Molecular virology of hepatitis C virus

Berwyn Clarke

Hepatitis Antiviral Research, Virology Research Unit, GlaxoWellcome Medicines Research Centre, Gunnels Wood Road, Stevenage, Herts SG1 2NY, UK

Introduction

Towards the end of the 1980s a significant number of cases of parenterally transmitted viral hepatitis could not be ascribed to any of the then known hepatic viruses (hepatitis A virus, hepatitis B virus and hepatitis delta virus). The discovery and characterization of a novel RNA virus with characteristics typical of the family Flaviviridae (Choo et al., 1989) provided the tools to study the epidemiological importance of this agent in these cases of non-A non-B hepatitis (NANBH). The virus, known as hepatitis C virus (HCV) was shown to be the cause of most cases of NANBH. Over 50% of patients become chronic carriers (Kuo et al., 1989). The persistent infection commonly results in chronic active hepatitis which may lead to liver cirrhosis and hepatocellular carcinoma (Saito et al., 1990). Random screening of blood donor populations has indicated that worldwide there could be as many as 500 million chronic carriers of the virus (Dhillon & Dusheiko, 1995) and this highlights HCV as a major human pathogen.

Basic virology

Current approaches to the treatment of NANBH are limited to the use of α-interferon (α-IFN) either alone (e.g. Davis et al., 1989) or in combination with other agents such as ribavirin (e.g. Brillanti et al., 1995), but the efficacy of these therapies is extremely poor and post-therapy relapse rate is very high such that sustained biochemical response is observed in less than 20% of patients. Consequently there is an urgent need for the development of novel therapeutics. Unfortunately, approaches to the development of in vitro replication systems for propagation of this virus have been particularly difficult with reliable systems for efficient long-term virus replication being irreproducible. Nevertheless, reports of virus replication have been made using systems from hepatic tissue (Ito et al., 1996) and peripheral blood mononuclear cells (PBMC) (Willems et al., 1994; Zignego et al., 1992) while in vitro infection systems have been reported from human T and B cell lines (Bertolini et al., 1993; Shimizu et al., 1992), human foetal liver cells (Iacovacci et al., 1993) and chimpanzee hepatocytes (Lanford et al., 1994). One of the most promising in vitro infection system so far reported uses Daudi cells (a human B-cell line) (Nakajima et al., 1996) where virus replication was reported to be detectable up to a year post-infection and where there was apparent selection of one particular viral sequence from the quasispecies present in the inoculum, indicative of selection for a cell-culture adapted virus. Very similar data were reported (Mizutani et al., 1995, 1996; Sugiyama et al., 1997) using a human T cell leukaemia virus type I-infected T cell line, MT-2. Again, in this study, there was apparent selection from the viral quasispecies of a single genomic sequence, presumably representing a virus which was naturally adapted for growth in this cell-line. Whether the viruses in these studies, which are apparently adapted for tissue culture growth, can replicate in other cell-lines remains to be established.

This system is one of the few in which the virion particle has been visualized (Shimizu et al., 1996b). The size of the virion had been previously estimated to be < 80 nm by ultracentrifugation (Bradley et al., 1985) while filtration studies suggested 30–60 nm (He et al., 1987). Several studies have shown that the buoyant density of HCV in infectious serum is lower than that in non-infectious serum (Hijikata et al., 1993c; Bradley et al., 1991) and the density (< 1.06 g/ml) suggested an association with the low density lipoprotein (LDL) fraction (Thomssen et al., 1992), which was subsequently verified by Prince et al. (1996). Typically, HCV-infected serum also contains a higher density fraction in which the virus is complexed with anti-viral antibodies (Hijikata et al., 1993c). As would be expected, virus in the lipoprotein fraction appears to be relatively more infectious than the antibody-coated fraction and has also been more amenable to electron-microscopic examination. Using very low density lipoprotein (VLDL)- associated virus, highly purified particles of 60–70 nm have been observed which appear to have prominent (6–8 nm) spikes on their surface (Kaito et al., 1994; Prince et al., 1996), probably representing glycoprotein portions of the viral envelope. The observation of an association between HCV and lipoproteins may be critical for the virus replication cycle because LDL and VLDL are taken up by hepatocytes through the LDL receptor, which may provide a unique and efficient mode of virus entry.

Molecular analysis of the HCV genome

The genomic organization of HCV is illustrated schematically in Fig. 1. Comparative analysis of the genomes of several HCV strains (e.g. Kato et al., 1990; Choo et al., 1991;
Okamoto et al., 1991; Takamizawa et al., 1991) indicated the virus to be closely related to both the pestivirus and flavivirus genera within the family Flaviviridae. The viral genome is a single-stranded RNA molecule approximately 9.5 kb in length which is positive sense and possesses a unique open reading frame, coding for a single polyprotein, flanked by untranslated regions at both its 5' and 3' ends. The length of the polyprotein-encoding region varies according to the isolate and genotype of the virus from 3008 to 3037 amino acids (see Chamberlain et al., 1997). Apart from differences in length HCV genotypes show diversities of around 30% in the nucleotide sequences of their whole genomes and comprehensive analysis of these sequences has revealed the existence of at least 6 genotypes and more than 30 subtypes throughout the world (see Bukh et al., 1995; Simmonds, 1995). The importance of the genomic heterogeneity lies in the fact that some genotypes appear to be associated with more severe pathology and are more refractory to treatment by current therapies (Bukh et al., 1995). However, other studies show no significant association between genotype and disease progression (e.g. Benvegnu et al., 1997).

5' Untranslated region (5' UTR)

An obvious characteristic of HCV is the presence of a long untranslated region (UTR) at the 5' end of the genome and detailed molecular analysis indicates that polyprotein synthesis is initiated at nucleotide 342 (Han et al., 1991). This region of the genome was predicted to be capable of forming extensive secondary structures (Tsukiyama-Kohara et al., 1992) which were biochemically confirmed by Brown et al. (1992). Furthermore, a survey of the sequences within the 5' UTR from 39 different isolates and genotypes of HCV showed that there was remarkable sequence conservation (Bukh et al., 1992). In fact, this region is the most conserved region of the whole genome, a characteristic which has allowed it to be used as a diagnostic marker for HCV by PCR (Smith et al., 1995). Within this region the majority of genotypes of HCV possess five AUG codons which are not used for initiation of translation. In these respects, the presence of a long, highly ordered untranslated region containing non-functional initiation codons closely resembles the situation in the picornaviruses. It is now well documented that picornaviral genomes are translated by a mechanism allowing ribosomes to bind internally on the genome and initiate translation at specific AUG codons (Jackson et al., 1990). These internal ribosomal entry sites (IRES) are dependent on the highly ordered structure of the 5'UTR regions and allow viral genome expression using a cap-independent mechanism.

Because of the obvious similarities with the picornaviral 5'UTR, several laboratories have looked for IRES elements in the HCV genome. Initially, Tsukiyama-Kohara et al. (1992) demonstrated that, using in vitro expression systems, the HCV 5' UTR could regulate translation initiation in a cap-independent manner, indicative of an internal initiation mechanism, and that it was possible to remove up to 101 nucleotides from the 5' end of the RNA without affecting the IRES function. Further studies from Wang et al. (1993) and Kettinen et al. (1994) confirmed the IRES activity but their mapping studies indicated that only 40 nucleotides could be removed, while subsequent work by Fukushi et al. (1994) indicated that RNA starting at nucleotide 62 was inactive as an IRES element. Analysis of the model of the structure of the HCV 5' UTR proposed by Brown et al. (1992) (see Fig. 2) indicates that a stem–loop structure is formed at nucleotide 46 which subsequent deletion mutagenesis experiments have shown to be critical for IRES function (Fukushi et al., 1994; Rijnbrant et al., 1995; Honda et al., 1996b). Although an additional hairpin loop is predicted to be located nearer the 5' end of the RNA (see Fig. 2), this structure is not essential for IRES function and probably fulfills a different replicative function.

In contrast to the general agreement on the presence of a functional IRES in HCV, Han et al. (1991) reported the presence of a subgenomic RNA initiating at nucleotide 145 which, given the highly ordered secondary structure of the full-length molecule, might be necessary to allow conventional scanning translation. This observation has not been substantiated by other groups and, indeed, Honda et al. (1996b) clearly demonstrated efficient IRES-directed translation from full-length genomic RNA using both in vitro systems and transiently transfected cells.

Despite these reproducible reports of IRES activity in the HCV 5'UTR Yoo et al. (1992) were unable to show any IRES effect using both in vitro and in vivo systems. The reasons for this discrepancy were unclear until studies by Kettinen et al. (1994) and Reynolds et al. (1995) indicated that the IRES of HCV was unique in that its 3' boundary did not lie in the 5'UTR but was located within the actual coding sequence of the capsid protein. Detailed mapping studies (Reynolds et al., 1995) have subsequently shown that the optimal IRES activity resides in a segment from nucleotides 40 to 370, thus...
incorporating at least 29 nucleotides from the capsid protein coding region. Extension of the computer-generated model of the 5′ UTR into the coding region indicates that this region is predicted to form a stable hairpin structure as shown in Fig. 2. Since the studies of Yoo et al. (1992) did not maintain the integrity of the 3′ end of the UTR this may have been responsible for the lack of IRES activity reported. Further studies by Honda et al. (1996a) have subsequently shown that an in-frame 12 nucleotide insertion placed within the capsid-coding region 9 nucleotides downstream of the initiator AUG strongly inhibited translation while multiple silent mutations within the first 42 nucleotides had the same effect. However the stem–loop structure in which the AUG is located (stem–loop IV, Fig. 2) is not essential for IRES activity but is able to regulate the efficiency of cap-independent translation both in vitro and in vivo (Honda et al., 1996a).

Apart from the mapping studies delineating the 5′ and 3′ boundaries of the IRES, several groups have reported the requirement for specific structural regions within the element itself. Fukushi et al. (1994) reported that the complete 5′ UTR containing all stem–loop structures was necessary for full IRES function while Rijnbrand et al. (1995) demonstrated that only the most 5′ located hairpin structure could be removed. Subsequent reports by Wang et al. (1994, 1995) and Le et al. (1995) indicated that, as well as simple hairpin–loop structures, more complex tertiary structures such as pseudoknots were also likely to be important.

The requirement for elements of the coding region to be involved in optimal IRES function has not been previously reported in any other RNA. Interestingly, Simmonds et al. (1993) observed that there appeared to be selective pressure on the HCV genome to maintain particular sequences up to the 50th codon of the core protein, including wobble positions. The fact that this area of the genome appears to be involved in highly ordered secondary structure would seem to provide a logical explanation for this extreme sequence conservation. Further studies by Reynolds et al. (1995) also revealed other unusual characteristics of the HCV IRES. In most studies on IRES function, mutation of the authentic initiator AUG residue has a dramatic inhibitory effect on initiation of translation. Systematic mutational analysis of the AUG codons within the putative IRES element of HCV revealed that the authentic initiator AUG at 342 could be mutated to AUU or CUG with very little effect on translation efficiency. Furthermore, Rijnbrand et al. (1996) showed that insertion of additional AUG residues within 8 nucleotides either side of the authentic AUG had no effect on its correct recognition by ribosomes.

The function of the HCV IRES, therefore, seems to be to provide a structure which is able to specifically direct ribosomes to the AUG codon at position 342 for translation initiation. As in other IRES examples, this interaction probably involves complex interactions with cellular protein cofactors and evidence has recently been reported for such specific factors which may critical in regulating translation (Yen et al., 1995; Fukushi et al., 1997).

A consequence of the cap-independent translation provided by the IRES function is that some picornaviruses are able to shut-off conventional cap-dependent scanning translation by proteolysis of eukaryotic translation initiation factor eIF4-F (Lloyd et al., 1985) by viral protein 2A. Although HCV is not known to have any obvious homologue to 2A, it is clear that cap-independent translation from the HCV IRES can occur in the presence of 2A from either Coxsackievirus (Tsukiyama-Kohara et al., 1992) or poliovirus (Rijnbrand et al., 1995). These observations are further indications of functional IRES activity in HCV but the significance of the resistance to inhibition of conventional translation mechanisms is, as yet, unclear.

Recent work has also confirmed the presence of a functional IRES element in the 5′ UTR of the pestiviruses (Poole et al., 1995) which has close structural similarity to the proposed structure of the HCV IRES (Brown et al., 1992; Le et al., 1995), further demonstrating the relationship between these two genera of the family Flaviviridae.

**Virus-encoded proteins**

The nascent viral polyprotein is processed by a combination of host and viral proteinases into the mature viral proteins (e.g. Hiiijikata et al., 1991, 1993b; Grakoui et al., 1993c) (Fig. 3); at
least 10 distinct viral proteins have so far been identified which are arranged as shown in Fig. 1, with the structural proteins located in the amino-terminal one-third and the replicative enzymes located within the carboxy-terminal two-thirds of the polyprotein.

Core

The protein located at the amino terminus of the polyprotein is highly basic in nature and is considered likely to be the viral capsid protein. It is released from the viral polyprotein by nascent proteolytic cleavage at amino acid 191 by host proteases (Hijikata et al., 1991; Ralston et al., 1993). The full-length protein, known as P21, has been identified by both in vitro and in vivo expression (Hijikata et al., 1991; Lo et al., 1994; Harada et al., 1991) but a second species (P19) generated by a secondary cleavage at amino acid 173 is the major product observed following expression in mammalian cells (Santolini et al., 1994; Moradpour et al., 1996). Both P21 and P19 are located in the endoplasmic reticulum (ER) membrane and the conversion of P21 to P19 is presumably mediated by membrane-associated cellular enzymes. A third collinear species of core which can also be detected in expression studies (Lo et al., 1995) is approximately 151 amino acids long (P16) and appears to be localized in the nucleus and more specifically in the nucleolus. The relative expression of P16 varies between different strains of HCV and, perhaps surprisingly, is closely associated with the nature of the amino acid at position 9 of the protein (Lo et al., 1994). The distinct difference in subcellular localization between P16 and P19/P21 suggests that the biological role of P16 may be different. In particular, the nucleolar localization of P16 may be due to its ability to bind to ribosomes, which are assembled in the nucleus (Santolini et al., 1994). Apart from its highly basic nature the core protein also possesses a number of distinct hydrophobic regions (residues 121–151, 170–191) which are involved in the association of P21 and P19 with the ER. It is likely that the carboxy-terminal hydrophobic region (170–191) plays a critical role in the translocation of the viral structural glycoproteins into the ER where the enzymes responsible for peptide processing and glycosylation are located. Recent reports have indicated additional cleavage species at 179 and 182 which are not dependent on prior cleavage at 191 (Hussy et al., 1996). It is currently unclear what roles these diverse core proteins play in virus morphogenesis. Since virion nucleocapsid formation involves multimerization of core protein and its interaction with viral RNA, these characteristics of the HCV core protein have been extensively studied and specific functions mapped to discrete portions of the molecule (Santolini et al., 1994; Matsumoto et al., 1996). The biological functions of the core found in the nucleus, if this also occurs in natural virus replication, are still unclear. Several studies have reported the suppression, by core, of transcription of several host genes as well as interference in expression of co-infecting genomes of hepatitis B and human immunodeficiency viruses (Shih et al., 1993; Srinivas et al., 1996). Perhaps the most interesting recent observations have been that core can specifically suppress apoptotic cell death in artificial systems (Ray et al., 1996) and also specifically interact with the cytoplasmic tail of the lymphotoxin-β (receptor (LTR), a member of the tumour necrosis factor family (Matsumoto et al., 1997). Since LTR is known to be involved in apoptotic signalling this strongly suggests that core may have an immunomodulatory function and play a critical role in the establishment of persistence and in disease pathogenesis.

Finally, a recent report (Barba et al., 1997) shows an association between the core protein and the surface of lipid droplets within the cytoplasm. Analysis of the triglyceride populations within the cell indicates that core protein expression stimulates a change in cellular metabolism of triglycerides. Since a characteristic of HCV infection is liver steatosis it is plausible that this occurs as a result of the direct effect of the core protein on lipid metabolism.

E1/E2

The major viral structural proteins are the glycoproteins E1 and E2, which are released from the viral polyprotein by the action of host-cell signal peptidases. Analysis of the amino termini of both E1 (gp35) and E2 (gp70) indicates that they are cleaved at amino acids 383 and 746 respectively (Hijikata et al., 1991; Grakoui et al., 1993; Mizushima et al., 1994). Both proteins are heavily glycosylated with 5/6 and 11 N-linked glycosylation sites respectively (Miyamura & Matsuura, 1993) and E2 is sometimes found extended at its carboxy terminus to include a smaller protein known as p7 such that multiple E2 species can be observed after expression in eukaryotic systems (Lin et al., 1994a; Mizushima et al., 1994; Selby et al., 1994; Lanford et al., 1993; Grakoui et al., 1993c). Unlike the other cleavages within the structural region, proteolytic cleavage between E2/p7 and p7/NS2 appears to occur post-translationally but the biological function of these diverse E2 species is currently unknown.

Numerous studies have reported interactions between the viral structural and non-structural proteins and it is clear that complex protein–protein association is a critical part of virus...
replication. Although E1 has been shown to associate through its carboxy terminus with the core protein (Lo et al., 1996) and E2 with NS2 (Matsuura et al., 1994; Selby et al., 1994), the focus of most studies have been on the direct interaction between E1 and E2 themselves since this interaction is likely to be a critical step in virus morphogenesis. Several studies have shown convincing evidence that E1 and E2 form complexes (Dubuisson et al., 1994; Grakoui et al., 1993c; Lanford et al., 1993) which are mainly non-covalently associated (Ralston et al., 1993) but in some lysates disulphide-linked complexes have also been reported (Grakoui et al., 1993c). Dubuisson et al. (1994) analysed these complexes and showed that the non-covalent forms were slow-forming stable species while the disulphide-linked forms were rapidly forming misfolded aggregates. Nevertheless, disulphide-dependent folding of both E1 and E2 appears to be a necessary prerequisite to complex formation and kinetic analysis indicates that the slow rate of folding of E1 may be the rate-limiting step in the process (Dubuisson et al., 1996). The efficient folding and assembly of the heterodimeric complexes has been shown to occur within the ER and is dependent on an initial prolonged association with the chaperone calnexin (Dubuisson et al., 1996). These complexes then slowly mature into non-calnexin bound complexes which may represent pre-budding native heterodimers (Deleersnyder et al., 1997).

Apart from the interest in understanding the biological processes of assembly and morphogenesis both E1 and E2 have been extensively studied in terms of antigenic variation and mechanisms of persistence, and are obviously major components of prototype vaccine studies for HCV. E2 represents the most variable region of the HCV genome (Kato et al., 1992; Weiner et al., 1991) and the variation is assumed to be caused by random mutation and selection of mutants capable of escaping from neutralizing antibodies produced in the host. Furthermore, antibodies against E2 correlate with protection by neutralizing antibodies using cell-binding of structural proteins (Rosa et al., 1996) or in vitro culture (Shimizu et al., 1996c). Within the E2 sequence are regions of extreme hypervariability (HVR) which have been the focus of more detailed study and one of these regions know as HVR-1 represents the amino-terminal 34 amino acids within E2, spanning residues 383–414 (e.g. Ogata et al., 1991; Weiner et al., 1991; Kumar et al., 1994). This region has been suggested to be particularly important in HCV neutralization because of its extreme variability and the fact that this variability was not observed in a patient with agammaglobulinaemia even over a period of 2–5 years (Kumar et al., 1993). Antibodies to HVR-1 in patient sera have been shown to neutralize binding of E2 (Zibert et al., 1995) while peptide antibodies raised to this region are able to block infection in tissue culture (Shimizu et al., 1996c) and have shown efficacy in chimpanzee challenge experiments (Farci et al., 1996). Despite this, the extreme heterogeneity in the sequence of this epitope in diverse virus strains means that its utility in a prophylactic vaccine strategy remains doubtful.

NS2

The NS2 protein has been shown to be a transmembrane protein with its carboxy terminus translocated into the lumen of the ER while its amino terminus lies in the cytosol (Santolini et al., 1995). Although immunoprecipitation studies (Grakoui et al., 1993b; Matsuura et al., 1994) have shown that NS2 is closely associated with the structural proteins the biological function of the majority of the NS2 protein is still unclear. However, detailed study of the proteolytic processing of this region of the polyprotein has revealed an unexpected proteolytic function contributed partially by the carboxy terminus of NS2.

By analogy with the genus Flavivirus it would be assumed that the NS3 serine protease was responsible for cleavage at the junction between NS2 and NS3 in HCV. Surprisingly, mutation of the active site Ser-1165 in NS3, although abolishing processing at downstream sites, did not affect proteolytic cleavage at the NS2/NS3 junction. The first report of the existence of a second virus-encoded protease was made by Hiijkata et al. (1993a) who showed that the protease responsible for cleavage at this position comprised elements of both the NS2 and NS3 proteins. Further mapping studies (Hijikata et al., 1993a; Grakoui et al., 1993b) delineated the proteolytic domain responsible for this activity to amino acids 827–1207. This region encompasses the carboxy terminus of NS2 and the amino terminus of NS3 and, therefore, overlaps the active domain of the serine protease. Apart from internal autocatalytic cleavage of itself, no other proteolytic processing functions have been ascribed to this enzyme. It has no known homology to other classes of proteases but has been proposed to be a metalloprotease since its activity can be inhibited by EDTA and stimulated by zinc (Zn²⁺) ions. Mutational analysis of the region surrounding the cleavage site (Hirowatari et al., 1993; Reed et al., 1995) (see Fig. 4) has shown that the NS2/3 protease is remarkably tolerant to amino acid mutations in the region from P5 to P3’, apart from substitutions such as proline which markedly affect the conformation of the substrate. However, Reed et al. (1995) noted a preference for hydrophobic amino acids at the P1 and P1’ positions similar to that shown by neutral metalloproteases. In particular, the preferences markedly resembled those of endopeptidase 24.15, a mammalian enzyme involved in processing of peptide hormones (Erdos & Skidgel, 1989). Despite this similarity there is, currently, no solid evidence to suggest that the NS2/3 protease is a metalloprotease.

Several groups have attempted to identify those regions of the enzyme involved in catalytic activity and mutation of His-952 and Cys-993 in NS2 has shown them to be involved. Nevertheless, other reports have also shown lack of NS2/3 protease activity even with wild-type residues at these
positions (D’Souza et al., 1994; R. M. Elliott, personal communication), indicating a critical role for other residues. Despite the fact that the mechanism of action of the NS2/3 protease during virus replication is an autocatalytic cis-cleavage process, Grakoui et al. (1993b) observed, using co-transfection experiments, that the NS2/3 junction could also be processed in trans. Although such a bimolecular cleavage is likely to be irrelevant to the natural process, this observation has allowed some characterization of the cleavage mechanism. Specifically, Reed et al. (1995) showed that the catalytic unit could be essentially separated into two domains (NS2 and NS3). If either of these domains was inactivated, by specific mutation (e.g., His-952) or truncation, then that molecule, although incapable of self-cleavage, could be processed in trans provided that one of the domains was functional. Therefore, this suggests that, in the trans-cleavage reaction, both molecules must contribute a functional subunit to the catalytic process.

Regardless of the mechanism of catalysis, it is likely that cleavage at the NS2/3 junction releasing the free amino terminus of the NS3 serine protease is a pivotal step in HCV replication. However, in the pestiviruses, virus replication is able to proceed in the absence of cleavage at this position and the resultant viruses exhibit a non-cytopathic phenotype. Furthermore, the HCV NS3 serine protease most closely resembles the pestivirus enzyme, rather than the flavivirus homologue, in that there appears to be no functional requirement for NS2 as a closely associated cofactor. In view of the similarity to the pestivirus process it will therefore be of interest to determine the biological relevance of the recent reports of HCV isolates which appear to be defective in processing at the NS2/3 junction (D’Souza et al., 1994; R. M. Elliott, personal communication).

**NS3**

The region coding for the NS3 protein has been the most extensively studied of the whole genome. The NS3 protein has been shown in numerous studies to be a protein of approximately 70 kDa in size and to possess several diverse biochemical functions. Initial genomic analysis of this region of the molecule showed that, by alignment with similar regions from flavivirus and pestivirus genomes, a conserved series of residues was present in all reported sequences. These residues, His-1083, Asp-1107 and Ser-1165, constitute the catalytic triad present in all characterized serine proteases and the conserved spatial arrangement between these residues also strongly suggested that this region of the NS3 protein did constitute a viral protease. Initial studies to substantiate this hypothesis focused on the use of *in vitro* transcription/translation and transient mammalian expression systems to observe proteolytic processing of the viral polyprotein (Grakoui et al., 1993a; Hijikata et al., 1993a; Bartenschlager et al., 1993; Tomei et al., 1993; D’Souza et al., 1994; Manabe et al., 1994; Eckert et al., 1993). In these experiments site-directed mutagenesis of the catalytic site serine was used as a control to clearly differentiate between NS3-mediated processing and other mechanisms. Studies such as these clearly demonstrated that the NS3 protease was entirely responsible for proteolytic processing of the whole downstream region of the viral polyprotein and that the protease activity resided completely in the amino-terminal one-third of the NS3
mediates proteolysis at the NS3 molecule. The overall proteolytic pathway for the whole viral protein are subtle biochemical differences in the exact mechanisms by which NS3 mediates these cleavages. Firstly, it was observed that cleavage of NS3 from NS4A was a spontaneous rapid autocatalytic event which, under normal conditions, could only be mediated in cis whereas all other cleavages could be carried out in trans if exogenous NS3 was added to substrates containing target cleavage sites. Detailed biochemical and mutagenesis experiments indicated that the point of association between NS3 and NS4A at the molecular level and several studies confirmed the formation of a stable complex between the two species (Bartenschlager et al., 1994; Failla et al., 1994; Lin et al., 1994b). Furthermore, efficient cleavage at the trans sites was dependent on the presence of the NS4A protein itself as a cofactor for NS3 (Bartenschlager et al., 1994; Failla et al., 1994; Lin et al., 1994b). However, the relative dependence on the presence of NS4A seemed to vary according to the particular cleavage site such that cleavage at NS5A/5B was relatively unaffected by the presence of NS4A whereas cleavage at NS4B/5A was absolutely dependent on its presence. These observations suggested a close association between NS3 and NS4A at the amino-terminal region of the NS3 molecule between residues 15 and 22 (Sato et al., 1995; Koch et al., 1996) and involved the central region of NS4A in which residues 22 to 31 were critical (Bartenschlager et al., 1995; Lin et al., 1995; Failla et al., 1995; Shimizu et al., 1996a; Butkiwieicz et al., 1996). The recent elucidation of the crystal structure of the NS3/4A complex has confirmed these observations indicating that NS4A forms an important structural feature of the active complex (Kim et al., 1996). A further unexpected feature of the crystal structure was the fact that a zinc atom appears to have a critical structural role (Kim et al., 1996; Love et al., 1996). This atom is closely bonded to four protease residues at Cys-97, Cys-99, Cys-145 and His-149 (via a water molecule). Although not located close to the active site of the enzyme, mutation of these residues has a dramatic effect on the proteolytic activity (Stempniak et al., 1997). Most of the original studies assessing the proteolytic activity of the NS3 protease used large precursors mimicking the natural substrates and amino-terminal analysis of the processed products revealed the specificity of the enzyme. As can be seen from Fig. 5, comparative analysis of the sequence surrounding the cleavage sites shows functionally important conservation at residues P1/P1′ and an acidic preference at P6 with a significant degree of variation at the other residues particularly on the carboxy side of the scissile bond. Interestingly, there is a significant difference between the P1 residue at the cis site (NS3/4A) which is always threonine and the trans sites which is always cysteine. Subsequent mutagenesis experiments using either natural substrates or synthetic peptides have allowed detailed analysis of the ability of the enzyme to cleave substrates with amino acid changes at specific positions around the cleavage sites and these largely support the observation that residues P1/P1′ and P6 are critical for efficient proteolysis (Kolykhlov et al., 1994; Komoda et al., 1994; Steinkuhler et al., 1996).

Analysis of the remainder of the sequence of the NS3 protein also revealed the presence of other motifs characteristic of NTPase and RNA helicase enzymatic functions. Expression of this portion of the NS3 protein in bacteria has allowed detailed evaluation of the presence of these activities. Initial studies (Suzich et al., 1993; D’Souza et al., 1993) confirmed the presence of the NTPase activity on the carboxy-terminal domain of the full-length NS3 protein and compared the activity with the related pestivirus and flavivirus enzymes. Subsequent studies (Kim et al., 1995; Gwack et al., 1996) confirmed the presence of the RNA helicase activity and also reported detailed evaluation of the biochemical specificity and kinetic characteristics of the enzyme (Preugschat et al., 1996). The purified enzyme was shown to actively bind RNA substrates with a minimal RNA binding size of between 7 and 20 nucleotides and to unwind both RNA/RNA, RNA/DNA and DNA/DNA heteroduplexes. Further analysis (Preugschat et al., 1996; Tai et al., 1996) indicated that the enzyme had a preference for duplex molecules containing single-stranded 3′ regions implying a 3′ to 5′ directionality and also that the enzyme bound preferentially to the poly(rU) sequence near the 3′ end of the viral genome (Kanai et al., 1995; Tanaka et al., 1995). Taken together, all these observations suggest that the HCV NS3 RNA helicase is unique among RNA helicases characterized so far including the related pestivirus enzyme. Further insight into the role that the RNA helicase plays in virus replication has recently been derived through the completion of the first crystal structure (Yao et al., 1997). The structure, refined to 2.1 Å, reveals that the molecule has
distinct NTPase and RNA binding domains and supports a mechanism of helicase activity involving initial recognition of a 3’ single-stranded region on the nucleic acid substrate by a conserved arginine-rich sequence on the RNA-binding domain. The RNA binding domain then appears to undergo a rotational event involved in physically unwinding the helicase and it is likely that this rotation is coupled to NTP hydrolysis through the other domain.

Despite the fact that there are three known catalytic activities in the NS3 protein there is no evidence to suggest that the two domains are separated by proteolysis in vivo. This could mean that there is a functional interdependence between the helicase and protease functions. Recent observations from Morgenstern et al. (1997) have shown evidence to support this in that all three activities of NS3 can be modulated by the addition of polynucleotides to the assay systems. Since these enzymes are likely to be components of a complex replication structure it is reasonable to assume that the interplay between the various domains of NS3, as well as other viral proteins, may be critical in regulating virus replication.

Apart from the reported biochemical functions of the NS3 protein there are also a number of intriguing reports on the possible role of NS3 in disease. In one study (Sakamuro et al., 1995) expression of the NS3 protein in NIH 3T3 cells resulted in cellular transformation and, on subsequent inoculation into nude mice, tumorigenesis. This may be related to a number of other reports (Borowski et al., 1996, 1997) that the NS3 protein is able to specifically interact with the catalytic subunit of protein kinase A. This molecule is involved in intracellular signal transduction processes and so interference by NS3 in these pathways would be expected to have a dramatic effect on normal cellular functions and would be closely associated with pathogenic mechanisms.

**NS4**

The NS4 region of the polyprotein comprises two proteins, namely NS4A and NS4B. Both of these are released from the viral polyprotein by the NS3 serine protease by cis cleavage at the NS3/NS4A and trans cleavage at the NS4A/NS4B and NS4B/NS5A junctions. NS4A is a small protein, approximately 8 kDa in size, and appears to have diverse functions such as anchorage of replication complexes and as a cofactor for the NS3 protease (see previous). Currently, there is no ascribed function for the NS4B protein but it is likely that it plays an integral role within HCV replication complexes.

**NS5A and NS5B**

The NS5 region of the polyprotein is composed of two major proteins, NS5A (p56) and NS5B (p65), which are released as mature products by the action of the NS3 protease in conjunction with NS4A. Early studies suggested that a precursor of NS5A existed (p58) from which a small third protein was released at the amino terminus resulting in mature NS5A (p56) and a small protein fragment NS5C (p2) (Hijikata et al., 1993b). Subsequent work indicated that this was incorrect and that p56 and p58 were both full-length NS5A and that both were phosphoproteins with a degree of hyperphosphorylation reflecting the difference in size between p58 and p56 (Kaneko et al., 1994). Both forms of NS5A are phosphorylated at serine residues and phosphorylation occurs after the mature NS5A protein is released from the polyprotein. Basal phosphorylation has been shown to occur in two regions (2200–2250 and the carboxy terminus of NS5A) and is independent of the presence of any other viral proteins (Tanji et al., 1995b). In contrast, hyperphosphorylation is extremely dependent on the presence of serine residues 2197, 2201 and 2204 and is enhanced by the presence of NS4A as a cofactor (Kaneko et al., 1994), again indicating the multifunctional nature of the NS4A protein (Tanji et al., 1995a). Further analysis has also shown that the NS4A-dependent phosphorylation is dependent on association between NS4A and amino acids 2135–2139 from NS5A (Asabe et al., 1997). Sequence comparison of the regions surrounding the sites of phosphorylation indicates an extremely high level of conservation between different strains of the virus but the biological significance of the phosphorylation is still unclear. However, the sites do not conform to recognized consensus motifs for known serine kinases. This indicates that either an unknown serine kinase is involved or that the protein has autokatalytic kinase activity. The functional role of either form of NS5A is currently unclear although both forms possess nuclear localization signals and are detected in the nuclear periplasmic membrane by immunofluorescence. Similarly, recent studies have reported a similar localization of the NS5B protein, suggesting that NS5A and NS5B may be closely associated components of a membrane-bound replication complex (Hwang et al., 1997). Apart from the probable role of NS5A in the replication cycle, recent evidence has suggested that it may be a critical factor in determining the susceptibility of the virus to treatment with IFN. It was initially reported that IFN sensitivity correlated with mutations within a discrete region of NS5A which was subsequently named the IFN sensitivity determining region (ISDR) (Enomoto et al., 1995, 1996). Subsequent analysis indicated that the likely mechanism by which this occurred was through a direct interaction of NS5A with the IFN-induced protein kinase, PKR, a mediator of IFN-induced antiviral resistance (Gale et al., 1997). Since PKR is a critical factor in the response to IFN (Katze, 1995) its inactivation by NS5A may be a major mechanism by which HCV evades the host immune response.

The sequence of the NS5B protein is highly conserved, not only between different strains of HCV but also in pestiviruses, flaviviruses and even in other RNA viruses. In particular, the amino acid motif G-D-D is totally conserved in HCV, flaviviruses, poliovirus and tobacco mosaic virus (Kamer & Argos, 1984). This motif is a characteristic of all known RNA-
dependent RNA polymerases and so the function of NS5B in HCV has been speculated to be the viral polymerase. In support of this Behrens et al. (1995) recently presented evidence of RNA-dependent RNA synthesis in extracts of baculovirus-infected insect cells expressing HCV proteins. As yet it is unclear whether NS5B alone can perform this role or whether it is merely a critical component of a multimolecular replication complex.

3’ Untranslated region (3 UTR)

A significant recent finding in the HCV area has been the revelation that the considered 3’ terminal region of the genome was incorrect. A number of studies had suggested that the 3’ terminus of the genome terminated in a poly(U) tract (e.g. Kato et al., 1990; Takamizawa et al., 1991; Okamoto et al., 1992) or, in a single report (Han et al., 1991), as a poly(A) tract. However, more detailed analysis of the 5’ end of the negative anti-genomic RNA strand in infected liver cells revealed the presence of a novel 98 nucleotide sequence downstream of the presumed genomic terminus (Tanaka et al., 1995; Kolykhalov et al., 1996). Detailed sequence analysis of the complete new 3’UTR shows that it can be considered as a tripartite structure comprising the conventional 3’ end, a poly(U) tract and the new highly conserved sequence known as the 3’X tail. Interestingly, the poly(U) region appears to be extremely heterogeneous between different virus isolates and even within the same infected liver. In contrast, the new sequence has been shown to be extremely highly conserved even between the two most genetically divergent HCV genotypes, 1B and 2A (Tanaka et al., 1996), although Yamada et al. (1996) have recently reported that some genotype 1B isolates possess an additional two uridine residues at the extreme 3’ terminus. Furthermore, computer modelling has predicted that this sequence can fold into an elaborate stem–loop structure (Kolykhalov et al., 1996) suggestive of a critical functional component in virus replication as shown for many positive-strand viruses. Surprisingly, the only reported infectious clone for HCV (Yoo et al., 1995) represents a virus which lacks this new region. The levels of genomic replication in this report were not high, however, and ligation of this new 3’ end to the clone may significantly improve replicative efficiency.

Conclusion

The identification of HCV as the aetiological agent of NANB hepatitis is certainly the most significant recent development in any area of viral disease. The clinical importance of the disease and the need to rapidly identify new therapeutic approaches have resulted in intensive study of the molecular properties of the virus. Although there are still large gaps in our understanding of some aspects of the virus and its replication this review has attempted to provide a snapshot of the current state of the art.

I would like to thank my colleagues at GlaxoWellcome, both past and present, for valuable contributions to this review, particularly Tony Carroll and Eddy Littler for critically reading the manuscript. Similarly, to Richard Elliott as well as providing access to unpublished information.

References


References


