged polar side chains (Ser, Tyr) and negatively charged, acidic side chains (Asp, Glu) are all found at this position in different cleavage sites and in different HCV genotypes.

At P4 an isoleucine residue has been reported, as have other non-polar residues (Met, Pro, Val), uncharged polar residues (Ser, Thr) and acidic residues (Glu). At P3 non-polar residues (Val, Iso, Ala), charged polar residues (Thr, Ser) and an acidic residues (Glu) are found.

The serine residue found at P2 of this potential cleavage site is not observed for any other site processed by the NS3 protease. An acidic residue (Glu) or a non-polar residue (Pro, Cys, Val) is observed but no HCV isolate contains a polar residues at this position, let alone a serine residue. As no serine residue was observed at this position it was hypothesised that this amino acid could be primarily responsible for the lack of processing observed for this potential cleavage site. Although it was not investigated further, the possibility that this serine residue may be phosphorylated and interferes with processing was considered.

An additional serine residue was also present at P2′ of this potential cleavage site. A serine residue was not documented at this position when a serine was also present at P1′, however a serine residue was present when there was an alanine residue at P1′. The effect of introducing a serine residue at P2′ was also studied within the NS5A/NS5B cleavage site in tNS5T.

To investigate the importance of the serine residue at P2 two potential cleavage sites were constructed. The first cleavage site constructed was essentially the wild type NS5A/NS5B cleavage site with the P2 residue substituted to a serine residue from the wild type cysteine. The second cleavage site constructed was derived from the LITS potential site with the P2 residue restored to the wild type cysteine to give a P5-P2 sequence of LITC. Both substrates were analysed for their susceptibility to cleavage by the NS34A protease complex (**6.3.**, Table 6.3.). The introduction of a serine residue at P2 of the wild type cleavage site had a significant effect on the

efficiency by which the substrate was processed. Although the efficiency of cleavage did not fall to the levels observed for the LITS cleavage site the drop in the degree of cleavage was still significant (~50%). This result implies that although the serine residue is a contributing factor to the fall in the efficiency of cleavage it is not the sole factor.

This observation is confirmed by the second potential cleavage site constructed to investigate the importance of a serine residue at P2. The substrate containing LITC was processed at an observable level, a significant increase in the level of processing compared to the cleavage site containing LITS. The level of processing was still significantly lower than that of the wild type NS5A/NS5B cleavage site (~25%).

The introduction of a serine residue at P2' of the wild type cleavage site produced a tNS5T substrate that was still processed by the NS34A complex. The efficiency by which this substrate was cleaved was lower (60%) than that observed for the tNS5T substrate containing the wild type NS5A/NS5B cleavage site. All these mutated substrates demonstrate the effect of introducing an additional serine residue in close proximity to the site of cleavage, although a serine residue at P2 is not the only factor involved in the almost complete abolition of processing observed for the LITS potential cleavage site.

To characterise further the reason for the minimal processing observed for the LITS cleavage site three further potential cleavage sites were introduced into tNS5T substrate in place of the wild type NS5A/NS5B cleavage site. These three potential sites were derived from the LITS sequence with pairs of amino acids replaced by the wild type residues, thus giving DVTS, LVVS and LIVC at P5-P2. These substrates were cotransfected with the NS34A protease construct and the efficiency of processing by the protease determined (**6.3.**, Table 6.4.)

When the amino acid sequence LITC was present at P5-P2 in the NS5A/NS5B cleavage site, processing was at an efficiency of 20-25% with respect to the wild type cleavage site. When this potential cleavage site was changed to contain LIVC at P5 to P2, the increase in the efficiency of processing was marginal, rising to 25-35% of the wild type efficiency. The increase in the efficiency of processing is attributable to the loss of the threonine residue at P3 as this is the only amino acid change between these two theoretical cleavage sites. It can be speculated that the presence of threonine at P3 has a small inhibitory effect on the cleavability of the site. Both serine and threonine have a polar hydroxyl group (OH) and the combined presence of these two similar amino acids may contribute to a fall in the efficiency of processing. These two factors are not alone in being responsible for the low level of processing observed when the LITS sequence is present from P5-P2 within the cleavage site. The efficiency with which the LIVC cleavage site is cleaved was still low in comparison to the wild type NS5A/NS5B cleavage site.

The threonine at P3 was also replaced when the two valine residues found at the wild type NS5A/NS5B cleavage site were introduced at P4 and P3 (giving the sequence LVVS at P5-P2). The substitution of these two amino acids resulted in an increase in the efficiency of processing with

respect to the LITS cleavage site by approximately 10%. As the isoleucine to valine substitution involves the replacement of one non-polar residue with another, the increase in the efficiency by which this substrate was processed is probably due to the loss of the threonine residue at P3. This increase agrees with the observation that the difference in the efficiency of processing between LITC and LIVC is also approximately 10%. Both these substitutions (LITC to LIVC and LITS to LVVS) involve the replacement of the threonine residue at P3 to the wild type valine residue, resulting in a more efficiently cleaved site. The presence of the threonine residue at P3 appears to have a definite but small negative effect on the efficiency by which the substrate is processed.

When the potential cleavage site containing LITS was replaced at P5 and P4 to give the wild type aspartic acid and valine respectively (to give DVTS at P5-P2), the increase in the efficiency of processing from that observed for the cleavage site containing the LITS sequence was negligible. Despite the introduction of the wild type acidic residue (Asp) at P5 there was barely any increase in the cleavability of this substrate. The replacement of isoleucine with valine would be expected to have little effect as both are non-polar residues of a similar size. The retention of the inhibitory serine residue at P2 may be the dominant negative factor and any positive influence on processing exerted by the acidic residue at P5 may be insignificant while the serine is present at P2. The serine side chain and the aspartic acid side chain together have the capability of forming a hydrogen bond (Fig. 8.7.).

The spatial arrangement of amino acids within this cleavage site has not been determined. It is not known whether these two side chains are in close enough proximity to interact and hence form a hydrogen bond, but if they are close enough the positive influence normally exerted by the acidic residue at P5 may be inhibited.

One further potential cleavage site was constructed with only the P5 leucine residue mutated from the wild type sequence. This cleavage site contained LVVC from P5 to P2 and was cotransfected with the NS34A construct before the efficiency with which this substrate was processed was calculated (**6.3.**, Table 6.5.). This cleavage site was processed with a slightly higher efficiency than the LVVS cleavage site although the increase was only approximately 5%. This cleavage site was processed at an efficiency of 15-25% of the wild type efficiency. The increase in efficiency corresponds to the loss of the serine residue at P2. The effect of replacing the serine at P2 was observed when the LITS cleavage site was mutated to LITC, with a similar rise in the efficiency of processing observed.

The "Plotstructures of all these LITS-derived cleavage sites were produced. All eight of these predicted secondary structures had the additional "squiggle" seen for the P1 Ser and P1 Gly substrates that were inefficiently processed by the NS3 serine protease (Fig. 8.8.)

These potential cleavage sites were all derived from the LITS sequence due to the effect on processing of the NS5A/NS5B cleavage site when this sequence was introduced from P5-P2. The presence of the serine residue at P2, and to a lesser extent the presence of threonine at P3, have

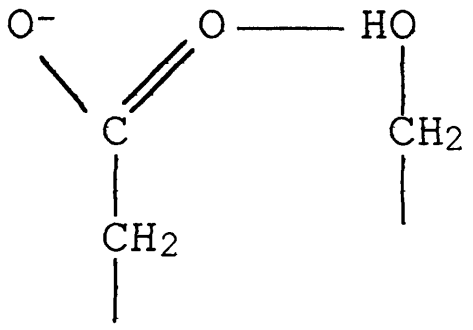


Fig. 8.7. Hydrogen bonding between two amino acid side chains. The negatively charged, acidic aspartic acid (left) can form a hydrogen bond (illustrated by the broken line) to the uncharged, polar side chain of the serine residue.

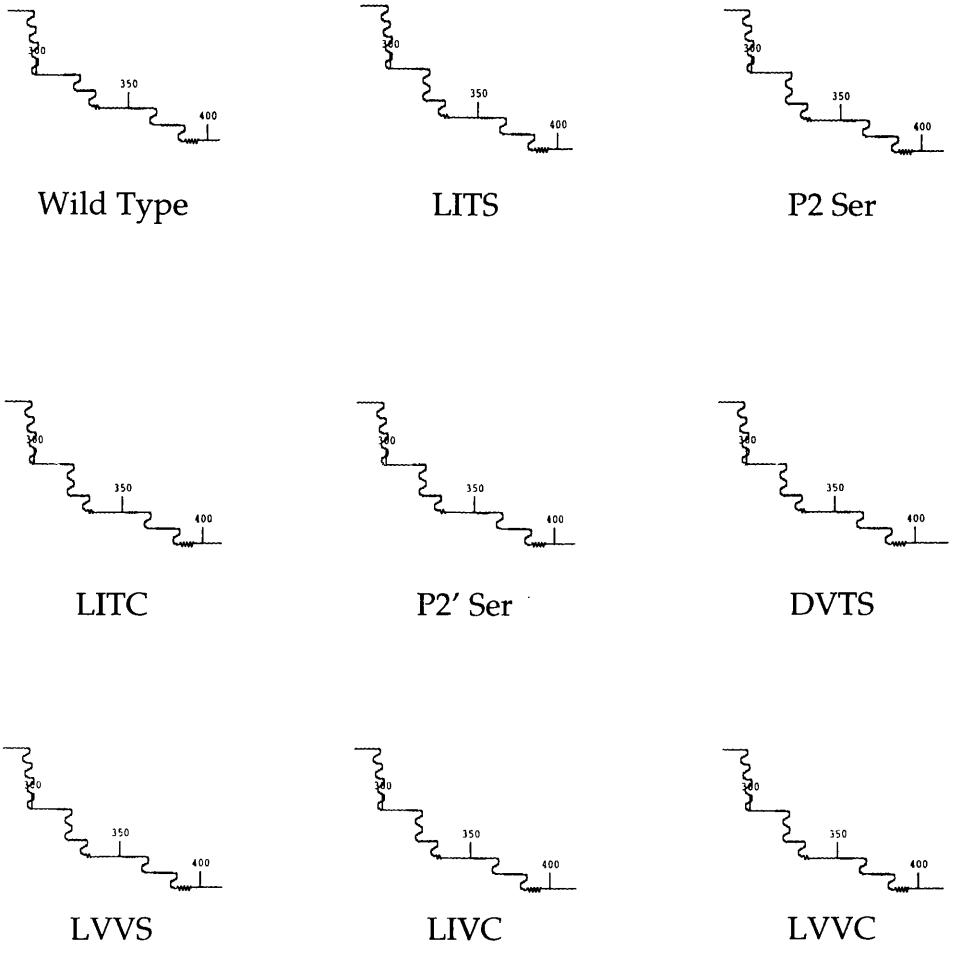


Fig. 8.8. Comparison of predicted secondary structures between the potential LITS cleavage site and its derivatives. The NS5A/NS5B cleavage site is processed at amino acid 315 of the tNS5T substrate. The predicted secondary structures for the cleavage sites containing the LITS and LITS-derived potential cleavage sites are significantly different to the wild type.

been shown to have a significant role in the fall in cleavage efficiency when they are introduced. Despite replacing these residues the two substrates (LIVC and LVVC) were processed with an efficiency 65-75% lower than the wild type substrate. Replacement of isoleucine at P4 with valine had little effect. This observation was not unexpected as the two residues are both non-polar. These analyses lead to the hypothesis that an acidic residue at P5 (aspartic acid for the NS5A/NS5B cleavage site) has a significant role in the substrate specificity of a cleavage site processed by the NS3 serine protease. The only cleavage site (derived from the LITS potential site) constructed with an aspartic acid at P5 also contained a serine residue at P2 and was not processed efficiently. The serine residue may inhibit the positive influence of the acidic residue at P5 on substrate specificity.

As well as the conserved residues at P6, P1 and P1', the NS5A/NS5B cleavage site is processed with a high efficiency apparently due to the presence of a positively charged, acidic residue at P5. Substitution of alanine at P5 of the NS5A/NS5B site resulted in a fall in the catalytic efficiency by a factor of 9 (Zhang *et al.*, 1997). This observation agrees with the conclusion that the acidic residue at P5 is important in the specificity of cleavage by the NS3 serine protease.

All these substrates derived from the LITS potential cleavage site were

Sequence analysis of NS5A/NS5B cleavage sites from different HCV isolates gives the following consensus sequence:

6	5	4	3	2	1	1'	2'	3'	4'
E(D)	D(S, G)	V	V(I, A)	C	C	/	S	M	S	Y

Although this sequence seems to illustrate a tolerance for amino acid changes only very few isolates exhibit sequence variation. Two type 2 isolates (HCV-J6 and HCV-J8) have an aspartic acid at P6 and a serine at P5 with one of these having an isoleucine at P3 (HCV-J8). A type 1b isolate (HCV-T) has a glycine at P5 and an isoleucine at P3. Another type 1b isolate (HCV-JK1) has an alanine at P3. All the remaining 14 isolates from genotypes 1 to 4 had the conserved EDVVCCSMSY sequence. The loss of the acidic residue at P5 has a significant effect on processing but the site is still processed. For the three isolates containing different amino acids at this position the NS5A/NS5B cleavage sites would still be processed, albeit with a lower level of efficiency than for cleavage sites containing the acidic P5 residue.

8.4. Investigation of processing at different sites within the HCV polyprotein.

An analysis of NS3 mediated cleavage at other sites in the HCV polyprotein was desirable to investigate the specificity of cleavage. Due to the lack of antisera against the other non-structural proteins within the

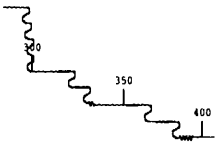
laboratory the processing of these other cleavage sites was not carried out at their normal positions within the HCV polyprotein. The cleavage sites processed within the HCV polyprotein (NS3/NS4A, NS4A/NS4B and NS4B/NS5A) were introduced into the tNS5T substrate in place of the NS5A/NS5B cleavage site. These substrates were cotransfected with the NS34A construct and the efficiency by which they were processed calculated. The result of this analysis is summarised below, as are the amino acid sequences of the cleavage sites introduced. The NS3/NS4A cleavage site was not processed *in trans* when expressed in this manner and the other sites were processed with varying efficiencies.

Cleavage Site	Sequence	Relative % Cleavage Efficiency
NS3/NS4A	DLEVVTS	0
NS4A/NS4B	DEMEECS	45-50
NS4B/NS5A	DCTTPCS	20-25
NS5A/NS5B	EDVVCCS	100

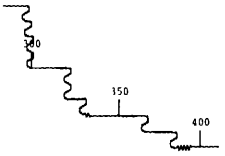
Table 8.1. Summary of HCV Gla cleavage sites processed by the NS3 serine protease. The amino acid sequence of the cleavage sites is illustrated as is the efficiency by which these cleavage sites are processed when they replaced the NS5A/NS5B site within tNS5T.

The "Plotstructures" of these substrates (including an NS4B/NS5A substrate containing aspartic acid at P6) were produced and all four predicted structures were different from the predicted wild type structure with the exception of the NS4A/NS4B structure (Fig. 8.9.). An interesting feature was observed for the NS3/NS4A cleavage site (which was not processed within the assay) compared to the potential cleavage site in NS5B (DLELITS). Both these cleavage site motifs are very similar with the only difference being the VV-LI change at P3 and P2. All three of these amino acids are non-polar and it would not be expected that replacement from one pair of amino acids to another would have a significant effect on secondary structure. The predicted structure for the potential NS5B site in tNS5T appears to be a structure susceptible to cleavage by the NS3 serine protease, however the NS3/NS4A tNS5T substrate has a significantly different predicted secondary structure from the wild type tNS5T substrate.

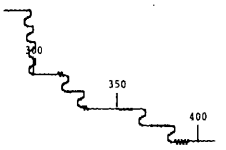
The most efficiently processed cleavage site (besides the wild type NS5A/NS5B site) was the NS4A/NS4B site, which was processed at an efficiency of 50% compared to the wild type site. Comparison between the amino acid sequence of both sites shows an acidic residue at P6 and P5 for both these cleavage sites. It had been speculated above that the presence of an acidic residue at P5 influenced the efficiency with which a site is



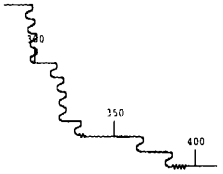
Wild Type



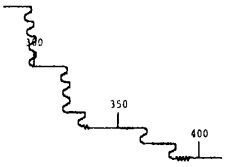
NS3/NS4A



NS4A/NS4B



NS4B/NS5A



NS4B/NS5A(P6 Asp)

Fig. 8.9. Comparison between predicted secondary structures of tNS5T substrates containing the different HCV NS3-mediated cleavage sites. With the exception of the NS4A/NS4B tNS5T substrate the predicted secondary structures are significantly different to the wild type tNS5T substrate.

processed. The rest of the amino acid sequence within the cleavage site motif is significantly different between the two sites. The NS5A/NS5B cleavage site has Val, Val, Cys at P4 to P2 whereas the NS4A/NS4B site contains Met, Glu, Glu at these positions. The NS5A/NS5B sequence contains three non-polar residues. The NS4A/NS4B sequence consists of a non-polar amino acid at P4 followed by two acidic residues. The presence of an acidic chain at P5 may be as important for efficient cleavage as the acidic residue at P6. As they are highly conserved, the remaining amino acids may also be important for efficient cleavage at this site. The introduction of the bulkier methionine residue at P4 instead of valine and the acidic residues at the other positions may have an inhibitory effect on the specificity of cleavage.

The NS4B/NS5A cleavage site is processed inefficiently (25%) in comparison to the NS5A/NS5B cleavage site. The NS4B/NS5A sequence does not have an acidic residue at P5. The absence of this residue was shown previously to produce a substrate processed at 15-25% compared to the wild type. This low efficiency of processing corroborates the results obtained for the NS4B/NS5A cleavage site. As there is little difference in efficiency between the LVVC NS5A/NS5B cleavage site and the NS4B/NS5A cleavage site the remaining amino acid differences are likely not to have any significant influence on substrate specificity. The two valine residues at P4 and P3 and cysteine at P2 are replaced by Thr, Thr, Pro. A valine to threonine substitution changes the nature of the amino acid from non-polar to uncharged polar although the sizes of the side chains are relatively similar with the valine side chain slightly larger. If these residues are not significantly important in cleavage site specificity then it is not surprising that these substitutions have little effect. The substitution of cysteine for proline, although not affecting the nature of the amino acid, may have an effect on the peptide backbone but as no significant effect on processing is observed the effect, if any, is likely to be negligible.

A cleavage site derived from the NS4B/NS5A site, containing a glutamic acid at P6 as opposed to an aspartic acid, was also introduced in place of the NS5A/NS5B cleavage site within the tNS5T substrate. This cleavage site was processed with a similar efficiency as the wild type NS4B/NS5A cleavage site, illustrating that the important feature at P6 is a negatively charged, acidic residue.

The NS3/NS4A cleavage site was not processed by the NS3 serine protease when it replaced the NS5A/NS5B cleavage site within tNS5T. As well as not having the acidic residue at P5, the NS3/NS4A cleavage site contains a threonine residue at P1 instead of the cysteine residue found for all cleavage sites processed *in trans*. The combination of these factors is probably enough to produce a cleavage site unsuitable for processing by the NS34A protease complex. As the NS3/NS4A site is normally processed *in cis* these factors are not as important. The threonine side chain is bulkier than the serine side chain and will not fit into the S1 specificity pocket of the NS3 protease as ideally as serine, despite having the correct sulphydryl group to interact with Phe-154 of the protease. The only other significant difference in the amino acids present is the acidic side chain at P4 as opposed to the

valine residue found in the NS5A/NS5B cleavage site. When processed *in cis*, the NS3/NS4A substrate is attached to the protease, which folds around the substrate and cleaves it despite the imperfect sequence of the cleavage site. It has been reported that the NS3/NS4A cleavage site can be processed in *trans* normally when the NS3 protease domain is cleaved form the N-terminus of the NS34A substrate (Lin *et al.*, 1994b). These data are in contrast to the results here. The reasons for this discrepancy are unclear but the folding of the NS34A substrate may overcome the negative influences exerted by the amino acid sequence of the cleavage site and permit cleavage in *trans*.

An investigation of this processing could be carried out by developing an truncated NS34A substrate and replacing the NS3/NS4A cleavage site with a cleavage site normally processed *in trans*. Comparison of the wild type NS34A substrate and this altered NS34A (NS5A/NS5B) substrate would perhaps illustrate the difference in efficiency obtained at this site. If the NS3/NS4A site is processed *in trans* within this substrate it may still be cleaved at a significantly lower level than a cleavage site normally processed *in trans*. Alternatively the reason that the NS3/NS4A cleavage site is processed *in trans* within the NS34A substrate may be entirely due to the amino acid sequence at the cleavage site within the context of the folded protein. A cleavage site processed normally *in trans* may not be efficiently processed within the NS34A substrate.

The introduction of the HCV cleavage sites into the tNS5T substrate provides some information into the specificity of cleavage by the NS3 protease. This analysis also seems to confirm the importance of the acidic residue at P5 for efficient cleavage *in trans*. For efficient processing at the NS5A/NS5B cleavage site it was hypothesised that an acidic residue at P5 was important. This deduction was the result of the analyses carried out with respect to the LITS potential cleavage site.

When the NS5A/NS5B cleavage site was replaced by the cleavage sites normally processed within the HCV polyprotein (NS3/NS4A, NS4A/NS4B, NS4B/NS5A) a significant variation in the efficiency with which these sites were processed was observed. Analysis of the amino acid sequences of these cleavage sites illustrated that the presence of an acidic residue at P5 was important for efficient processing *in trans* by the NS3 serine protease. The most efficient site cleaved by the NS3 protease contained an acidic residue at P5 whereas those sites that did not contain an acidic P5 residue were inefficiently processed (if at all) by NS3. These cleavage sites were all processed at a significantly lower level in comparison to the NS5A/NS5B cleavage site (in this analysis). If this analysis illustrates processing by the NS3 protease at its optimum, then the amino acid sequence, as well as the folding of the polyprotein, is important in determining the order in which the HCV non-structural proteins are released.

Analyses of processing at other cleavage sites within the HCV polyprotein have led to the deduction that these cleavage sites are less sensitive to mutation (Kolykhalov *et al.*, 1994). If these cleavage sites are processed with a lower efficiency within the HCV polyprotein due to the

lack of an acidic residue at P5 then any further mutation of this amino acid may not yield any information. Alternatively the introduction of an aspartic acid or a glutamic acid at this position may have a positive influence on the processing of these other sites at their normal positions. Substitution of the P5 residue within an NS4A/NS4B or an NS4B/NS5A substrate may reveal a significant increase in the efficiency of processing. Due to the nature of the cleavage at the NS3/NS4A cleavage site, mutation of the P5 residue within an NS34A substrate may not contribute any valuable information.

As well as an acidic residue at P5, the remaining amino acids within the NS5A/NS5B cleavage site have small, non-polar side chains (Val, Val, Cys). The NS4A/NS4B cleavage site was the site most efficiently processed next to the NS5A/NS5B cleavage site. This site contained an acidic residue at P5 but contained an additional two aspartic acid residues at P3 and P2. The only other cleavage site processed was the NS4B/NS5A cleavage site. This site did not have an acidic P5 and contained a proline residue at P2 that can have a significant effect on the secondary structure of the peptide backbone. The NS3/NS4A site, which was not processed *in trans*, had a leucine at P5, an aspartic acid at P4 and a threonine at P1.

The rapid release of the viral polymerase may be important in the viral replication cycle and so the NS5A/NS5B cleavage site is quickly and efficiently processed. The lesser importance of the remaining non-structural proteins or the importance in delaying the release of these proteins may be a factor in the lower substrate specificity of these cleavage sites.

8.5. Investigation of the protease requirements for processing by the NS3 serine protease.

Processing at NS3/NS4A does not require any other viral proteins and both NS3 and NS4A are required for the cleavage event to take place (Failla *et al.*, 1994). The cleavage event between NS3 and NS4A occurs *in cis*, rapidly and is the first NS3 mediated cleavage (Bartenschlager *et al.*, 1994) to take place during polyprotein processing. Protease inactivation through a serine to alanine mutation of the catalytic serine was observed to abolish cleavage at this site. Introduction of an enzymatically active NS3 protein was not observed to mediate cleavage of this site *in trans*. An inactivated NS3 may still form a complex around the NS3/NS4A site, protecting it from processing *in trans*. Removal of the protease domain permits processing of the NS3/NS4A site *in trans* (Lin *et al.*, 1994b).

NS4A is essential for processing at NS3/NS4A and at NS4B/NS5A (Lin *et al.*, 1994b; Failla *et al.*, 1995; Bouffard et al., 1995) but seems to have little effect at the NS4A/NS4B junction (Failla *et al.*, 1994). The lack of requirement for the presence of NS4A for processing at NS4A/NS4B may be because an endogenous NS4A is present within the substrate. The NS4A/NS4B site cannot be processed *in vitro* in the absence of microsomal membranes (Lin and Rice, 1995).

NS5A/NS5B processing is the most efficient *trans*-cleavage event (D'Souza *et al.*, 1994). Processing at the NS5A/NS5B cleavage site does not require NS4A (Bouffard *et al.*, 1995) although the presence of NS4A significantly enhances the efficiency of processing (Failla *et al.*, 1994). Within the context of the polyprotein the NS5A/NS5B site is processed efficiently with NS5B being produced either cotranslationally or from an unstable processing intermediate (Bartenschlager *et al.*, 1994). The rate of cleavage of the different *trans*-cleavage sites varies between different sites (D'Souza *et al.*, 1995). The NS4A protein can actively stimulate processing by NS3 even when uncleaved (Failla *et al.*, 1994; Bartenschlager *et al.*, 1995a). For *in vitro* analysis uncleaved NS4A cannot act as a cofactor (Bouffard *et al.*, 1995) for NS3 mediated processing.

The protease requirements for processing *in trans* of the HCV Gla NS5A/NS5B cleavage site were examined in Chapter 7. The importance of the NS3 N-terminus for *trans*-cleavage activity was investigated in **7.1.** A cDNA encoding the HCV Gla polyprotein NS234A region was cotransfected with the tNS5T substrate and the ability of this protease complex to process determined. This NS234A polyprotein contained a point mutation within the NS2 region that abolished the activity of the NS2/3 protease. The NS3/NS4A cleavage site was processed *in cis* producing an NS3 protease that had an additional 290 amino acids attached at its N-terminus. The effect of this large protein fragment on the ability of NS3 to process *in trans* was unknown. This additional protein fragment encoded not only the NS2 protein but also the p7 protein. The presence of this extra sequence was shown not abolish the NS3 *trans*-cleavage activity (**7.1.**).

The level of processing observed with cotransfection of the NS234Mut construct was significantly lower than for cotransfection with the NS34A construct. Both these protease cDNAs were under the control of the T7 promoter alone (NS34A in pTF7.5 and NS234Mut in pTZ18) and are likely to be expressed at similar levels. As the NS234A is a significantly longer cDNA, less of the NS3 protease will be produced relative to the NS34A construct. A quantitative comparison between these two protease constructs could be carried out so that the amount of NS3 protease introduced was the same for both constructs. This analysis would determine the effect, if any, of having the additional sequence at the N-terminus of NS3.

What can be determined from this analysis is that the folding of the NS3 serine protease is not affected by the additional N-terminal sequence in such a way that the NS3 protease is inactive. The quantitative analysis would give a more complete insight into this question.

A more detailed characterisation of the protease requirement was carried out to determine the minimal protease region required to process the NS5A/NS5B cleavage site. The tNS5T substrate was cotransfected with a variety of different combinations of protease constructs (**7.2.**). This analysis showed that only cotransfection with NS34A resulted in processing of the tNS5T substrate. Cotransfection of the substrate with either NS3 or NS3P did not produce cleavage products. When these cotransfections were repeated with the NS4A cofactor also introduced, no cleavage products were observed. These results are in contrast to the reported data that demonstrate

that processing of the NS5A/NS5B cleavage site does not require NS4A (Bouffard *et al.*, 1995).

Cotransfection of the tNS5T substrate with an NS3 construct derived from HCV-A resulted in cleavage of the tNS5T substrate. This experiment yielded two pieces of information. The NS5A/NS5B cleavage site could be processed by NS3 alone and that an NS3 construct from a different subtype (1b) could process the HCV Gla polyprotein.

As the HCV-A NS3 was able to process the tNS5T substrate alone and the HCV Gla NS3 required NS4A there must be some factor that differentiates these two proteases. The amino acid sequence of these two protease constructs was compared to see if any differences were present that could contribute to the difference in processing ability.

Sequence analysis demonstrated that there were a number of amino acid differences between the protease constructs. Most of these differences were either substitution to a similar amino acid and/or observed for other HCV isolates. A positively charged basic amino acid, lysine, was present immediately next to the aspartic acid of the catalytic triad. The negative charge of this amino acid is important as it stabilises the correct conformation of the histidine residue of the catalytic triad. The presence of the positive charge adjacent to this residue may have a detrimental effect on the catalytic triad. In the presence of the stabilising cofactor NS4A this negative effect may be overcome, producing an active NS3 protease. When this lysine residue was substituted with the glutamine residue normally found at this position the HCV Gla NS3 protease was able to process the tNS5T substrate alone. When this amino acid was substituted within the NS34A construct the efficiency of cleavage was slightly higher than for the wild type NS34A. This increase in efficiency may be due to the existence of a more stable catalytic triad within the HCV Gla NS3. Sequence comparison of NS3 proteins from different HCV isolates illustrated that although the residue adjacent to the catalytic aspartic acid was not completely conserved a basic residue was never found at this position. The majority of isolates contained a glutamine at this position although a leucine residue and a glycine residue were found in two isolates each. It would seem that it is important that there is no complementary charged residue in close proximity to the aspartic acid of the catalytic triad.

In Figs. 8.10. and 8.11. the active site of the NS3 protease domain is illustrated in the determined crystal structure of NS3. As well as being in close proximity to the catalytic histidine residue, the aspartic acid is also near to an arginine residue with which it could also interact. The presence of an additional positively charged residue adjacent to the aspartic acid may not only affect the interaction with the catalytic histidine residue but also this other amino acid, altering the structure of the protease. The presence of NS4A may stabilise the structure of the NS3 protein, overcome the inhibitory influence of the lysine residue and induce the proteolytic activity of the protease.

One further amino acid difference between the two NS3 constructs was investigated. Eleven amino acids downstream of the catalytic triad there was an arginine residue within the HCV Gla NS3 sequence whereas

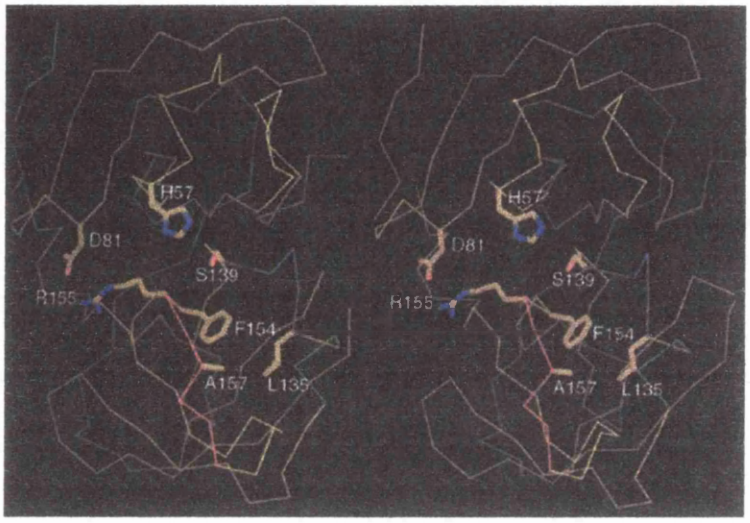


Fig. 8.10. Active site of NS3P (taken from Love *et al.*, 1996). Side-by-side view of the active site and surroundings of HCV NS3P. A Cα trace is given for clarity. Side chains are labeled by the single-letter amino acid code. Shown here is the active site triad (H57, D81, S139) and residues defining the S1 specificity pocket (L135, F154, A157). Overall, this S1 pocket is relatively small and non-polar.

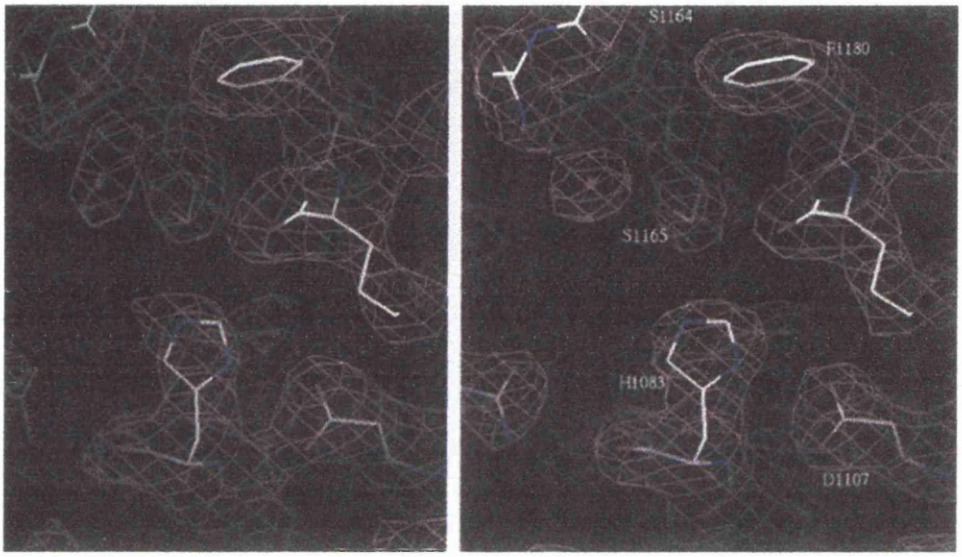


Fig. 8.11. Electron density maps around the active site of tNS3:NS4A (taken from Kim *et al.*, 1996). Electron density map at 2.7 angstroms (left) and 2.5 angstroms (right). The active site residues His-1083, Asp-1107, and Ser-1165 are labeled, as well as Phe-1180, which defines part of the P1 substrate-binding pocket. A well-ordered water molecule occupies the oxyanion hole.

there was a glycine residue within the HCV-A sequence. Analysis of other isolates showed that in all but one of the sequences there was a basic amino acid (arginine for the type 1 HCV isolates and lysine for types 2 and 3) with the exception of one isolate that contained an alanine residue. The relatively widespread conservation of a positive charge at this position was investigated to determine if this feature was of any importance.

The HCV Gla arginine residue was substituted with the glycine residue both within the wild type HCV NS3 sequence and the HCV NS3 sequence containing the glutamine residue beside the catalytic aspartic acid to produce four new protease constructs (NS3[R-G], NS3[K-Q, R-G], NS34A[R-G], NS34A[K-Q, R-G]. When tested for their ability to process, the tNS5T substrate all four protease constructs successfully produced cleavage products.

Although the Arg-Gly substitution enabled the NS3 alone to process it was not as efficient as the NS3[K-Q] protease. When this substitution was present alongside the Lys-Gln substitution the NS3 protease processed with an efficiency between that of the NS3[K-Q] and the NS3[R-G] proteases. The NS34A[R-G] protease processed at essentially the same efficiency as the wild type NS34A. The NS34A[K-Q, R-G] protease processed at the same efficiency as the NS34A[K-Q] protease.

The presence of the glycine 11 amino acids downstream of the catalytic aspartic acid enables the NS3 alone to process the tNS5T substrate despite the presence of the adjacent, inhibitory lysine residue. The loss of the arginine downstream (or replacement with glycine) may have a small effect on the folding and/or the structure of the NS3 and orient the lysine residue further from the aspartic acid, abolishing its inhibitory effect on processing. When the two substitutions are present together the influences exerted may be slightly in opposition, hence the fall in processing efficiency from the NS3[K-Q] protease.

The NS34A[R-G] protease processes at a similar level to the wild type NS34A and this may be a feature of the stabilising nature of the NS4A cofactor. Any influence on structure exerted by the presence of the glycine may be negated by the presence of the NS4A cofactor. The NS34A[K-Q, R-G] protease processes at the same efficiency as the NS34A[K-Q] protease and this may be due to combined effect of the more stable catalytic triad (due to the Lys-Gln substitution) and the presence of NS4A. The possible compensatory effect of the Arg-Gly substitution is not substantial and when the stabilising NS4A cofactor is present the effect is negated.

8.6. Essential nature of viral proteases.

Not every virus encodes a protease but those that do ensure that the proteolytic process occurs at a specific place and at a specific time throughout the viral life cycle. These proteolytic events take place at a particular location within the cell and the protease is targetted to a specific site(s) within the target protein. The activity of the viral protease is

independent of the host cell. The action of the protease is such that it plays an essential role in the maturation of the virus particle, processing the viral polyprotein into proteins required for replication and particle assembly.

Virus-encoded proteases are found in a variety of different viruses and their presence is not dependent on the nature of the genome, complexity of the capsid or presence of a lipid envelope. These proteases have been observed in non-enveloped ssRNA viruses (picornaviruses), enveloped ssRNA viruses (flaviviruses and retroviruses), non-enveloped dsDNA viruses (adenoviruses) and enveloped dsDNA viruses (herpesviruses). The proteases encoded within all these different virus families are essential for the replication of the virus and exhibit a diversity in protease function.

The nature of the flavivirus serine proteases has been discussed previously (**1.8.4**). The activity of other viral proteases can also be demonstrated to be essential for virus replication despite having a significantly different method of action.

Herpes virus particles contain a capsid shell that is initially constructed by the assembly of scaffolding proteins within the capsid core before the viral DNA is packaged. These scaffolding proteins are removed from the herpesvirus particle after DNA introduction to produce the mature virion. For the human cytomegalovirus (CMV) the scaffolding proteins are encoded by two overlapping genes that contain the same 3′ terminus. The processing of these two proteins is carried out by a protease encoded within the N-terminal domain of the larger gene. The C-terminal domain contains the assembly protein. This CMV protease (and all the other equivelant herpesvirus proteases) has been shown to be a serine protease and is highly conserved between different herpesviruses despite having no sequence homology to known serine proteases. This cleavage event is essential for herpesvirus replication. Herpesviruses containing an inactive protease are severely restricted in their replication. The second, downstream gene encodes the assembly protein required for assembly of the scaffold in hepresvirus capsid formation.

The larger protein is cleaved at two sites in all herpesviruses. The M-site (maturation site) is located upstream of the C-terminus of the assembly protein, whereas the R-site (release site) is located upstream of the N-terminus of the assembly protein. The remaining N-terminal fragment contains the catalytically competent enzyme. Both these cleavage events take place at a conserved motif - (V,L) - X - A / S. There is an additional site cleaved within the HCMV protease region called the internal site (I-site). Cleavage at this site produces two fragments that associate together to form an active, "two-chain" protease (Holwedra *et al.*, 1997).

For all herpesviruses, however, the major substrate cleaved by the protease is the smaller, assembly protein translated from the downstream start. After the B-capsid is formed, the protease cleaves the assembly protein at the M-site which enables removal of the scaffold before packaging of the viral DNA.

The CMV protease crystal structure was determined and found to contain a novel active site (Qiu *et al.*, 1996) and a previously unseen fold.

This structural determination showed that the CMV serine protease consisted of a single β-barrel structure as opposed to the two β-barrel structure observed for chymotrypsin and other classical serine proteases. Seven strands composed the β-barrel core that was surrounded by seven α-helices on three sides. The β-barrel core contains two parallel strands with strand B3 forming a β-bend that closes one corner of the barrel. The other corner is closed by two hydrogen bonds between strands B5 and B7. The first four stands (B1-B4) of the HCMV protease form a Greek key motif quite different to that of trypsin. The HCMV protease is believed to be evolutionary unrelated to other serine protease β-barrels and is folded in a unique way compared to other serine proteases.

The dimer interface between the two subunits is composed of one helix from one monomer and four helices of the other. This interface is predominantly hydrophobic and is conserved between herpesviruses. The dimer interface is not close to the active site and it is unclear why dimerisation is required to form an active protease.

The catalytic triad was composed of His/His/Ser instead of Asp/His/Ser normally observed for serine proteases. The CMV protease and the other herpesvirus serine proteases have a low turnover number that may be due to the presence of the His residue as opposed to the usual Asp residue. Although this His residue orients the catalytic His residue it does not increase the nucleophilicity of the serine residue (a feature of the Asp residue when present within the catalytic triad). This His residue, although important for catalytic activity, is not essential. The relatively low turnover rate may be advantageous for correct particle assembly of the hepresvirus virion as scaffold maturation is a relatively slow process. The encapsidation of the viral DNA and the removal of the scaffold proteins takes place only if the serine protease mediated processing takes place. A faster assembly of the virus scaffold may result in incorrectly organised virions leading to less infectious particles.

The CMV serine protease assembles as an allosteric homodimer which processes at the nucleus. The protease is only active as a dimer. The protease cleaves between an alanine residue and a serine residue.

The HSV-1 protease is able to cleave a small peptide substrate containing 5-8 amino acids either side of the cleaved peptide bond. The HCMV protease requires only 5 amino acids either side of the target peptide bond to efficiently process. Mutation of the residues at P4-P1 result in a large decrease in cleavage efficiency. Substitution with the serine at P1′ with alanine did not drastically reduce cleavage. As the I-site contains an alanine at P1′ this observation is not unexpected. The hepresvirus proteases are reviewed in Holwedra (1997).

The adenovirus protease is essential for virus replication and creation of infectious virus. The adenovirus protease (AVP) has a cysteine protease catalytic triad (Glu/His/Cys), in a similar arrangement to cysteine proteases like papain, which is located within a completely novel protein fold. This fold also incorporates a peptide cofactor (Ding *et al.*, 1996). The AVP acts at the nucleus but recognises the GX/G G/X motif. Three α-

helices and a single five-stranded β-sheet form a unique "β-sheet sandwich". Three α-helices from the blunt end of the structure and a seventh α-helix forms the pointed end. An eleven amino acid peptide seems to act as a sixth β-strand next to the core β-sheet and interacts with a number of residues, including a disulphide bond. This factor does not encounter any catalytic residues but may play a conformational role. The AVP processes the core proteins of an assembled adenovirus virion within the nucleus of an infected cell. The AVP binds to the viral DNA non-specifically using four large, positively charged clusters located on the surface of the molecule. This binding, as well as the use of a peptide cofactor, can be viewed as a regulatory step to ensure that the protease activity is not activated until the protease has entered the virion.

As for flaviviruses, the picornavirus replication cycle involves the processing of a single polyprotein to produce the mature virus proteins. The core protein is processed cotranslationally by the 2A protease before protease-independent cleavage during final core maturation. The non-structural region of the picornavirus polyprotein is processed by the 3C protease. Both the 2A and the 3C proteases have an essential cysteine residue. The 3C protease was shown to fold into a chymotrypsin-like double β-barrel (Bazan and Fletterick, 1988).

The picornaviruses are single stranded, positive sense RNA viruses. The RNA genome contains a single open reading frame that encodes a polyprotein that is rapidly processed cotranslationally. Throughout the picornaviruses there are three protease types (L^{pro}, $2A^{pro}$, $3C^{pro}$). The Aphthovirus Lpro is found at the N-terminus of the polyprotein and cleaves cotranslationally at its own C-terminus (Burroughs et al., 1984). L^{pro} can exist as two forms, translated from two in frame AUG codons separated by 84 nucleotides. Both forms cleave at the L/P1 junction. There is a similarity to thiol proteases with the highly conserved residues Cys-51, His-148, Asp-164 forming the catalytic triad. Glu-76 and Asp-164 may play a role in substrate binding. L^{pro} also can cleave the host cell protein eIF-4G *in trans* in a similar manner to entero- and rhino - viruses $2A^{pro}$.

The enterovirus and rhinovirus $2A^{pro}$ cleaved between the capsid precursor P1 and the replicative domains of the polyprotein. The 17kDa $2A^{pro}$ cleaves at its own N-terminus between a tyrosine and a glycine residue although there seem to be a preference rather than requirement for the amino acids present. This protease also cleaves inefficiently at another site within the 3D protein depending on the virus strain. The nature of $2A^{pro}$ is not known but it has sequence homology to a sub-class of cellular serine proteases despite the nucleophile being a cysteine residue. $2A^{pro}$ is suggested to have a His, Asp, Cys catalytic triad with a zinc atom in a structural role. $2A^{pro}$ also inhibits host-cell protein synthesis by cleaving eIF-4G, a protein associated in the cap binding protein complex. Inhibition of cap-dependent translation results in host cell shut off. $2A^{pro}$ also transactivates IRES-mediated translation. The aphthovirus and cardiovirus $2A^{pro}$ cleaves at the C-terminus of 2A, has no known protease motifs and has no similarity to the $2A^{pro}$ to enterovirus $2A^{pro}$.

$3C^{pro}$ is found in all picornaviruses and exhibits high sequence similarity across the genera. This protease is required to cleave at a distal site between 2C and 3A as well as for a series of secondary cleavages that produce the capsid and replicative protein precursors.

The human rhinovirus (HRV) 3C protease structure illustrated that the catalytic triad Glu/His/Cys was positioned in a similar manner of the Asp/His/Ser triad of serine proteases (Matthews *et al.*, 1994). The HRV 3C protease folded in a very similar manner to chymotrypsin but with a number of slight differences. The precise orientation of the catalytic triad, regions required for transition state stabilisation and some connecting loops were different for the 3C protease. The 3C proteases, although more efficient than the CMV protease, is not highly efficient. This factor may be due to the weaker Cys/His catalytic dyad as opposed to the His/Ser dyad seen for serine proteases. The third residue in the catalytic triad is not strictly conserved (Glu or Asp) and from structural analyses seems to point away from the active site. The 3C protease processes between a glutamine residue and a glycine residue.

The 3C protease binds directly to viral RNA and is important for the initiation of RNA replication. The region involved in RNA binding is located distal from the active site. The 3C protease is fused to the viral RNA polymerase (3D) for a large portion of the replication cycle with no effect on the protease activity. The 3C polymerase is inactive in this uncleaved state. As the release of the 3D polymerase from the polyprotein is essential for viral replication the functional activity of the 3C protease is required. The picornavirus proteases are reviewed in Ryan and Flint (1997).

Sindbis virus is an enveloped virus containing a positive sense, single-stranded RNA genome. The Sindbis virus core protein (SCP) has two functions (Tong *et al.*, 1993). As well as binding the RNA genome the SCP contains a proteolytic activity. The Sindbis virus genomic RNA produces two polyproteins that are processed post-translationally by both cellular and viral proteases. The p130 polyprotein encodes the structural proteins including the SCP. At the N-terminus, SCP is released from the polyprotein through a single autocatalytic cleavage *in cis* to release itself from the polyprotein. This cleavage occurs between a tryptophan residue and a serine residue. After this first step in maturation the remainder of the polyprotein is processed by other proteases.

The active SCP structure is a chymotrypsin-like serine protease. The SCP folds into a chymotrypsin-like double β-barrel that flank the substrate binding site. The polypeptide backbone of SCP is similar to chymotrypsin, consisting of two similar Greek key β-barrel domains. The β-barrel is closed when strand D forms hydrogen bonds with strands A and E. The C-terminal β-barrel contains a central, hydrophobic core as does the N-terminal β-barrel, although this core is smaller. The SCP can form a dimer with a relatively small interface and the interaction between the two monomers is mostly hydrophobic. This dimeric form of the protein may also be found in the Sindbis virus core. The N-terminal region associates with the viral RNA and is essential for encapsidation.

The catalytic triad was shown to be Asp/His/Ser. The arrangement of the triad residues is conserved between SCP and other proteases although the orientation of the scissile bond relative to the catalytic residues is different compared to chymotrypsin. The specificity pocket of SCP is a deep, essentially hydrophobic groove on the surface. Once the SCP has been cleaved from the polyprotein the C-terminal tryptophan residue of SCP inserts itself into this groove to inactivate the SCP enzymatic activity. This new structure then assembles into the icosahedrally symmetric core protein.

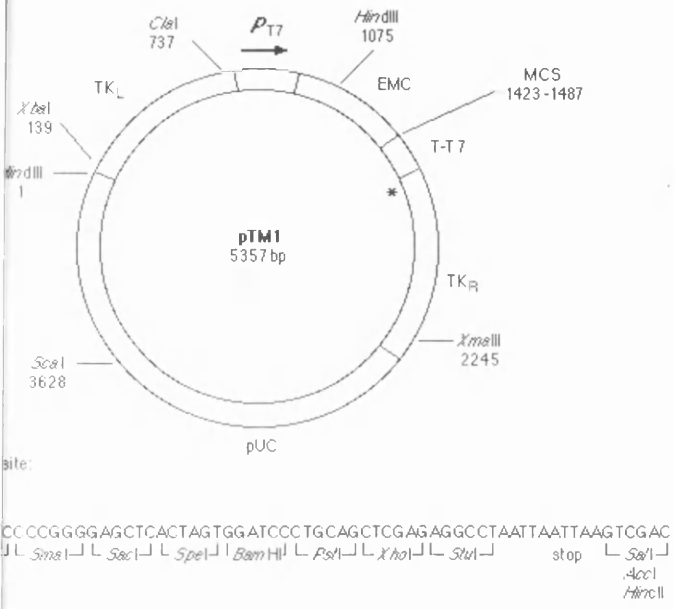
All retroviruses encode a single protease. The HIV protease (PR) processes the gag and gag-pol polyproteins to produce the mature structural proteins and viral protein itself. Sequence alignment with other proteases suggests that the HIV protease could fold to form a single domain of an aspartic protease similar to pepsin. The main difference to pepsin is in the interface region where two HIV-1 PR monomers have significantly shorter N- and C-terminal strands that interact to form a dimer whose structure and properties similar to pepsin. The active site triad (Asp-25, Thr-26 and Gly-27) is located deep within a loop stabilised by a network of hydrogen bonds, a feature of aspartic proteases (Wlodawer et al., 1989). The requirement of dimerisation for protease activity introduces a regulatory element into the virus life cycle. Only when the virions are assembling on the plasma membranes are the polyproteins close enough to permit protease dimerisation and activation. Once dimerised, the proteolytically active homodimer processes between a phenylalanine (or tyrosine residue) and a proline residue.

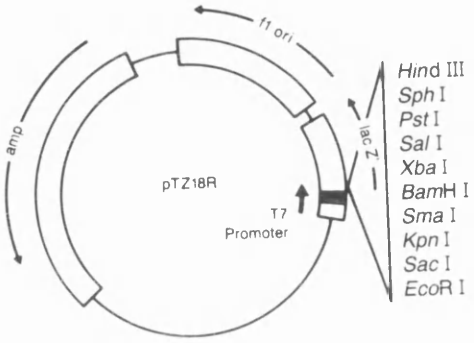
These proteases, although varied in structure and functional requirement, are all essential for the virus maturation and replication.

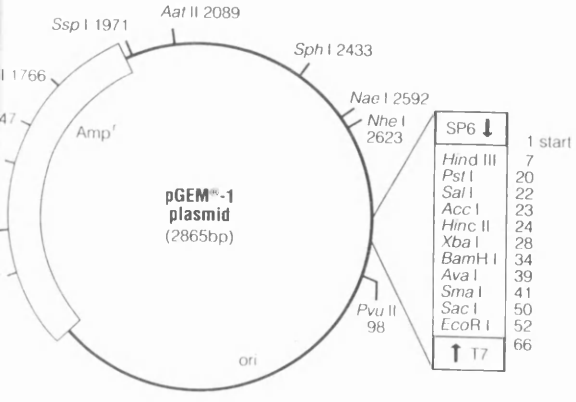
8.7. The importance of antiviral therapy.

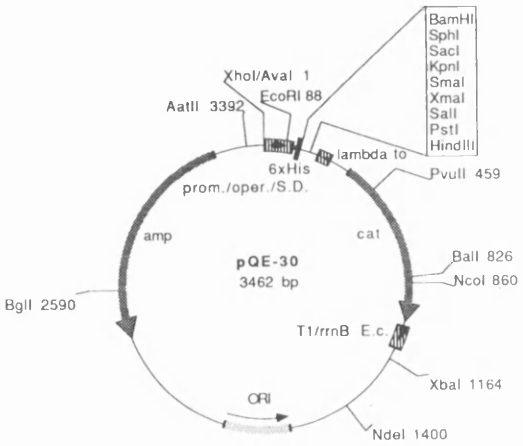
Viral proteases play an important and usually essential role in viral replication. For the majority of RNA viruses e.g. HIV and HCV, the virus population is constantly changing such that a preventative or therapeutic vaccine is unlikely to be of any use. Targetting of essential proteins such as the viral protease or polymerase with antivirals is the only realistic treatment. To inhibit protease function the use of substrate mimics or other antiviral compounds could be effective. A detailed understanding of substrate specificity and protease function and requirements is invaluable for the development of antivirals.

Appendix


ClaI
737
P_{T7}
HindIII
1075
TK$_L$
EMC
MCS
1423-1487
XbaI
139
T-T 7
HindIII
1
pTM1
5357 bp
TK$_R$
XmaIII
2245
ScaI
3628
pUC
site:
CC CCGGGGAGCTCACTAGTGGATCCCTGCAGCTCGAGAGGCCTAATTAATTAAGTCGAC
SmaI SacI SpeI BamHI PstI XhoI StuI
stop
SalI
AccI
HincII


f1 ori
amp
lac Z
pTZ18R
T7
Promoter
Hind III
Sph I
Pst I
Sal I
Xba I
BamH I
Sma I
Kpn I
Sac I
EcoR I


Ssp I 1971
Aat II 2089
Sph I 2433
Nae I 2592
Nhe I
2623
Amp^r
pGEM®-1
plasmid
(2865bp)
Pvu II
98
ori
SP6
T7
1 start
Hind III 7
Pst I 20
Sal I 22
Acc I 23
Hinc II 24
Xba I 28
BamH I 34
Ava I 39
Sma I 41
Sac I 50
EcoR I 52
66


BamHI
SphI
SacI
KpnI
SmaI
XmaI
SalI
PstI
HindIII
XhoI/AvaI 1
EcoRI 88
AatII 3392
lambda to
6xHis
prom./oper./S.D.
PvuII 459
amp
cat
pQE-30
3462 bp
BalI 826
NcoI 860
BglI 2590
T1/rrnB E.c.
XbaI 1164
ORI
NdeI 1400

References

Adams, N. J., Chamberlain, R. W., Taylor, L. A., Davidson, F., Lin, C. K., Elliott, R. M. & Simmonds, P. (1997). Complete coding sequence of hepatitis C virus genotype 6a. *Biochemical and Biophysical Research Communications* **234**, 393-396.

Ali, N. & Siddiqui, A. (1995). Interaction of Polypyrimidine Tract-Binding Protein With the 5'- Noncoding Region of the Hepatitis-C Virus-RNA Genome and its Functional Requirement in Internal Initiation of Translation. *Journal of Virology* **69**, 6367-6375.

Alter, H. J., Purcell, R. H., Shih, J. W., Melpolder, J. C., Houghton, M., Choo, Q. L. & Kuo, G. (1989). Detection of Antibody to Hepatitis-C Virus In Prospectively Followed Transfusion Recipients With Acute and Chronic Non-A-Hepatitis, Non-B- Hepatitis. *New England Journal of Medicine* **321**, 1494-1500.

Asabe, S. I., Tanji, Y., Satoh, S., Kaneko, T., Kimura, K. & Shimotohno, K. (1997). The N-Terminal Region of Hepatitis-C Virus-Encoded NS5A is Important for NS4A-Dependent Phosphorylation. *Journal of Virology* **71**, 790-796.

Bartenschlager, R., Ahlbornlaake, L., Mous, J. & 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**, 3835-3844.

Bartenschlager, R., Ahlbornlaake, L., Mous, J. & Jacobsen, H. (1994). Kinetic and Structural-Analyses of Hepatitis-C Virus Polyprotein Processing. *Journal of Virology* **68**, 5045-5055.

Bartenschlager, R., Ahlbornlaake, L., Yasargil, K., Mous, J. & Jacobsen, H. (1995a). Substrate Determinants for Cleavage *in cis* and *in trans* by the Hepatitis-C Virus NS3 Proteinase. *Journal of Virology* **69**, 198-205.

Bartenschlager, R., Lohmann, V., Wilkinson, T. & Koch, J. O. (1995b). Complex-Formation Between the NS3 Serine-Type Proteinase of the Hepatitis-C Virus and NS4A and its Importance for Polyprotein Maturation. *Journal of Virology* **69**, 7519-7528.

Bartholomeusz, A. I. & Wright, P. J. (1993). Synthesis of Dengue Virus-RNA *in vitro* - Initiation and the Involvement of Proteins NS3 and NS5. *Archives of Virology* **128**, 111-121.

Bastie, A., Pawlotsky, J., Roudotthoraval, F. & Dhumeaux, D. (1995). Hepatitis-C Virus Infection - Epidemiology. *Pathologie Biologie* **43**, 674-680.

Bazan, J. F. & Fletterick, R. J. (1988). Viral Cysteine Proteases are Homologous to the Trypsin-Like Family of Serine Proteases - Structural and Functional Implications. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7872-7876.

Behrens, S. E., Tomei, L. & Defrancesco, R. (1996). Identification and Properties Of the RNA-Dependent RNA-Polymerase of Hepatitis-C Virus. *Embo Journal* **15**, 12-22.

Bellobuono, A., Mondazzi, L., Tempini, S., Silini, E., Vicari, F. & Ideo, G. (1997). Ribavirin and interferon-a combination therapy vs interferon-a alone in the retreatment of chronic hepatitis C: A randomized clinical trial. *Journal of Viral Hepatitis* **4**, 185-191.

Blackwell, J. L. & Brinton, M. A. (1995). BHK Cell-Proteins that Bind to the 3' Stem-Loop Structure of the West Nile Virus Genome RNA. *Journal of Virology* **69**, 5650-5658.

Blight, K. J. & Rice, C. M. (1997). Secondary structure determination of the conserved 98-base sequence at the 3' terminus of hepatitis C virus genome RNA. *Journal of Virology* **71**, 7345-7352.

Bode, W., Walter, J., Huber, R., Wenzel, H. R. & Tschesche, H. (1984). The Refined 2.2-a (0.22-Nm) X-Ray Crystal-Structure of the Ternary Complex formed by Bovine Trypsinogen, Valine-Valine and the Arg15 Analog of Bovine Pancreatic Trypsin-Inhibitor. *European Journal of Biochemistry* **144**, 185-190.

Bouffard, P., Bartenschlager, R., Ahlbornlaake, L., Mous, J., Roberts, N. & Jacobsen, H. (1995). An *In-Vitro* Assay for Hepatitis-C Virus NS3 Serine Proteinase. *Virology* **209**, 52-59.

Boussif, O., Lezoualch, F., Zanta, M. A., Mergny, M. D., Scherman, D., Demeneix, B. & Behr, J. P. (1995). A Versatile Vector for Gene and Oligonucleotide Transfer into Cells in Culture and *in-vivo* - Polyethylenimine. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 7297-7301.

Bradley, D. W., McCaustland, K. A., Cook, E. H., Schable, C. A., Ebert, J. W. & Maynard, J. E. (1985). Posttransfusion Non-A, Non-B Hepatitis in Chimpanzees - Physiochemical Evidence that the Tubule-Forming Agent is a Small, Enveloped Virus. *Gastroenterology* **88**, 773-779.

Brandon, C. & Tooze, J. (1991) An example of enzyme catalysis: serine proteinases. *Introduction to Protein Structure.* 231-245.

Brillanti, S., Garson, J., Foli, M., Whitby, K., Deaville, R., Masci, C., Miglioli, M. & Barbara, L. (1994). Pilot-Study of Combination Therapy with Ribavirin Plus Interferon- Alfa for Interferon Alfa-Resistant Chronic Hepatitis-C. *Gastroenterology* **107**, 812-817.

Brillanti, S., Miglioli, M. & Barbara, L. (1995). Combination Antiviral Therapy With Ribavirin and Interferon-Alfa in Interferon-Alfa Relapsers and Nonresponders - Italian Experience. *Journal of Hepatology* **23**, 13-16.

Brinton, M. A., Fernandez, A. V. & Dispoto, J. H. (1986). The 3'-Nucleotides of Flavivirus Genomic RNA form a Conserved Secondary Structure. *Virology* **153**, 113-121.

Bukh, J., Purcell, R. H. & Miller, R. H. (1993a). At Least 12 Genotypes Of Hepatitis-C Virus Predicted by Sequence- Analysis of the Putative E1-Gene of Isolates Collected Worldwide. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8234-8238.

Bukh, J., Wantzin, P., Krogsgaard, K., Knudsen, F., Purcell, R. H. & Miller, R. H. (1993b). High Prevalence of Hepatitis-C Virus (HCV) RNA in Dialysis Patients - Failure of Commercially Available Antibody Tests to Identify a Significant Number of Patients with HCV Infection. *Journal of Infectious Diseases* **168**, 1343-1348.

Buratti, E., Gerotto, M., Pontisso, P., Alberti, A., Tisminetzky, S. G. & Baralle, F. E. (1997). In vivo translational efficiency of different hepatitis C virus 5'-UTRs. *Febs Letters* **411**, 275-280.

Burroughs, J. N., Sangar, D. V., Clarke, B. E., Rowlands, D. J., Billiau, A. & Collen, D. (1984). Multiple Proteases in Foot-and-Mouth-Disease Virus-Replication. *Journal of Virology* **50**, 878-883.

Butkiewicz, N. J., Wendel, M., Zhang, R. M., Jubin, R., Pichardo, J., Smith, E. B., Hart, A. M., Ingram, R., Durkin, J., Mui, P. W., Murray, R. G., Ramanathan, L. & Dasmahapatra, B. (1996). Enhancement of Hepatitis-C Virus NS3 Proteinase Activity by Association with NS4A-Specific Synthetic Peptides - Identification of Sequence and Critical Residues of NS4A for the Cofactor Activity. *Virology* **225**, 328-338.

Chamberlain, R. W., Adams, N., Saeed, A. A., Simmonds, P. & Elliott, R. M. (1997a). Complete nucleotide sequence of a type 4 hepatitis C virus variant, the predominant genotype in the Middle East. *Journal of General Virology* **78**, 1341-1347.

Chamberlain, R. W., Adams, N. J., Taylor, L. A., Simmonds, P. & Elliott, R. M. (1997b). The complete coding sequence of hepatitis C virus genotype 5a, the predominant genotype in South Africa. *Biochemical and Biophysical Research Communications* **236**, 44-49.

Chambers, T. J., Grakoui, A. & Rice, C. M. (1991). Processing of the Yellow-Fever Virus Nonstructural Polyprotein - a Catalytically Active NS3-Proteinase Domain and NS2B are Required for Cleavages at Dibasic Sites. *Journal of Virology* **65**, 6042-6050.

Chambers, T. J., Hahn, C. S., Galler, R. & Rice, C. M. (1990). Flavivirus Genome Organization, Expression, and Replication. *Annual Review of Microbiology* **44**, 649-688.

Chambers, T. J., Nestorowicz, A., Amberg, S. M. & Rice, C. M. (1993). Mutagenesis of the Yellow-Fever Virus NS2B Protein - Effects on Proteolytic Processing, NS2B-NS3 Complex-Formation, and Viral Replication. *Journal of Virology* **67**, 6797-6807.

Chambers, T. J., Nestorowicz, A. & Rice, C. M. (1995). Mutagenesis of the Yellow-Fever Virus NS2B/3 Cleavage Site - Determinants of Cleavage Site-Specificity and Effects on Polyprotein Processing and Viral Replication. *Journal of Virology* **69**, 1600-1605.

Chang, S. C., Yen, J. H., Kang, H. Y., Jang, M. H. & Chang, M. F. (1994). Nuclear Localisation Signals in the Core Protein of Hepatitis C. *Biochemical and Biophysical Research Communications* **205**, 1284-1290.

Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). Isolation Of a cDNA Clone Derived from a Blood-Borne Non-A, Non-B Viral-Hepatitis Genome. *Science* **244**, 359-362.

Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medinaselby, A., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G. & Houghton, M. (1991). Genetic Organization and Diversity of the Hepatitis-C Virus. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 2451-2455.

Chu, P. W. G. & Westaway, E. G. (1985). Replication Strategy of Kunjin Virus - Evidence for Recycling Role of Replicative Form RNA as Template in Semiconservative and Asymmetric Replication. *Virology* **140**, 68-79.

Cleaves, G. R., Ryan, T. E. & Schlesinger, R. W. (1981). Identification and Characterization of Type-2 Dengue Virus Replicative Intermediate and Replicative Form RNAs. *Virology* **111**, 73-83.

Collett, M. S. (1992). Molecular Genetics of Pestiviruses. *Comparative Immunology Microbiology and Infectious Diseases* **15**, 145-154.

D'Souza, E. D. A., Grace, K., Sangar, D. V., Rowlands, D. J. & Clarke, B. E. (1995). *In-Vitro* Cleavage of Hepatitis-C Virus Polyprotein Substrates by Purified Recombinant NS3 Protease. *Journal of General Virology* **76**, 1729-1736.

D'Souza, E. D. A., Osullivan, E., Amphlett, E. M., Rowlands, D. J., Sangar, D. V. & Clarke, B. E. (1994). Analysis of NS3-Mediated Processing of the Hepatitis-C Virus Nonstructural Region *in-vitro*. *Journal of General Virology* **75**, 3469-3476.

Degos, F. (1994). Epidemiology of Hepatitis-C Virus in Europe. *Fems Microbiology Reviews* **14**, 267-271.

Deng, R. T. & Brock, K. V. (1993). 5' and 3' Untranslated Regions of Pestivirus Genome - Primary and Secondary Structure Analyses. *Nucleic Acids Research* **21**, 1949-1957.

Dhillon, A. P. & Dusheiko, G. M. (1995). Pathology of Hepatitis-C Virus-Infection. *Histopathology* **26**, 297-309.

Di Bisceglie, A. M., Krawczynski, K., Brazzeal, D. & Hoofnagle, J. H. (1992). Hepatitis-C Viral-Antigen (HCVag) in Liver - Effect of Antiviral Therapy. *Hepatology* **16**, A 131-A 131.

Ding, J. Z., McGrath, W. J., Sweet, R. M. & Mangel, W. F. (1996). Crystal-Structure of the Human Adenovirus Proteinase with its 11- Amino-Acid Cofactor. *Embo Journal* **15**, 1778-1783.

Duarte, E. A., Novella, I. S., Ledesma, S., Clarke, D. K., Moya, A., Elena, S. F., Domingo, E. & Holland, J. J. (1994a). Subclonal Components of Consensus Fitness in an RNA Virus Clone. *Journal of Virology* **68**, 4295-4301.

Duarte, E. A., Novella, I. S., Weaver, S. C., Domingo, E., Wainhobson, S., Clarke, D. K., Moya, A., Elena, S. F., Delatorre, J. C. & Holland, J. J. (1994b). RNA Virus Quasi-Species - Significance for Viral Disease and Epidemiology. *Infectious Agents and Disease-Reviews Issues and Commentary* **3**, 201-214.

Dubuisson, J., Hsu, H. H., Cheung, R. C., Greenberg, H. B., Russell, D. G. & Rice, C. M. (1994). Formation and Intracellular-Localization of Hepatitis-C Virus Envelope Glycoprotein Complexes Expressed by Recombinant Vaccinia and Sindbis Viruses. *Journal of Virology* **68**, 6147-6160.

Dubuisson, J. & Rice, C. M. (1996). Hepatitis-C Virus Glycoprotein Folding - Disulfide Bond Formation and Association with Calnexin. *Journal of Virology* **70**, 778-786.

Elbers, K., Tautz, N., Becher, P., Stoll, D., Rumenapf, T. & Thiel, H. J. (1996). Processing in the Pestivirus E2-NS2 Region - Identification of Proteins P7 and E2p7. *Journal of Virology* **70**, 4131-4135.

Enomoto, N., Sakuma, I., Asahina, Y., Kurosaki, M., Murakami, T., Yamamoto, C. & Izumi, N. (1996). Interferon Sensitivity Determining Sequence of the Hepatitis-C Virus Genome. *Hepatology* **24**, 460-460.

Enomoto, N., Sakuma, I., Asahina, Y., Kurosaki, M., Murakami, T., Yamamoto, C., Izumi, N., Marumo, F. & Sato, C. (1995). Comparison of Full-Length Sequences of Interferon-Sensitive and Resistant Hepatitis-C Virus 1b - Sensitivity to Interferon is Conferred by Amino-Acid Substitutions in the NS5A Region. *Journal of Clinical Investigation* **96**, 224-230.

Failla, C., Tomei, L. & Defrancesco, R. (1994). Both NS3 and NS4A are Required for Proteolytic Processing of Hepatitis-C Virus Nonstructural Proteins. *Journal of Virology* **68**, 3753-3760.

Failla, C., Tomei, L. & Defrancesco, R. (1995). An Amino-Terminal Domain of the Hepatitis-C Virus NS3 Protease is Essential for Interaction with NS4A. *Journal of Virology* **69**, 1769-1777.

Farci, P., Alter, H. J., Ogata, N., Wong, D., Engle, R., Miller, R., Dawson, G., Lesmiewski, R., Mushahwar, I. & Purcell, R. (1991). Lack of Protection Against Reinfection with Hepatitis-C Virus (HCV) in Multiple Cross-Challenges of Chimpanzees. *Hepatology* **14**, A 90-A 90.

Fuerst, T. R., Niles, E. G., Studier, F. W. & Moss, B. (1986). Eukaryotic Transient-Expression System Based on Recombinant Vaccinia Virus that Synthesizes Bacteriophage-T7 RNA-Polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 8122-8126.

Fukushi, S., Kurihara, C., Ishiyama, N., Hoshino, F. B., Oya, A. & Katayama, K. (1997). The sequence element of the internal ribosome entry site and a 25-kilodalton cellular protein contribute to efficient internal initiation of translation of hepatitis C virus RNA. *Journal of Virology* **71**, 1662-1666.

Gale, M. J., Korth, M. J., Tang, N. M., Tan, S. L., Hopkins, D. A., Dever, T. E., Polyak, S. J., Gretch, D. R. & Katze, M. G. (1997). Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* **230**, 217-227.

Ganem, D. (1996). *Hepadenaviridae:* The Viruses and Their Replication. *Fields Virology.* **3rd Edition,** 2703-2737.

Gong, Y. H., Trowbridge, R., Macnaughton, T. B., Westaway, E. G., Shannon, A. D. & Gowans, E. J. (1996). Characterization of RNA-Synthesis During a One-Step Growth Curve and of the Replication Mechanism of Bovine Viral Diarrhea Virus. *Journal of General Virology* **77**, 2729-2736.

Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M. & Rice, C. M. (1993a). A 2nd Hepatitis-C Virus-Encoded Proteinase. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 10583-10587.

Grakoui, A., McCourt, D. W., Wychowski, C., Feinstone, S. M. & Rice, C. M. (1993b). Characterization of the Hepatitis-C Virus-Encoded Serine Proteinase - Determination of Proteinase-Dependent Polyprotein Cleavage Sites. *Journal of Virology* **67**, 2832-2843.

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

Grun, J. B. & Brinton, M. A. (1986). Characterization of West Nile Virus RNA-Dependent RNA-Polymerase and Cellular Terminal Adenylyl and Uridylyl Transferases In Cell-Free- Extracts. *Journal of Virology* **60**, 1113-1124.

Hahm, B., Han, D. S., Back, S. H., Song, O. K., Cho, M. J., Kim, C. J., Shimotohno, K. & Jang, S. K. (1995). NS3-4A of Hepatitis-C Virus is a Chymotrypsin-Like Protease. *Journal of Virology* **69**, 2534-2539.

Hamatake, R., Wang, H. H., Butcher, J. A., Bifano, M., Clark, G., Hernandez, D., Zhang, S., Racela, J., Standring, D. & Colonno, R. (1996). Establishment of an *in vitro* assay to Characterise Hepatitis C Virus NS3-4A Protease *Trans*-Processing Activity. *Intervirology*, 249-258.

Han, D. S., Hahm, B., Rho, H. M. & Jang, S. K. (1995). Identification of the Protease Domain in NS3 of Hepatitis-C Virus. *Journal of General Virology* **76**, 985-993.

Hanke, T., Szawlowski, P. & Randall, R. E. (1992). Construction of Solid Matrix Antibody Antigen Complexes Containing Simian Immunodeficiency Virus-P27 using Tag-Specific Monoclonal- Antibody and Tag-Linked Antigen. *Journal of General Virology* **73**, 653-660.

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

Healey, C. J., Read, S., Kurtz, J., Fleming, K. A. & Chapman, R. W. G. (1995a). Hepatitis-C Infection and Alpha-Interferon Therapy in Patients with Normal Transaminases - a Pilot-Study. *Gastroenterology* **108**, A1082-A1082.

Healey, C. J., Smith, D. B., Walker, J. L., Holmes, E. C., Fleming, K. A., Chapman, R. W. G. & Simmonds, P. (1995b). Acute Hepatitis-C Infection after Sexual Exposure. *Gut* **36**, 148-150.

Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S. & Shimotohno, K. (1991a). Hypervariable Regions in the Putative Glycoprotein of Hepatitis-C Virus. *Biochemical and Biophysical Research Communications* **175**, 220-228.

Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M. & Shimotohno, K. (1991b). Gene-Mapping of the Putative Structural Region of the Hepatitis-C Virus Genome by *In vitro* Processing Analysis. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 5547-5551.

Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K. & Shimotohno, K. (1993a). 2 Distinct Proteinase Activities Required for the Processing of a Putative Nonstructural Precursor Protein of Hepatitis-C Virus. *Journal of Virology* **67**, 4665-4675.

Hijikata, M., Mizushima, H., Tanji, Y., Komoda, Y., Hirowatari, Y., Akagi, T., Kato, N., Kimura, K. & Shimotohno, K. (1993b). Proteolytic Processing and Membrane Association of Putative Nonstructural Proteins of Hepatitis-C Virus. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 10773-10777.

Hitomi, Y., McDonnell, W. M., Killeen, A. A. & Askari, F. K. (1995). Sequence Analysis of the Hepatitis C Virus (HCV) Core Gene Suggests the Core Protein as an Appropriate Target for HCV Vaccine Strategies. *Journal of Viral Hepatitis* **2**, 235-241.

Hollinger, F. B. & Ticehurst, J. R. (1996). Hepatitis A virus. *Fields Virology* **3rd Edition**, 735-782

Hollinger, F. B. (1996). Hepatitis B virus. *Fields Virology* **3rd Edition**, 2738-2808.

Holwerda, B. C. (1997). Herpesvirus proteases: Targets for novel antiviral drugs. *Antiviral Research* **35**, 1-21.

Honda, M., Ping, L. H., Rijnbrand, R. C. A., Amphlett, E., Clarke, B., Rowlands, D. & Lemon, S. M. (1996). Structural Requirements for Initiation of Translation by Internal Ribosome Entry within Genome-Length Hepatitis-C Virus-RNA. *Virology* **222**, 31-42.

Hussy, P., Langen, H., Mous, J. & Jacobsen, H. (1996). Hepatitis-C Virus Core Protein - Carboxy-Terminal Boundaries of 2 Processed Species Suggest Cleavage by a Signal Peptide Peptidase. *Virology* **224**, 93-104.

Ito, T. & Lai, M. M. C. (1997). Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. *Journal of Virology* **71**, 8698-8706.

Kage, M., Ogasawara, S., Kosai, K. I., Nakashima, E., Shimamatsu, K., Kojiro, M., Kimura, A., Fujisawa, T., Matsukuma, Y., Ito, Y., Kondo, S., Kawano, K. & Sata, M. (1997). Hepatitis C virus RNA present in saliva but absent in breast-milk of the hepatitis C carrier mother. *Journal of Gastroenterology and Hepatology* **12**, 518-521.

Kapoor, M., Zhang, L. W., Mohan, P. M. & Padmanabhan, R. (1995). Synthesis and Characterization of an Infectious Dengue Virus Type-2 RNA Genome (New-Guinea C-Strain). *Gene* **162**, 175-180.

Kato, N., Lan, K. H., OnoNita, S. K., Shiratori, Y. & Omata, M. (1997). Hepatitis C virus nonstructural region 5A protein is a potent transcriptional activator. *Journal of Virology* **71**, 8856-8859.

Kato, N., Nakazawa, T., Ootsuyama, Y., Sugiyama, K., Ohkoshi, S. & Shimotohno, K. (1994). Virus Isolate-Specific Antibodies against Hypervariable Region-1 of the Hepatitis-C Virus 2nd-Envelope Protein, gp70. *Japanese Journal of Cancer Research* **85**, 987-991.

Kato, N., Ootsuyama, Y., Tanaka, T., Nakagawa, M., Nakazawa, T., Muraiso, K., Ohkoshi, S., Hijikata, M. & Shimotohno, K. (1992). Marked Sequence Diversity in the Putative Envelope Proteins of Hepatitis-C Viruses. *Virus Research* **22**, 107-123.

Khorsi, H., Castelain, S., Wyseur, A., Izopet, J., Canva, V., Rombout, A., Capron, D., Capron, J. P., Lunel, F., Stuyver, L. & Duverlie, G. (1997). Mutations of hepatitis C virus 1b NS5A 2209-2248 amino acid sequence do not predict the response to recombinant interferon-alfa therapy in French patients. *Journal of Hepatology* **27**, 72-77.

Kim, D. W., Gwack, Y., Han, J. H. & Choe, J. (1995). C-Terminal Domain of the Hepatitis-C Virus NS3 Protein Contains an RNA Helicase Activity. *Biochemical and Biophysical Research Communications* **215**, 160-166.

Kim, D. W., Gwack, Y., Han, J. H. & Choe, J. (1997a). Towards defining a minimal functional domain for NTPase and RNA helicase activities of the hepatitis C virus NS3 protein. *Virus Research* **49**, 17-25.

Kim, D. W., Kim, J., Gwack, Y., Han, J. H. & Choe, J. (1997b). Mutational analysis of the hepatitis C virus RNA helicase. *Journal of Virology* **71**, 9400-9409.

Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., Omalley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R. & Thomson, J. A. (1996). Crystal-Structure of the Hepatitis-C Virus NS3 Protease Domain Complexed with a Synthetic NS4A Cofactor Peptide. *Cell* **87**, 343-355.

Kiyosawa, K., Tanaka, E., Sodeyama, T. & Furuta, S. (1994). Natural-History of Hepatitis-C. *Intervirology* **37**, 101-107.

Koch, J. O. & Bartenschlager, R. (1997). Determinants of substrate specificity in the NS3 serine proteinase of the hepatitis C virus. *Virology* **237**, 78-88.

Koch, J. O., Lohmann, V., Herian, U. & Bartenschlager, R. (1996). *In-Vitro* Studies on the Activation of the Hepatitis-C Virus NS3 Proteinase by the NS4A Cofactor. *Virology* **221**, 54-66.

Kolykhalov, A. A., Agapov, E. V., Blight, K. J., Mihalik, K., Feinstone, S. M. & Rice, C. M. (1997). Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA. *Science* **277**, 570-574.

Kolykhalov, A. A., Agapov, E. V. & Rice, C. M. (1994). Specificity of the Hepatitis-C Virus NS3 Serine-Protease - Effects of Substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B Cleavage Sites on Polyprotein Processing. *Journal of Virology* **68**, 7525-7533.

Kolykhalov, A. A., Feinstone, S. M. & Rice, C. M. (1996). Identification of a Highly Conserved Sequence Element at the 3'- Terminus of Hepatitis-C Virus Genome RNA. *Journal of Virology* **70**, 3363-3371.

Komoda, Y., Hijikata, M., Sato, S., Asabe, S. I., Kimura, K. & Shimotohno, K. (1994a). Substrate Requirements of Hepatitis-C Virus Serine Proteinase for Intermolecular Polypeptide Cleavage in Escherichia-Coli. *Journal of Virology* **68**, 7351-7357.

Komoda, Y., Hijikata, M., Tanji, Y., Hirowatari, Y., Mizushima, H., Kimura, K. & Shimotohno, K. (1994b). Processing of Hepatitis-C Viral Polyprotein in Escherichia-Coli. *Gene* **145**, 221-226.

Koonin, E. V. (1991). Similarities In RNA Helicases. *Nature* **352**, 290-290.

Koshy, R. & Inchauspe, G. (1996). Evaluation of Hepatitis-C Virus Protein Epitopes for Vaccine Development. *Trends in Biotechnology* **14**, 364-369.

Kumar, P. K. R., Machida, K., Urvil, P. T., Kakiuchi, N., Vishnuvardhan, D., Shimotohno, K., Taira, K. & Nishikawa, S. (1997). Isolation of RNA aptamers specific to the NS3 protein of hepatitis C virus from a pool of completely random RNA. *Virology* **237**, 270-282.

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, M., Lee, W. S., Kuo, C., Berger, K., Shuster, J. R., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). An Assay For Circulating Antibodies to a Major Etiologic Virus of Human Non-A, Non-B-Hepatitis. *Science* **244**, 362-364.

Kurosaki, M., Enomoto, N., Murakami, T., Sakuma, I., Asahina, Y., Yamamoto, C., Ikeda, T., Tozuka, S., Izumi, N., Marumo, F. & Sato, C. (1997). Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. *Hepatology* **25**, 750-753.

Laemmli, (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lanford, R. E., Notvall, L., Chavez, D., White, R., Frenzel, G., Simonsen, C. & Kim, J. (1993). Analysis of Hepatitis-C Virus Capsid, E1, and E2/NS1 Proteins expressed in Insect Cells. *Virology* **197**, 225-235.

Leinbach, S. S., Bhat, R. A., Xia, S. M., Hum, W. T., Stauffer, B., Davis, A. R., Hung, P. P. & Mizutani, S. (1994). Substrate-Specificity of the NS3 Serine Proteinase of Hepatitis-C Virus as Determined by Mutagenesis at the NS3/NS4A Junction. *Virology* **204**, 163-169.

Lin, C., Lindenbach, B. D., Pragai, B. M., McCourt, D. W. & Rice, C. M. (1994a). Processing in the Hepatitis-C Virus E2-NS2 Region - Identification of P7 and 2 Distinct E2-Specific Products with Different C-Termini. *Journal of Virology* **68**, 5063-5073.

Lin, C., Pragai, B. M., Grakoui, A., Xu, J. & Rice, C. M. (1994b). Hepatitis-C Virus NS3 Serine Proteinase - Trans-Cleavage Requirements and Processing Kinetics. *Journal of Virology* **68**, 8147-8157.

Lin, C. & Rice, C. M. (1995). The Hepatitis-C Virus NS3 Serine Proteinase and NS4A Cofactor - Establishment of a Cell-Free Trans-Processing Assay. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 7622-7626.

Lin, C., Thomson, J. A. & Rice, C. M. (1995). A Central Region in the Hepatitis-C Virus NS4A Protein allows formation of an Active NS3-NS4A Serine Proteinase Complex *In-Vivo* and *In-Vitro*. *Journal of Virology* **69**, 4373-4380.

Lin, C., Wu, J. W., Hsiao, K. & Su, M. S. S. (1997). The hepatitis C virus NS4A protein: Interactions with the NS4B and NS5A proteins. *Journal of Virology* **71**, 6465-6471.

Lo, S. Y., Masiarz, F., Hwang, S. B., Lai, M. M. C. & Ou, J. H. (1995). Differential Subcellular-Localization of Hepatitis-C Virus Core Gene-Products. *Virology* **213**, 455-461.

Lo, S. Y., Selby, M., Tong, M. & Ou, J. H. (1994). Comparative-Studies of the Core Gene-Products of 2 Different Hepatitis-C Virus Isolates - 2 Alternative Forms Determined by a single Amino-Acid Substitution. *Virology* **199**, 124-131.

Lo, S. Y., Selby, M. J. & Ou, J. H. (1996). Interaction Between Hepatitis-C Virus Core Protein and E1 Envelope Protein. *Journal of Virology* **70**, 5177-5182.

Lohmann, V., Korner, F., Herian, U. & Bartenschlager, R. (1997). Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *Journal of Virology* **71**, 8416-8428.

Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T. & Hostomska, Z. (1996). The Crystal-Structure of Hepatitis-C Virus NS3 Proteinase Reveals a Trypsin-Like Fold and a Structural Zinc-Binding Site. *Cell* **87**, 331-342.

Lu, H. H. & Wimmer, E. (1996). Poliovirus Chimeras Replicating under the Translational Control of Genetic Elements of Hepatitis-C Virus Reveal Unusual Properties of the Internal Ribosomal Entry Site of Hepatitis-C Virus. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1412-1417.

Manabe, S., Fuke, I., Tanishita, O., Kaji, C., Gomi, Y., Yoshida, S., Mori, C., Takamizawa, A., Yosida, I. & Okayama, H. (1994). Production of Nonstructural Proteins of Hepatitis-C Virus Requires a Putative Viral Protease encoded by NS3. *Virology* **198**, 636-644.

Markland, W., Petrillo, R. A., Fitzgibbon, M., Fox, T., McCarrick, M. A., McQuaid, J. A., Fulgham, J. R., Chen, W. Y., Fleming, M. A., Thomson, J. A., Chambers, S.P (1997). Purification and Characterisation of the NS3 serine protease domain of Hepatitis C Virus expressed in Saccharomyces cerevisiae. *Journal of General Virology* **78**, 39-43.

Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J. & Gomez, J. (1992). Hepatitis-C Virus (HCV) circulates as a Population of Different but Closely Related Genomes - Quasi-Species Nature of HCV Genome Distribution. *Journal of Virology* **66**, 3225-3229.

Matsumoto, M., Hsieh, T. Y., Zhu, N. L., VanArsdale, T., Hwang, S. B., Jeng, K. S., Gorbalenya, A. E., Lo, S. Y., Ou, J. H., Ware, C. F. & Lai, M. M. C. (1997). Hepatitis C virus core protein interacts with the cytoplasmic tail of lymphotoxin-beta receptor. *Journal of Virology* **71**, 1301-1309.

Matsumoto, M., Hwang, S. B., Jeng, K. S., Zhu, N. L. & Lai, M. M. C. (1996). Homotypic Interaction and Multimerization of Hepatitis-C Virus Core Protein. *Virology* **218**, 43-51.

Matsuura, Y. & Miyamura, T. (1993). The Molecular-Biology of Hepatitis-C Virus. *Seminars in Virology* **4**, 297-304.

Matsuura, Y., Suzuki, T., Suzuki, R., Sato, M., Aizaki, H., Saito, I. & Miyamura, T. (1994). Processing of E1 and E2 Glycoproteins of Hepatitis-C Virus Expressed In Mammalian and Insect Cells. *Virology* **205**, 141-150.

Matthews, D., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C., McElroy, H., Gribskov, C. & Worland, S. (1994). Crystal-Structure of Human Rhinovirus type-14 3C-Protease. *Journal of Cellular Biochemistry*, 170-170.

Miller, R. H. & Purcell, R. H. (1990). Hepatitis-C Virus shares Amino-Acid-Sequence Similarity with Pestiviruses and Flaviviruses as well as Members of 2 Plant-Virus Supergroups. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 2057-2061.

Mizushima, H., Hijikata, M., Asabe, S. I., Hirota, M., Kimura, K. & Shimotohno, K. (1994). 2 Hepatitis-C Virus Glycoprotein E2 Products with Different C-Termini. *Journal of Virology* **68**, 6215-6222.

Moormann, R. J. M., Vangennip, H. G. P., Miedema, G. K. W., Hulst, M. M. & Vanrijn, P. A. (1996). Infectious RNA Transcribed from an Engineered Full-Length cDNA Template of the Genome of a Pestivirus. *Journal of Virology* **70**, 763-770.

Morgenstern, K. A., Landro, J. A., Hsiao, K., Lin, C., Gu, Y., Su, M. S. S. & Thomson, J. A. (1997). Polynucleotide modulation of the protease, nucleoside triphosphatase, and helicase activities of a hepatitis C virus NS3-NS4A complex isolated from transfected COS cells. *Journal of Virology* **71**, 3767-3775.

Moriya, T., Sasaki, F., Mizui, M., Ohno, N., Mohri, H., Mishiro, S. & Yoshizawa, H. (1995). Transmission of Hepatitis-C Virus from Mothers to Infants - Its Frequency and Risk-Factors Revisited. *Biomedicine & Pharmacotherapy* **49**, 59-64.

Muramatsu, S., Ishido, S., Fujita, T., Itoh, M. & Hotta, H. (1997). Nuclear localization of the NS3 protein of hepatitis C virus and factors affecting the localization. *Journal of Virology* **71**, 4954-4961.

Nakatsuji, Y., Matsumoto, A., Tanaka, E., Ogata, H. & Kiyosawa, K. (1992). Detection of Chronic Hepatitis-C Virus-Infection by 4 Diagnostic Systems - 1st-Generation and 2nd-Generation Enzyme-Linked- Immunosorbent-Assay, 2nd-Generation Recombinant Immunoblot Assay and Nested Polymerase Chain-Reaction Analysis. *Hepatology* **16**, 300-305.

Ohto, H., Terazawa, S., Sasaki, N., Sasaki, N., Hino, K., Ishiwata, C., Kako, M., Ujiie, N., Endo, C., Matsui, A., Okamoto, H., Mishiro, S., Kojima, M., Aikawa, T., Shimoda, K., Sakamoto, M., Akahane, Y., Yoshizawa, H., Tanaka, T., Tokita, H. & Tsuda, F. (1994). Transmission of Hepatitis-C Virus from Mothers to Infants. *New England Journal of Medicine* **330**, 744-750.

Overton, H., McMillan, D., Gillespie, F. & Mills, J. (1995). Recombinant Baculovirus-Expressed NS3 Proteinase of Hepatitis-C Virus shows Activity in Cell-Based and *In-Vitro* Assays. *Journal of General Virology* **76**, 3009-3019.

Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W. & Wittinghofer, A. (1990). Refined Crystal-Structure of the Triphosphate Conformation of H-Ras P21 at 1.35 A Resolution - Implications for the Mechanism of GTP Hydrolysis. *Embo Journal* **9**, 2351-2359.

Pieroni, L., Santolini, E., Fipaldini, C., Pacini, L., Migliaccio, G. & LaMonica, N. (1997). *In vitro* study of the NS2-3 protease of hepatitis C virus. *Journal of Virology* **71**, 6373-6380.

Pizzi, E., Tramontano, A., Tomei, L., Lamonica, N., Failla, C., Sardana, M., Wood, T. & Defrancesco, R. (1994). Molecular-Model of the Specificity Pocket of the Hepatitis-C Virus Protease - Implications for Substrate Recognition. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 888-892.

Poole, T. L., Wang, C. Y., Popp, R. A., Potgieter, L. N. D., Siddiqui, A. & Marc, S. (1995). Pestivirus Translation Initiation occurs by Internal Ribosome Entry. *Virology* **206**, 750-754.

Preugschat, F. & Strauss, J. H. (1991). Processing of Nonstructural Protein-NS4A and Protein-NS4B of Dengue-2 Virus *In-vitro* and *In-vivo*. *Virology* **185**, 689-697.

Proutski, V., Gould, E. A. & Holmes, E. C. (1997). Secondary structure of the 3' untranslated region of flaviviruses: Similarities and differences. *Nucleic Acids Research* **25**, 1194-1202.

Purcell, R. H. (1996). Hepatitis E Virus. *Fields Virology* **3rd Edition,** 2831-2844.

The QIAexpressionist (1992) **2nd Edition.**

Qiu, X. Y., Culp, J. S., Dilella, A. G., Hellmig, B., Hoog, S. S., Janson, C. A., Smith, W. W. & Abdelmeguid, S. S. (1996). Unique Fold and Active-Site In Cytomegalovirus Protease. *Nature* **383**, 275-279.

Ralston, R., Thudium, K., Berger, K., Kuo, C., Gervase, B., Hall, J., Selby, M., Kuo, G., Houghton, M. & Choo, Q. L. (1993). Characterization of Hepatitis-C Virus Envelope Glycoprotein Complexes Expressed by Recombinant Vaccinia Viruses. *Journal of Virology* **67**, 6753-6761.

Rasi, G., Mutchnick, M. G., Divirgilio, D., Sinibaldivallebona, P., Pierimarchi, P., Colella, F., Favalli, C. & Garaci, E. (1996). Combination Low-Dose Lymphoblastoid Interferon and Thymosin Alpha(1) Therapy in the Treatment of Chronic Hepatitis-B. *Journal of Viral Hepatitis* **3**, 191-196.

Ravaggi, A., Natoli, G., Primi, D., Albertini, A., Levrero, M. & Cariani, E. (1994). Intracellular-Localization of Full-Length and Truncated Hepatitis-C Virus Core Protein Expressed in Mammalian-Cells. *Journal of Hepatology* **20**, 833-836.

Ray, R. B., Lagging, L. M., Meyer, K., Steele, R. & Ray, R. (1995). Transcriptional Regulation of Cellular and Viral Promoters by the Hepatitis-C Virus Core Protein. *Virus Research* **37**, 209-220.

Ray, R. B., Meyer, K. & Ray, R. (1996). Suppression of apoptotic cell death by hepatitis C virus core protein. *Virology* **226**, 176-182.

Ray, R. B., Steele, R., Meyer, K. & Ray, R. (1997). Transcriptional repression of p53 promoter by hepatitis C virus core protein. *Journal of Biological Chemistry* **272**, 10983-10986.

Reed, K. E., Grakoui, A. & Rice, C. M. (1995). Hepatitis-C Virus-Encoded NS2-3 Protease - Cleavage-Site Mutagenesis and Requirements for Bimolecular Cleavage. *Journal of Virology* **69**, 4127-4136.

Reed, K. E., Xu, J. & Rice, C. M. (1997). Phosphorylation of the hepatitis C virus NS5A protein *in vitro* and *in vivo*: Properties of the NS5A-associated kinase. *Journal of Virology* **71**, 7187-7197.

Resti, M., Azzari, C., Lega, L., Rossi, M. E., Zammarchi, E., Novembre, E. & Vierucci, A. (1995). Mother-to-Infant transmission of Hepatitis-C Virus. *Acta Paediatrica* **84**, 251-255.

Reynolds, J. E., Kaminski, A., Carroll, A. R., Clarke, B. E., Rowlands, D. J. & Jackson, R. J. (1996). Internal Initiation of Translation of Hepatitis-C Virus-RNA - the Ribosome Entry Site is at the Authentic Initiation Codon. *RNA-a Publication of the RNA Society* **2**, 867-878.

Reynolds, J. E., Kaminski, A., Kettinen, H. J., Grace, K., Clarke, B. E., Carroll, A. R., Rowlands, D. J. & Jackson, R. J. (1995). Unique Features of Internal Initiation of Hepatitis-C Virus-RNA Translation. *Embo Journal* **14**, 6010-6020.

Rice, C. M. (1996). Flaviviridae: The Viruses and their Replication. In *Fields Virology*, 3 edn.

Rijnbrand, R., Bredenbeek, P., Vanderstraaten, T., Whetter, L., Inchauspe, G., Lemon, S. & Spaan, W. (1995). Almost the Entire 5' Non-Translated Region of Hepatitis-C Virus is Required for Cap-Independent Translation. *Febs Letters* **365**, 115-119.

Rijnbrand, R., Vanderstraaten, T., Vanrijn, P. A., Spaan, W. J. M. & Bredenbeek, P. J. (1997). Internal Entry of Ribosomes is Directed by the 5'-Noncoding Region of Classical Swine Fever Virus and is Dependent on the Presence of an RNA Pseudoknot Upstream of the Initiation Codon. *Journal of Virology* **71**, 451-457.

Rijnbrand, R. C. A., Abbink, T. E. M., Haasnoot, P. C. J., Spaan, W. J. M. & Bredenbeek, P. J. (1996). The influence of AUG Codons in the Hepatitis-C Virus 5'-Nontranslated Region on Translation and Mapping of the Translation Initiation Window. *Virology* **226**, 47-56.

Rose, J. K., Buonocore, L. & Whitt, M. A. (1991). A New Cationic Liposome Reagent Mediating Nearly Quantitative Transfection of Animal-Cells. *Biotechniques* **10**, 520-525.

Rumenapf, T., Unger, G., Strauss, J. H. & Thiel, H. J. (1993). Processing of the Envelope Glycoproteins of Pestiviruses. *Journal of Virology* **67**, 3288-3294.

Ryan, M. D. & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* **78**, 699-723.

Saito, I., Miyamura, T., Ohbayashi, A., Harada, H., Katayama, T., Kikuchi, S., Watanabe, Y., Koi, S., Onji, M., Ohta, Y., Choo, Q. L., Houghton, M. & Kuo, G. (1990). Hepatitis-C Virus-Infection is Associated with the Development of Hepatocellular-Carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 6547-6549.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**, 161-178.

Santolini, E., Migliaccio, G. & Lamonica, N. (1994). Biosynthesis and Biochemical-Properties of the Hepatitis-C Virus Core Protein. *Journal of Virology* **68**, 3631-3641.

Santolini, E., Pacini, L., Fipaldini, C., Migliaccio, G. & Lamonica, N. (1995). The NS2 Protein of Hepatitis-C Virus is a Transmembrane Polypeptide. *Journal of Virology* **69**, 7461-7471.

Satoh, S., Tanji, Y., Hijikata, M., Kimura, K. & Shimotohno, K. (1995). The N-Terminal Region of Hepatitis-C Virus Nonstructural Protein-3 (NS3) is Essential for Stable Complex-Formation with NS4A. *Journal Of Virology* **69**, 4255-4260.

Scotto, G., Savastano, A. M., Fazio, V., Conte, P. E., Ferrara, S., Mangano, A. & Tantimonaco, G. (1996). Sexual Transmission of Hepatitis-C Virus-Infection. *European Journal of Epidemiology* **12**, 241-244.

Selby, M. J., Glazer, E., Masiarz, F. & Houghton, M. (1994). Complex Processing and Protein-Protein Interactions in the E2-NS2 Region of HCV. *Virology* **204**, 114-122.

Shih, C. M., Chen, C. M., Chen, S. Y. & Lee, Y. H. W. (1995). Modulation of the Trans-Suppression Activity of Hepatitis-C Virus Core Protein by Phosphorylation. *Journal of Virology* **69**, 1160-1171.

Simmonds, P., Holmes, E. C., Cha, T. A., Chan, S. W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J. & Urdea, M. S. (1993). Classification of Hepatitis-C Virus into 6 Major Genotypes and a Series Of Subtypes by Phylogenetic Analysis of the NS-5 Region. *Journal of General Virology* **74**, 2391-2399.

Smith, J. P. (1997). Treatment of chronic hepatitis C with amantadine. *Digestive Diseases and Sciences* **42**, 1681-1687.

Speight, G., Coia, G., Parker, M. D. & Westaway, E. G. (1988). Gene-Mapping and Positive Identification of the Non-Structural Proteins NS2A, NS2B, NS3, NS4B and NS5 of the Flavivirus Kunjin and their Cleavage Sites. *Journal of General Virology* **69**, 23-34.

Srinvas, R. V., Ray, R. B., Meyer, K. & Ray, R. (1996). Hepatitis-C Virus Core Protein Inhibits Human-Immunodeficiency-Virus Type-1 Replication. *Virus Research* **45**, 87-92.

Steinkuhler, C., Tomei, L. & Defrancesco, R. (1996a). In-Vitro Activity of Hepatitis-C Virus Protease NS3 Purified From Recombinant Baculovirus-Infected SF9 Cells. *Journal of Biological Chemistry* **271**, 6367-6373.

Steinkuhler, C., Urbani, A., Tomei, L., Biasiol, G., Sardana, M., Bianchi, E., Pessi, A. & Defrancesco, R. (1996b). Activity Of Purified Hepatitis-C Virus Protease NS3 On Peptide- Substrates. *Journal Of Virology* **70**, 6694-6700.

Stempniak, M., Hostomska, Z., Nodes, B. R. & Hostomsky, Z. (1997). The NS3 proteinase domain of hepatitis C virus is a zinc-containing enzyme. *Journal Of Virology* **71**, 2881-2886.

Suzich, J. A., Tamura, J. K., Palmerhill, F., Warrener, P., Grakoui, A., Rice, C. M., Feinstone, S. M. & Collett, M. S. (1993). Hepatitis-C Virus NS3 Protein Polynucleotide-Stimulated Nucleoside Triphosphatase and Comparison With the Related Pestivirus and Flavivirus Enzymes. *Journal Of Virology* **67**, 6152-6158.

Suzuki, R., Matsuura, Y., Suzuki, T., Ando, A., Chiba, J., Harada, S., Saito, I. & Miyamura, T. (1995). Nuclear-Localization of the Truncated Hepatitis-C Virus Core Protein with its Hydrophobic C-Terminus Deleted. *Journal of General Virology* **76**, 53-61.

Tabor, S. & Richardson, C. C. (1987) Sequence Analysis with a Modified Bacteriophage-T7 DNA-polymerase. *Proceedings of the National Academy of Science of the United States if America* **84**, 4767-4771.

Tai, C. L., Chi, W. K., Chen, D. S. & Hwang, L. H. (1996). The Helicase Activity Associated with Hepatitis-C Virus Nonstructural Protein-3 (Ns3). *Journal of Virology* **70**, 8477-8484.

Takada, A., Tsutsumi, M. & Sawada, M. (1996). Effect of a Combination Therapy with Interferon and Ofloxacin for Chronic Type-C Hepatitis - a Multicenter Randomized Trial. *Hepatology* **23**, 249-P 249.

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

Takahashi, K., Okamoto, H., Kishimoto, S., Munekata, E., Tachibana, K., Akahane, Y., Yoshizawa, H. & Mishiro, S. (1992b). Demonstration of a Hepatitis-C Virus-Specific Antigen Predicted from the Putative Core Gene in the Circulation of Infected Hosts. *Journal of General Virology* **73**, 667-672.

Tamura, J. K., Warrener, P. & Collett, M. S. (1993). RNA-Stimulated NTPase Activity Associated with the p80 Protein of the Pestivirus Bovine Viral Diarrhea Virus. *Virology* **193**, 1-10.

Tanaka, T., Kato, N., Cho, M. J. & Shimotohno, K. (1995). A Novel Sequence Found at the 3'-Terminus of Hepatitis-C Virus Genome. *Biochemical and Biophysical Research Communications* **215**, 744-749.

Tanaka, T., Kato, N., Cho, M. J., Sugiyama, K. & Shimotohno, K. (1996). Structure of the 3'-Terminus of the Hepatitis-C Virus Genome. *Journal of Virology* **70**, 3307-3312.

Tanimoto, A., Ide, Y., Arima, N., Sasaguri, Y. & Padmanabhan, R. (1997). The amino terminal deletion mutants of hepatitis C virus nonstructural protein NS5A function as transcriptional activators in yeast. *Biochemical and Biophysical Research Communications* **236**, 360-364.

Tanji, Y., Hijikata, M., Hirowatari, Y. & Shimotohno, K. (1994a). Hepatitis-C Virus Polyprotein Processing - Kinetics and Mutagenic Analysis of Serine Proteinase-Dependent Cleavage. *Journal of Virology* **68**, 8418-8422.

Tanji, Y., Hijikata, M., Hirowatari, Y. & Shimotohno, K. (1994b). Identification of the Domain Required for Trans-Cleavage Activity of Hepatitis-C Viral Serine Proteinase. *Gene* **145**, 215-219.

Tanji, Y., Hijikata, M., Satoh, S., Kaneko, T. & Shimotohno, K. (1995a). Hepatitis-C Virus-Encoded Nonstructural Protein NS4A has versatile functions in Viral Protein Processing. *Journal of Virology* **69**, 1575-1581.

Tanji, Y., Kaneko, T., Satoh, S. & Shimotohno, K. (1995b). Phosphorylation of Hepatitis-C Virus-Encoded Nonstructural Protein NS5A. *Journal of Virology* **69**, 3980-3986.

Tautz, N., Elbers, K., Stoll, D., Meyers, G. & Thiel, H. J. (1997). Serine protease of pestiviruses: Determination of cleavage sites. *Journal of Virology* **71**, 5415-5422.

Taylor, J. M. (1996). Hepatitis Delta Virus and Its Replication. *Fields Virology* **3rd Edition,** 2809-2818.

Thomas, H. C., Pickering, J. & Karayiannis, P. (1997). Identification, prevalence and aspects of molecular biology of hepatitis G virus. *Journal of Viral Hepatitis* **4**, 51-54.

Tokushige, K., Wakita, T., Pachuk, C., Moradpour, D., Weiner, D. B., Zurawski, V. R. & Wands, J. R. (1996). Expression and Immune-Response to Hepatitis-C Virus Core DNA-Based Vaccine Constructs. *Hepatology* **24**, 14-20.

Tomei, L., Failla, C., Santolini, E., Defrancesco, R. & Lamonica, N. (1993). NS3 Is a Serine-Protease Required for Processing of Hepatitis-C Virus Polyprotein. *Journal of Virology* **67**, 4017-4026.

Tong, L., Wengler, G. & Rossmann, M. G. (1993). Refined Structure of Sindbis Virus Core Protein and Comparison With other Chymotrypsin-Like Serine Proteinase Structures. *Journal of Molecular Biology* **230**, 228-247.

Tong, M. J., Lai, P. P. C., Hwang, S. J., Lee, S. Y., Co, R. L., Chien, R. N. D. & Kuo, G. (1995). Evaluation of Sexual Transmission in Patients with Chronic Hepatitis- C Infection. *Clinical and Diagnostic Virology* **3**, 39-47.

Tremolada, F. (1992). Long-term follow-up of non-A, non-B (type C) posttransfusion hepatitis. *Journal of Hepatology* **16**, 273-281.

Trent, D. W. & C.W., N. (1980). Biochemistry and replication. In *St. Louis encephalitis*, pp. 159-199. Edited by T. Monath. Washington D.C.: American Public Health Association.

Uehara, S., Abe, Y., Saiti, T., Yoshida, Y., Wagatsuma, S. & Okamura, K. (1993). The Incidence of Vertical Transmission of Hepatitis C Virus. *Tohoku Journal of Experimental Medicine* **171**, 195-202.

Urbani, A., Bianchi, E., Narjes, F., Tramontano, A., DeFrancesco, R., Steinkuhler, C. & Pessi, A. (1997). Substrate specificity of the hepatitis C virus serine protease NS3. *Journal of Biological Chemistry* **272**, 9204-9209.

Urvil, P. T., Kakiuchi, N., Zhou, D. M., Shimotohno, K., Kumar, P. K. R. & Nishikawa, S. (1997). Selection of RNA aptamers that bind specifically to the NS3 protease of hepatitis C virus. *European Journal of Biochemistry* **248**, 130-138.

van der Poel, C. L., Cuypers, H. T. & Reesink, H. W. (1994). Hepatitis-C Virus 6 Years On. *Lancet* **344**, 1475-1479.

Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982). Distantly Related Sequences in the Alpha-Subunits and Beta-Subunits of ATP Synthase, Myosin, Kinases and other ATP-Requiring Enzymes and a Common Nucleotide Binding Fold. *Embo Journal* **1**, 945-951.

Wallner, G., Mandl, C. W., Kunz, C. & Heinz, F. X. (1995). The Flavivirus 3'-Noncoding Region - Extensive Size Heterogeneity Independent of Evolutionary Relationships among strains of Tick-Borne Encephalitis-Virus. *Virology* **213**, 169-178.

Wang, C. Y., Le, S. Y., Ali, N. & Siddiqui, A. (1995). An RNA Pseudoknot is an Essential Structural Element of the Internal Ribosome Entry Site located within the Hepatitis-C Virus 5'-Noncoding Region. *RNA-a Publication Of the RNA Society* **1**, 526-537.

Warrener, P., Tamura, J. K. & Collett, M. S. (1993). RNA-Stimulated NTPase Activity associated with Yellow-Fever Virus NS3 Protein Expressed in Bacteria. *Journal of Virology* **67**, 989-996.

Watson, H. G., Ludlam, C. A., Rebus, S., Zhang, L. Q., Peutherer, J. F. & Simmonds, P. (1992). Use of Several 2nd-Generation Serological Assays to Determine the True Prevalence of Hepatitis-C Virus-Infection in Hemophiliacs Treated with Non-Virus Inactivated Factor-VIII and Factor-IX concentrates. *British Journal of Haematology* **80**, 514-518.

Weiner, A. J., Brauer, M. J., Rosenblatt, J., Richman, K. H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q. L., Houghton, M. & Han, J. H. (1991). Variable and Hypervariable Domains are Found in the Regions of HCV Corresponding to the Flavivirus Envelope and NS1 Proteins and the Pestivirus Envelope Glycoproteins. *Virology* **180**, 842-848.

Weiner, A. J., Truett, M. A., Han, J., Polito, A. J., Choo, Q. L., Page, E., Rosenblatt, J., Quan, S., Kuo, G., Houghton, M., Agius, C. & Nelles, M. J. (1990). HCV Testing In Low-Risk Population. *Lancet* **336**, 695-695.

Wengler, G. & Wengler, G. (1991). The Carboxy-Terminal Part of the NS3-Protein of the West Nile Flavivirus can be isolated as a Soluble-Protein after Proteolytic Cleavage and Represents an RNA-Stimulated NTPase. *Virology* **184**, 707-715.

Westaway, E. G. (1987). Flavivirus Replication Strategy. *Advances in Virus Research* **33**, 45-90.

Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J. & Kent, S. B. H. (1989). Conserved Folding in Retroviral Proteases - Crystal-Structure of a Synthetic HIV-1 Protease. *Science* **245**, 616-621.

Xu, J. A., Mendez, E., Caron, P. R., Lin, C., Murcko, M. A., Collett, M. S. & Rice, C. M. (1997). Bovine viral diarrhea virus NS3 serine proteinase: Polyprotein cleavage sites, cofactor requirements, and molecular model of an enzyme essential for pestivirus replication. *Journal of Virology* **71**, 5312-5322.

Yamada, N., Tanihara, K., Takada, A., Yorihuzi, T., Tsutsumi, M., Shimomura, H., Tsuji, T. & Date, T. (1996). Genetic Organization and Diversity of the 3'-Noncoding Region of the Hepatitis-C Virus Genome. *Virology* **223**, 255-261.

Yanagi, M., Purcell, R. H., Emerson, S. U. & Bukh, J. (1997). Transcripts from a single full-length cDNA clone of hepatitis C virus are infectious when directly transfected into the liver of a chimpanzee. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 8738-8743.

Yi, M. Y., Kaneko, S., Yu, D. Y. & Murakami, S. (1997). Hepatitis C virus envelope proteins bind lactoferrin. *Journal of Virology* **71**, 5997-6002.

Yoshioka, K., Kakumu, S., Wakita, T., Ishikawa, T., Itoh, Y., Takayanagi, M., Higashi, Y., Shibata, M. & Morishima, T. (1992). Detection of Hepatitis-C Virus by Polymerase Chain-Reaction and Response to Interferon-Alpha Therapy - Relationship to Genotypes of Hepatitis-C Virus. *Hepatology* **16**, 293-299.

Yuasa, T., Ishikawa, G., Manabe, S., Sekiguchi, S., Takeuchi, K. & Miyamura, T. (1991). The Particle-Size of Hepatitis-C Virus Estimated by Filtration through Microporous Regenerated Cellulose Fiber. *Journal of General Virology* **72**, 2021-2024.

Yuki, N., Hayashi, N., Moribe, T., Matsushita, Y., Tabata, T., Inoue, T., Kanazawa, Y., Ohkawa, K., Kasahara, A., Fusamoto, H. & Kamada, T. (1997). Relation of disease activity during chronic hepatitis C infection to complexity of hypervariable region 1 quasispecies. *Hepatology* **25**, 439-444.

Zanetti, A. R., Tanzi, E., Paccagnini, S., Principi, N., Pizzocolo, G., Caccamo, M. L., Damico, E., Cambie, G., Vecchi, L., Bresciani, S., Marin, M. G., Padula, D., Rodella, A., Bulgarelli, I., Chiodo, F., Magliano, E., Miotto, G., Muggiasca, M. L., Pilloton, E., Pozzoli, R., Pregliasco, F., Romano, L., Stringhi, C., Dagostino, F., Paolillo, F. & Zapparoli, B. (1995). Mother-to-Infant Transmission of Hepatitis-C Virus. *Lancet* **345**, 289-291.

Zeuzem, S., Lee, J. H. & Roth, W. K. (1997). Mutations in the nonstructural 5A gene of European hepatitis C virus isolates and response to interferon Alfa. *Hepatology* **25**, 740-744.

Zhang, L. W., Mohan, P. M. & Padmanabhan, R. (1992). Processing and Localization of Dengue Virus Type-2 Polyprotein Precursor NS3-NS4A-NS4B-NS5. *Journal of Virology* **66**, 7549-7554.

Zhang, R. M., Durkin, J., Windsor, W. T., McNemar, C., Ramanathan, L. & Le, H. V. (1997). Probing the substrate specificity of hepatitis C virus NS3 serine protease by using synthetic peptides. *Journal of Virology* **71**, 6208-6213.

Zuccotti, G. V., Ribero, M. L., Giovannini, M., Fasola, M., Riva, E., Portera, G., Biasucci, G., Decarlis, S., Profeta, M. L. & Tagger, A. (1995). Effect of Hepatitis-C Genotype on Mother-to-Infant Transmission of Virus. *Journal of Pediatrics* **127**, 278-280.