

Review

Correspondence
Peter Simmonds
Peter.Simmonds@ed.ac.uk

Genetic diversity and evolution of hepatitis C virus – 15 years on

Peter Simmonds

Centre for Infectious Diseases, University of Edinburgh, Summerhall, Edinburgh EH9 1QH, UK

In the 15 years since the discovery of hepatitis C virus (HCV), much has been learned about its role as a major causative agent of human liver disease and its ability to persist in the face of host-cell defences and the immune system. This review describes what is known about the diversity of HCV, the current classification of HCV genotypes within the family *Flaviviridae* and how this genetic diversity contributes to its pathogenesis. On one hand, diversification of HCV has been constrained by its intimate adaptation to its host. Despite the >30% nucleotide sequence divergence between genotypes, HCV variants nevertheless remain remarkably similar in their transmission dynamics, persistence and disease development. Nowhere is this more evident than in the evolutionary conservation of numerous evasion methods to counteract the cell's innate antiviral defence pathways; this series of highly complex virus–host interactions may represent key components in establishing its 'ecological niche' in the human liver. On the other hand, the mutability and large population size of HCV enables it to respond very rapidly to new selection pressures, manifested by immune-driven changes in T- and B-cell epitopes that are encountered on transmission between individuals with different antigen-recognition repertoires. If human immunodeficiency virus type 1 is a precedent, future therapies that target virus protease or polymerase enzymes may also select very rapidly for antiviral-resistant mutants. These contrasting aspects of conservatism and adaptability provide a fascinating paradigm in which to explore the complex selection pressures that underlie the evolution of HCV and other persistent viruses.

Introduction

Since its discovery 15 years ago (Choo *et al.*, 1989; Kuo *et al.*, 1989), hepatitis C virus (HCV) has been the subject of intense research and clinical investigations as its major role in human disease has emerged. Globally, HCV is estimated to infect 170 million people (3% of the world's population) and creates a huge disease burden from chronic, progressive liver disease. HCV has become a major cause of liver cancer and one of the commonest indications for liver transplantation (reviewed by Hoofnagle, 2002; Seeff, 2002; Pawlotsky, 2003b). HCV infection can be treated, but this is costly and requires long-term medical support and follow-up; current therapies are impractical for the majority of HCV carriers worldwide. The development of a protective vaccine remains, at best, a distant prospect.

HCV is an enveloped virus with an RNA genome of approximately 9400 bp in length. Most of the genome forms a single ORF that encodes three structural (core, E1, E2) and seven non-structural (p7, NS2–NS5B) proteins (Fig. 1). Short untranslated regions (UTRs) at each end of the genome are required for replication of the genome, a

process that has recently been found to additionally require a *cis*-acting replication element in the coding sequence of NS5B (You *et al.*, 2004). Translation of viral proteins is dependent on an internal ribosomal entry site in the 5' UTR, which comprises a complex RNA structure element that interacts directly with the 40S ribosomal subunit during translation initiation (Pestova *et al.*, 1998).

HCV is classified in the family *Flaviviridae*, although it differs in many details of its genome organization from the original (vector-borne) members of the family. HCV is additionally distinct and somewhat unusual for an RNA virus in being able to establish persistent infections in the majority of exposed individuals. This phenomenon has attracted the greatest interest in HCV research, not least because long-term, chronic infections underlie its disease manifestations and effective therapy must break this ongoing cycle of replication in the liver. Understanding the mechanism of persistence is also of fundamental immunological interest and, as discussed, represents an important new paradigm in which to explore the genetic basis for this highly adapted interaction with its host.

The evolution of HCV is shaped by distinct selection pressures that are associated with, on one hand, the historical events underlying the adaptation of HCV to its human host

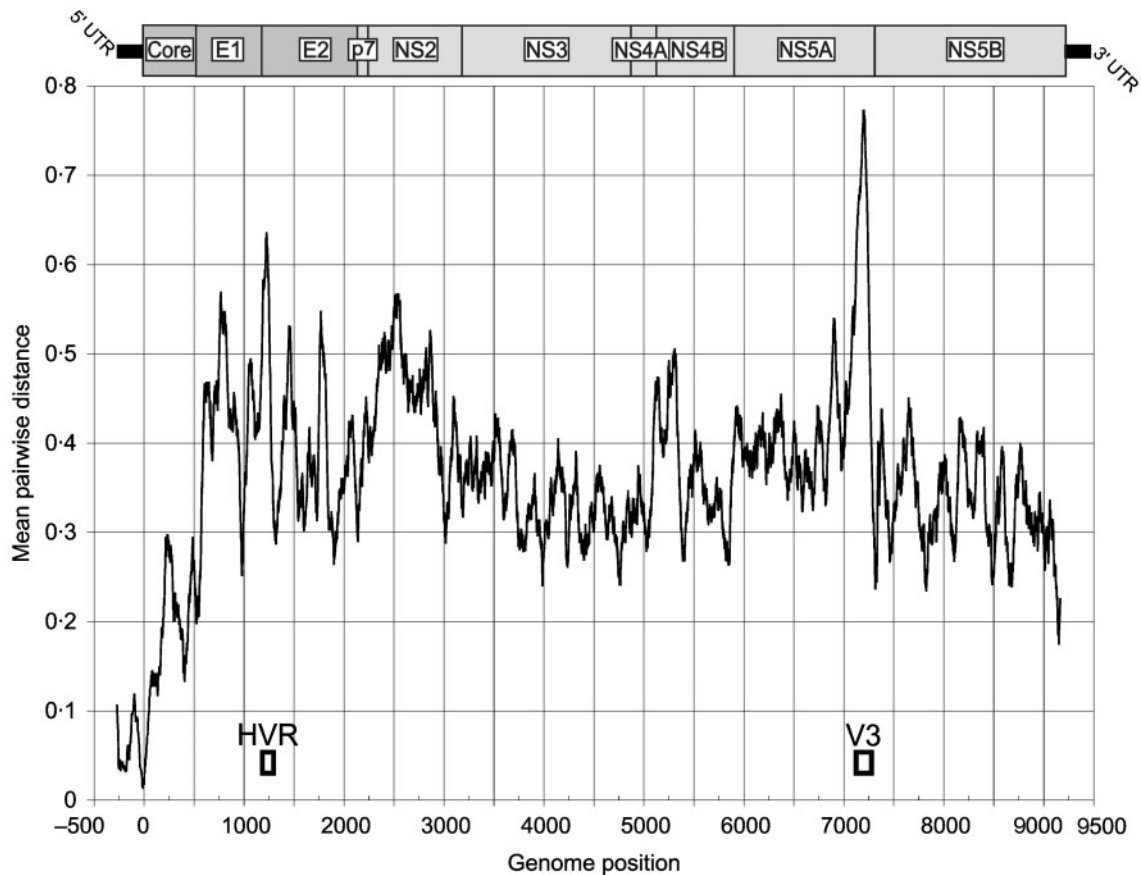


Fig. 1. Diagrammatic representation of the HCV genome (top panel), showing positions of structural and non-structural proteins. The lower panel plots sequence diversity between different HCV genotypes in windows of 150 bp across the genome. This analysis reveals highly conserved regions in the 5' UTR and core regions, and high viral diversity in envelope genes and NS5A.

that have ensured its successful ongoing transmission. On the other hand, HCV is capable of very rapid, adaptive changes that are associated with *de novo* infection of each individual in response to immunological selection pressures to antiviral therapy. HCV also accumulates sequence changes as a result of 'neutral' sequence drift over time and this process, rather than the adaptive changes, accounts for much of the sequence diversity that is observed between its different genotypes. This review will describe the different mechanisms of evolutionary change, their relationship with selection (see the next section) and the importance of neutral and adaptive evolution in the diversification of HCV and its persistence and treatment resistance.

Virus sequence change

The evolution of viruses resembles that of all organisms; it is a process that is ultimately dependent on mutations in their genetic material. In many ways, however, viruses differ from commonly studied organisms such as the geneticist's mouse or fruit fly, particularly in their speed of sequence

change, large population size and the nature of the selection pressures that they encounter.

In popular use, the word 'evolution' describes the process of adaptive change whereby organisms change in their phenotype (such as body shape or behaviour) in response to external, sometimes changing, selection pressures and by competition with other organisms for limited resources in a shared environment. Random mutations from copying errors or chromosomal damage occasionally (and entirely by chance) might improve organism fitness, allowing the mutated gene to spread and, eventually, to predominate in the population where the advantage it confers, in terms of reproductive success, is significant. In this model, the evolution of distinct species of animals, plants and bacteria results from large numbers of incremental changes in phenotype that are associated with adaptation to the wide range of separate contemporary and previous environments.

Surprisingly, this 'Darwinian' type of evolution makes very little contribution to the genetic diversity of organisms when measured at the level of DNA or RNA sequences. Although

highly controversial when proposed (Kimura, 1968; King & Jukes, 1969), neutral theory demonstrates that the majority of sequence change in and between species has no significant effect on phenotype, i.e. it is 'neutral' (Kimura, 1983). Nucleotide changes in coding and non-coding sequences that have little or no effect on organism fitness become fixed in the population by chance. Thus, geographically isolated members of species can become very different genetically, whilst often remaining unchanged morphologically and behaviourally. The frequency of fixation of neutral changes can be predicted to be relatively constant over time and underlies the remarkably close correlation between sequence divergence of certain genes, such as haemoglobin, with the established chronology of splitting of different mammalian species and orders over the past 150 million years.

Evidence for both 'Darwinian' and neutral evolution can be found in the sequences of HCV. One possible example of adaptive change in HCV is the rapid evolution of the hypervariable region of the E2 envelope glycoproteins to prevent recognition by antibodies that are induced by infection (see the section entitled 'Sequence variability within genotypes'). In contrast, 'neutral' sequence drift undoubtedly accounts for much of the genetic diversity that is observed between geographically or epidemiologically separated populations of HCV. This process of divergence resulting from the fixation of neutral sequence changes does not alter the phenotype of the viruses greatly. Despite the >30% sequence difference that is observed between genotypes of HCV (see the section entitled 'HCV genotypes'), each retains a similar replication cycle in human hosts. Indeed, their shared abilities to establish persistent infections in humans with high infectivity titres in blood and to cause only slowly progressive and largely asymptomatic infection are key factors in their ongoing transmission. This lack of phenotypic innovation over an extremely long period of divergent evolution demonstrates, perhaps, how the evolution of HCV is shaped and constrained entirely by its close adaptation to the particular ecological niche it inhabits, the human liver.

HCV genetic diversity and genotypes

Genetic variability of HCV exists at several different levels. Most obvious is the substantial genetic divergence of the main genotypes of HCV, which frequently show specific geographical ranges in the human population and associations with particular risk groups for infection. Below this and arising from sequence drift over a much shorter period is the variability that is observed between individual variants (or strains). Much of the sequence diversity that is observed between such strains (such as the 5–8% divergence observed between variants in epidemiologically unlinked infections by HCV genotypes 1a, 1b and 3a) reflects processes of neutral sequence drift over time after the introduction of HCV into new risk groups in the 20th century. Some of the sequence divergence may represent

phenotypically selected changes that are associated with adaptation for replication in individuals with different immune responses to infection (see the section entitled 'Sequence variability within genotypes'). Finally, HCV diversifies measurably within an infected individual over time, forming what has been described as a 'quasispecies'. This pre-existing genetic variability, combined with an extremely large replicating population size of HCV in a chronically infected individual, provides a large pool of genetic variants that can adapt to new selection pressures, such as immunological recognition and antiviral treatment.

HCV genotypes. Comparison of nucleotide sequences of variants recovered from infected individuals in different risk groups for infection and from different geographical regions has revealed the existence of at least six major genetic groups. On average over the complete genome, these differ in 30–35% of nucleotide sites, with more variability concentrated in regions such as the E1 and E2 glycoproteins, whereas sequences of the core gene and some of the non-structural protein genes, such as NS3, are more conserved (Fig. 1). The lowest sequence variability between genotypes is found in the 5' UTR, where specific sequences and RNA secondary structures are required for replication and translation functions.

Despite the sequence diversity of HCV, all genotypes share an identical complement of collinear genes of similar or identical size. However, contrasting with this general observation is the marked variation in their capability to express a further protein that is generated by a translational frameshift at codon 11 of the core gene (Walewski *et al.*, 2001; Xu *et al.*, 2001; Varaklioti *et al.*, 2002); both the frameshift site and potential size of this novel coding sequence are very poorly conserved between and within genotypes. This contrast with the evolutionarily conserved nature of so many other aspects of HCV replication supports the idea that this 'gene' is more likely to be a computational artefact that has arisen from RNA structure-imposed constraints on third-codon position variability in the core gene (Tuplin *et al.*, 2004).

Each of the six major genetic groups of HCV contains a series of more closely related subtypes that typically differ from each other by 20–25% in nucleotide sequences, compared with the >30% divergence between genotypes (Fig. 1; Simmonds *et al.*, 1993). Some, such as genotypes 1a, 1b and 3a, have become distributed very widely as a result of transmission through blood transfusion and needle-sharing between infecting drug users (IDUs) over the past 30–70 years and now represent the vast majority of infections in Western countries (Fig. 2). These are the genotypes that are encountered most commonly in the clinical setting and for which most information has been collected on response to interferon (IFN) and other antiviral treatments (see the section entitled 'Biological differences').

A different pattern of sequence diversity is observed in parts of Africa and South-East Asia. Here, there are close

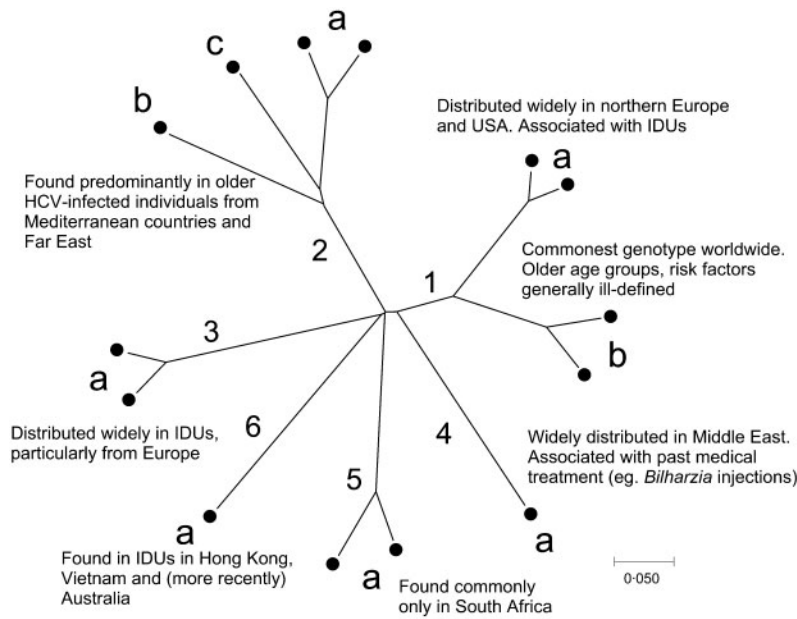


Fig. 2. Evolutionary tree of the principal genotypes of HCV that are found in industrialized countries and their main epidemiological associations with specific risk groups. These genotypes of HCV are believed to have become prevalent over the course of the 20th century. The tree was constructed by using the neighbour-joining method as implemented in the MEGA package, using Jukes–Cantor-corrected distances between complete genome sequences.

associations between genotypes and specific geographical regions (Fig. 3). For example, infections in western Africa are caused predominantly by HCV genotype 2 (Mellor *et al.*, 1995; Ruggieri *et al.*, 1996; Jeannel *et al.*, 1998; Wansbrough-Jones *et al.*, 1998; Candotti *et al.*, 2003), whereas those in central Africa, such as the Democratic Republic of Congo and Gabon, are caused by genotypes 1 and 4 (Bukh *et al.*, 1993; Stuyver *et al.*, 1993; Xu *et al.*, 1994; Fretz *et al.*, 1995; Mellor *et al.*, 1995; Menéndez *et al.*, 1999; Ndjomou *et al.*, 2003). In both regions, there is a remarkable diversity of

subtypes; for example, 20 of 23 HCV-seropositive blood donors in Ghana (western Africa) were infected by genotype 2, but each corresponded to a different and previously undescribed subtype (Candotti *et al.*, 2003). This diversity is reproduced in Guinea, Benin and Burkina Faso (central/western Africa), where 18 different subtypes of genotypes 1 and 2 were found in samples from 41 HCV-infected individuals (Jeannel *et al.*, 1998). These field observations reflect both the huge genetic diversity of genotypes 1, 2 and 4 and, also, its probable long-term presence in human

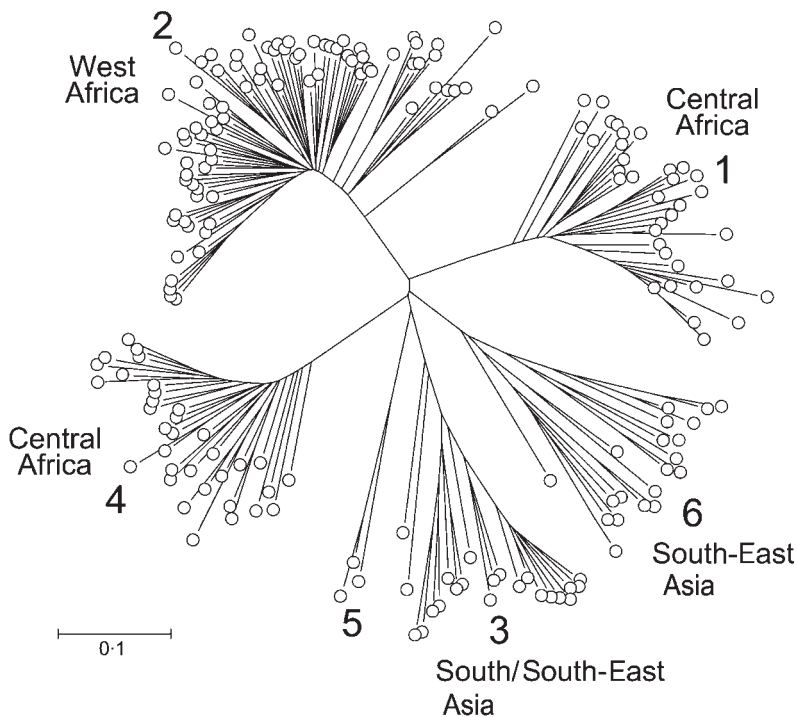


Fig. 3. Evolutionary tree of all known subtypes and genotypes of HCV, including those found in high-diversity areas of genotypes 1, 2 and 4 (sub-Saharan Africa) and 3 and 6 (South-East Asia). HCV variants still fall into six distinct clades, but with far greater numbers of genetic variants corresponding to subtypes in industrialized countries. The tree was constructed by using the neighbour-joining method as implemented in the MEGA package, using Jukes–Cantor-corrected distances between partial NS5B sequences (320 bp).

populations in these parts of Africa. Taking this geographical mapping further, genotypes 3 and 6 show similar genetic diversity in southern and eastern Asia (Tokita *et al.*, 1994a, b, 1995; Mellor *et al.*, 1995).

The model that is suggested by these genotype distributions is that HCV has been endemic in sub-Saharan Africa and South-East Asia for a considerable time, and that the occurrence of infection in Western and other non-tropical countries represents a relatively recent emergence of infection in new risk groups for infection (Simmonds, 2001; Ndjomou *et al.*, 2003). In the 20th century, parenteral exposure to blood-borne viruses became frequent through the widespread adoption of blood transfusion since the 1940s, the medical use of often unsterilized needles for injections and vaccinations (a practice that continues in many developing countries) and, most specifically, to industrialized countries, injecting drug use and the sharing of injection equipment. These new routes for transmission plausibly account for the epidemiological and genetic evidence for recent epidemic spread of HCV over the past 50 years in Europe, Egypt and elsewhere (Pybus *et al.*, 2001, 2003; Cochrane *et al.*, 2002; Ndjomou *et al.*, 2003).

However, one of the puzzles about the origins of HCV is the absence of obvious transmission routes in those areas where the greatest genetic diversity is observed. Transmission by either sexual contact or from mother to child is inefficient (Wasley & Alter, 2000; Pradat & Trepo, 2000; Thomas, 2000) and there is little historical evidence for the type of widespread parenteral exposure that fuelled the epidemic in Western countries. However, recent field investigations in southern Burkino Faso (central/western Africa), where genotypes 1 and 2 are prevalent and highly diverse in sequence, have shown associations between HCV infection with previous sexually transmitted diseases (STDs), circumcision, excision and scarification practices (D. Jeannel, personal communication). Although the association with STDs has not been documented in studies carried out in Western countries, it is possible that lack of mucosal integrity during STD episodes may facilitate the entry of HCV into the genital tract. Determining how long HCV has been in these 'endemically infected' populations would clearly be of value in understanding its epidemiology in these regions (see the section entitled 'Genotype origins').

Even less is known about the earlier divergence of the six major genotypes of HCV, the origins of infection in humans and the underlying basis of the current geographical distribution of genotypes. Areas where HCV is endemic and highly diverse correspond closely to those where hepatitis B virus (HBV) is also prevalent (reviewed by Simmonds, 2001) and also represent regions where human and ape population ranges overlap. However, in contrast to HBV, there is currently no evidence that HCV or HCV-like variants infect Old World ape or monkey species (Makuwa *et al.*, 2003). Therefore, despite tempting analogies with the introduction and spread of human

immunodeficiency virus type 1 (HIV-1) and HIV-2 infections in humans through cross-species transmission of simian viruses from chimpanzees and mangabeys (Gao *et al.*, 1992, 1999), it would be highly speculative and currently unjustified to imagine that HCV originated in these human populations as a result of similar cross-species transmission. On the other hand, it has been discovered that a very distantly related, HCV-like virus, GB virus B (Simons *et al.*, 1995), infects tamarins and/or other New World primates. The existence of this homologue in such a distantly related primate species certainly allows for the possibility that HCV or HCV-like viruses may indeed be distributed more widely in primates than was thought previously. Larger-scale serological and PCR-based surveys of a much larger range of primates in Africa, Asia and South America are required to resolve this issue.

Sequence variability within genotypes. Several studies have described the rapid sequence drift of HCV over time, a process of diversification that leads ultimately to the existence of identifiably separate strains or isolates within human populations. By comparing HCV sequences from sequential samples from chronically infected individuals or from those infected by a common source, rates of sequence change were measured to be 1.44×10^{-3} nucleotide changes per site per year over the whole genome, or 4.1 and 7.1×10^{-4} changes per site per year in the NS5 and E1 regions, respectively (Okamoto *et al.*, 1992; Smith *et al.*, 1997). In coding regions of the genome, changes occur predominantly at synonymous sites (sites that do not alter the encoded amino acid) and are therefore likely to represent the accumulation of phenotypically neutral changes. The expectation from neutral theory that such diversification should occur at a constant rate over time is implicit in attempts to use this rate to estimate times of spread of HCV in specific transmission networks, such as IDUs (Pybus *et al.*, 2001, 2003; Cochrane *et al.*, 2002), and, indeed, its extrapolation to calculating times of introduction of specific genotypes, such as 1a, 1b, 3a and 4a, into new risk groups for infection in Western countries (Smith *et al.*, 1997). For example, the current sequence diversity and phylogenetic tree structure of genotype 4a in Egypt is compatible with the introduction of HCV into that population through parenteral treatment for schistosomiasis (*Bilharzia*) (tragically, with non-disposable and poorly sterilized needles) in the 1950s and 1960s (Frank *et al.*, 2000; Ray *et al.*, 2000; Pybus *et al.*, 2003). The increasing sequence diversity within genotypes 3a, 1a, 1b, 2a and 2b, respectively, suggests times of introduction of these viruses at increasingly earlier times in the 20th century; this may have been associated with other parenteral risk factors for infection, such as injecting drug use, blood transfusion, large-scale immunizations and syphilis treatment (Mortimer, 1995).

Whilst regions of the genome such as NS5B have been used frequently for epidemiological reconstruction, other parts of the genome, such as the 'hypervariable' regions

(HVRs) of E2 and NS5A, show much greater variability and much more rapid amino acid sequence change over time. This variability may arise through specific selection (Darwinian) mechanisms operating on the virus that are associated with immune escape; for example, the HVR in E2 may be a target for neutralizing antibody and persistence may therefore require continuous virus sequence change to evade B-cell responses (Weiner *et al.*, 1992; Kumar *et al.*, 1993; Taniguchi *et al.*, 1993; Farci *et al.*, 2000; Kantzanou *et al.*, 2003). For HIV-1, it is known that initial infection is accompanied by a number of amino acid changes in class I binding motifs in potential T-cell epitopes, such as the Arg→Lys or Gly change in two different immunodominant epitopes in p24^{gag} that are recognized by the B27 allele (Kelleher *et al.*, 2001) and by B58 and B5801 (Leslie *et al.*, 2004). In the latter case, it has been demonstrated that the immune selection mediated by B8 and B5801 occurred only with likely significant fitness cost to the virus; it invariably reverted to the original, 'wild-type' sequence on transmission to other individuals with different human leukocyte antigen (HLA) types.

By analogy, it is therefore possible that many of the amino acid polymorphisms that are observed in HCV are also driven sequentially by selection from different major histocompatibility complex class I or II alleles or by antibody recognition that is encountered during a virus's passage through human populations. Indeed, direct evidence for the occurrence of immune-selected changes in cytotoxic T-lymphocyte (CTL) epitopes has been obtained on experimental infection of chimpanzees (Erickson *et al.*, 2001). Also supporting this hypothesis is the observation that sequence change was slower in individuals with defects in T- or B-cell immunity (Booth *et al.*, 1998), interpreted as indicating reduced immune selection on CTL or B-cell epitopes. Indeed, recovery from infection is associated with strong and sustained CTL responses around the time of primary infection (Cooper *et al.*, 1999; Lechner *et al.*, 2000; Thimme *et al.*, 2001), a time where there is evidence for specific changes in CTL epitopes and accelerated sequence change in the coding sequence in those who become chronic carriers (Chang *et al.*, 1997; Cantaloube *et al.*, 2003; Sheridan *et al.*, 2004). Immune selection may also underlie the high degree of sequence variability between and within genotypes in the NS5A region (Fig. 1) and lead to differences in the ability of different variants and genotypes to evade intracellular defences (see the section entitled 'Interaction with the host cell').

'Quasispecies'. The processes of neutral and adaptive evolution of HCV operate during the course of chronic infection within an individual, leading to both continued fixation of nucleotide changes over time and the development of variable degrees of sequence diversity within the replicating population at a given time point. Sequence diversity is generated continually during virus replication, as RNA copying by the virally encoded RNA polymerase (NS5B) is error-prone and the replicating population is so

large. For example, ongoing error rates of between 1 in 10 000 and 1 in 100 000 bp copied, which are typically found for RNA polymerases (reviewed by Domingo *et al.*, 1996; Drake *et al.*, 1998), combined with a rate of virus production of up to 10^{12} virions per day (Neumann *et al.*, 1998), would produce a highly genetically diverse population of variants, containing mutants that differed at every nucleotide position and every combination of paired differences from the population mean or consensus.

Even though the consensus sequence may be close to the fitness peak at any one time, the existence of a large and diverse population would allow rapid, adaptive (Darwinian) changes in response to changes in the replication environment. This might take the form of evolving immune responses that select against viruses with specific T- or B-cell epitopes; it might also confer resistance to antiviral agents. The rapid and reproducible independent appearance of specific amino acid changes that are associated with the acquisition of HIV-1 resistance to reverse transcriptase and protease inhibitors is a dramatic demonstration of Darwinian evolution of the 'quasispecies'. In the future, this may be reproduced in HCV infections that are treated with the new generation of protease inhibitors (such as BILN 2061) and RNA polymerase inhibitors (Lamarre *et al.*, 2003; Pause *et al.*, 2003; Trozzi *et al.*, 2003; Lu *et al.*, 2004; Sarisky, 2004).

Recombination. Recombination occurs in many families of RNA viruses, its occurrence requiring both epidemiological opportunity and biological compatibility. In positive-stranded RNA viruses, recombination generally occurs through a process of template-switching during RNA genomic replication. To detect such occurrences, a single cell must be infected with two or more genetically identifiable variants of the virus. *In vivo*, this requires both coinfection of the same individual with more than one such variant and substantial overlap in their geographical distributions, in order to enable recombinant forms to be detected.

The genotype epidemiology and natural history of infection with HCV clearly fulfils both of these criteria. A wide range of genotypes circulates in the main risk groups for HCV in Western countries, including 1a and 3a in IDUs and 1b, 2a–2c and 4a throughout the Mediterranean area. In these areas, infection is often characterized by multiple exposures around the time of primary infection, such as frequently repeated needle-sharing with several infected individuals over short time-intervals in the case of IDUs and the contamination of blood products, such as factor VIII clotting factor concentrates, with multiple HCV-positive plasma units. Indeed, even ongoing, chronic HCV infection does not protect from reinfection in experimentally challenged chimpanzees (Farci *et al.*, 1992) or in HCV-contaminated blood or blood-product recipients, such as thalassaemics and haemophiliacs (Kao *et al.*, 1993; Jarvis *et al.*, 1994; Lai *et al.*, 1994).

The viability and pathogenicity of inter- and intra-genotype recombinants are more difficult to assess and are likely to vary considerably between virus families. Amongst the monopartite, positive-stranded RNA viruses, recombination is best-documented in the family *Picornaviridae*. Recombination between different enteroviruses in species B is known to generate a huge number of naturally occurring recombinant viruses with novel combinations of capsid (serotype-determining) and non-structural proteins (Santti *et al.*, 1999), which can show subtle differences in pathogenicity in mouse models (Harvala *et al.*, 2002). Whilst the coding regions of these evidently compatible enterovirus B sequences differ by approximately 25 % (nucleotide) or 9 % (amino acid), recombination between enterovirus species (which differ by 40–42 % in nucleotide sequence and by 40–45 % in amino acid sequence) are not observed, possibly because these sequences differ too much to be biologically compatible.

There is little experimental information on the potential viability of inter- or intra-genotype recombinants of HCV, although it has recently been observed that most combinations of genotype 1a and 1b sequences in the non-structural region of the genome fail to generate a viable replicon (Gates *et al.*, 2004), implying the existence of incompatibilities between variants that show approximately 20 % sequence divergence. Despite these *in vitro* observations, recombinant forms of HCV have been observed in nature, including a variant formed from structural genes of genotype 2 and non-structural genes from genotype 1b that was found in infected IDUs in St Petersburg, Russia (Kalinina *et al.*, 2002, 2004), and a possible 1a/1b recombinant in Peru (Colina *et al.*, 2004).

Despite the number of studies that have been carried out to investigate this issue, the true frequency of recombination of HCV may have been considerably underestimated. For example, recombination would not be detected easily between variants of the same subtype (such as between two infecting genotype 1a strains in an IDU). Similarly, it would be difficult to document inter-subtype recombinants where HCV is highly diverse, such as within genotype 2 in western Africa, because in such regions, we lack a full catalogue of sequence variants within which to observe recombination events.

The existence of widespread recombination would place a considerable limitation on the use of genotyping assays that are based on single genome regions, such as the 5' UTR or core gene. Should more recombinant viruses emerge in the coming years as a result of increasing geographical overlap in genotype distributions, this would cause major problems in the interpretation of genotyping assay results. For recombinant viruses, only those assays that genotyped samples in regions of the genome that determined IFN susceptibility would be able to predict treatment response, which is one of the main applications of genotyping assays (see the section entitled 'Interaction with the host cell').

Classification. The *Seventh Report of the International Committee on Taxonomy of Viruses* (ICTV) currently classifies HCV and GB virus B as members of the genus *Hepacivirus* within the family *Flaviviridae* (Heinz *et al.*, 2000). The six main genotypes of HCV have been designated 'clades', reflecting their phylogenetic group (Fig. 3). As proposed previously, based on this type of phylogenetic analysis (Mellor *et al.*, 1995; Mizokami *et al.*, 1996; de Lamballerie *et al.*, 1997), 'new' genotypes (such as 7a, 8a, 9a and 11a) have been assigned as members of clade 6 and genotype 10a has been assigned to clade 3.

The ICTV report acknowledges the existence of the extremely large number of subtypes within the main HCV genetic groups and the difficulty in establishing criteria for their classification and nomenclature (Fig. 3). Indeed, in the future, it might be more appropriate to regard the subtype tier in previously published HCV classification proposals as being of significance only in Western countries and other regions where HCV has spread recently. Here, discrete subtypes are identifiable, as they are descendants of founder viruses that have been seeded into these new transmission networks (as exemplified by those analysed in Fig. 2).

The existence of identifiable subtypes within the main clades of HCV is, therefore, an epidemiological phenomenon that is associated with recent spread. Thus, there seems to be little justification for the continued efforts to catalogue and name individual subtypes (up to the letter 'r' in the case of genotype 4; http://hcv.lanl.gov/content/hcv-db/combined_search/search) in high-diversity areas, such as central Africa. In my opinion, future assignment of subtypes is only worth pursuing where it is epidemiologically relevant, and should therefore be restricted to those that are distributed widely and show specific geographical, risk group or other epidemiological associations.

Another remaining classification issue is the procedure for identifying and naming inter- or intra-genotype recombinant viruses. For HIV-1, designation of inter-subtype recombinant viruses [or circulating recombinant forms (CRFs)] requires the detection and complete genome sequences of the recombinant in three or more individuals with no epidemiological connection and the demonstration of recombination breakpoints in identical positions in each sequence. For nomenclature, each CRF is numbered sequentially in order of discovery, followed by subtype identification letters to indicate its approximate composition. This procedure might be adapted easily for HCV, in which case the recombinant virus circulating in St Petersburg, Russia (Kalinina *et al.*, 2002, 2004), would officially be designated CRF01_2k1b (or RF1_2k/1b as proposed by the authors).

Genotype origins

It is difficult to estimate the length of time that HCV has been present in human populations. As described above,

the diversity of variants within genotypes 1, 2 and 4 in sub-Saharan Africa and of genotypes 3 and 6 in South-East Asia suggests that HCV may have been endemic in these populations for considerably longer than in Western countries. As the evolutionary process of sequence divergence that led to the diversity of subtypes in these regions is likely to have been predominantly neutral in mechanism, it may therefore be possible to calculate the times of splitting of subtypes and, possibly, also the times of divergence of the six main clades of HCV through use of published rates of sequence change over time (Okamoto *et al.*, 1992; Smith *et al.*, 1997).

Extrapolation of these rates to time the 20 and 30% sequence divergence that is observed between subtypes and genotypes, respectively, produces relatively recent times of origin that, in many ways, are difficult to reconcile with the epidemiology of HCV and its global distribution. For example, the diversity of variants observed in west African genotype 2 sequences predicts a time of origin for this endemic pattern of infection of approximately 200–250 years ago, whilst different genotypes would have diverged from each other about 100 years earlier. Even by using complex methods for correction for multiple substitutions and allowing rate variation between sites, the current diversity of genotypes predicts an origin no earlier than 1000 years ago (Smith *et al.*, 1997). This seems to be too recent for such a widely distributed virus that infects often relatively isolated human populations in equatorial Africa and South-East Asia.

We have argued that there may be far greater constraints on sequence change in HCV genomes than are found in eukaryotic and prokaryotic gene sequences, on which the neutral theory was first developed and where a molecular clock has been shown to operate over long periods of evolution. We have discovered that the genome of HCV is highly ordered, forming complex RNA secondary structures throughout the coding sequence of the genome (Simmonds & Smith, 1999; Tuplin *et al.*, 2002; Simmonds *et al.*, 2004). This characteristic, termed ‘genome-scale ordered RNA structure’ (GORS), is a conserved feature of several genera and/or families of positive-stranded RNA viruses that infect animals and plants. Although the function of GORS is unknown, its correlation with host persistence raises the intriguing possibility of its role in the modulation of recognition or inhibition of innate cell-defence recognition or effector mechanisms that depend on the detection of double-stranded (ds)RNA (Simmonds *et al.*, 2004). The requirement for base-pairing in such structured viruses severely limits the number of ‘neutral’ sites in the genome, as most sequence changes, even at synonymous sites, would disrupt RNA folding. Given the complexity and large scale of these HCV RNA secondary structures, truly ‘neutral’ sites, where sequence changes have no significant effect on virus phenotype, may be rare indeed.

These limitations on sequence change can be illustrated

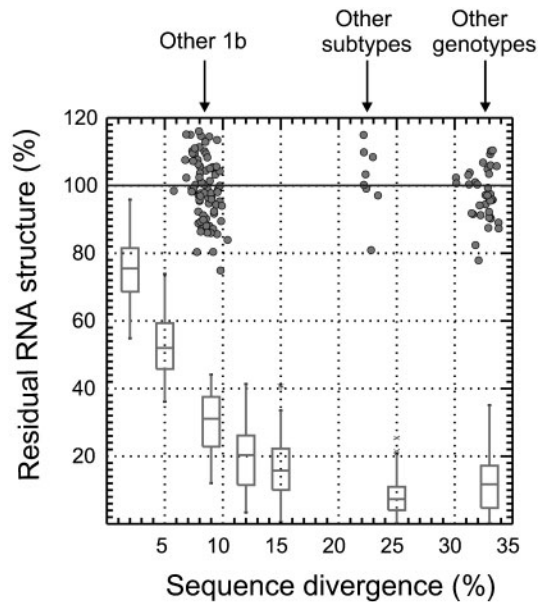


Fig. 4. Effect of simulated ‘neutral’ sequence drift on the formation of RNA secondary structure in the HCV genome [modified from Simmonds *et al.* (2004)]. A genotype 1b sequence was mutated through the introduction of nucleotide substitutions that reproduced naturally occurring variability in HCV (such as transition/transversion ratio, synonymous/non-synonymous substitution ratio and base composition) to create variants with 2–35% divergence from the original sequence (*x* axis). The formation of RNA structure in the mutant sequences was compared in extent to that of the native sequence by using MFOLD (*y* axis; expressed as a percentage). Mutants differing by as little as 2% from the original sequence showed evidence for disruption of RNA structure formation (boxed symbols); introduced sequence drift of >10% produced mutants that were no more capable of RNA structure formation than sequence order-randomized controls. In contrast, the diversification of naturally occurring variants of HCV differing by up to 33% from the 1b sequence (round symbols) preserved RNA structure, with equivalent levels found in other 1b variants, other genotype 1 subtypes and other genotypes. Thus, RNA structure in HCV is highly evolutionarily conserved and its evident requirement for internal base-pairing must constrain even its very short-term evolution.

dramatically through simulation of constrained neutral drift on HCV sequences and measurement of its effect on retention of RNA structure (Fig. 4; Simmonds *et al.*, 2004). The coding sequence of a genotype 1b variant was mutated by using an algorithm that introduced random changes into the sequence, but preserved specific characteristics of naturally occurring virus diversity within HCV. Despite the close simulation of expected neutral evolutionary drift, mutated sequences showed marked and progressive reductions in RNA structure, which was apparent even in sequences that differed by only 2% from the original sequence (Fig. 4). As GORS is conserved in all genotypes of HCV, this loss of RNA structure clearly does

not occur during the natural evolution of HCV. Pathways followed during virus diversification over time that retain GORS must therefore be extremely constrained, and lead to substantial homoplasy and sequence convergence in the limited number of sites where substitutions can occur without damaging RNA structure. The measurable loss of GORS in sequences that have been drifted artificially by 2% indicates that even the very recent evolution of HCV, such as within the Irish 'anti-D cohort' of women that was used for measuring the rate of HCV sequence change (see the section entitled 'Genotype origins'), is subject to the same severe constraints.

Applying a molecular clock to extrapolate times of origin of more divergent HCV variants, such as subtypes and genotypes, is clearly pointless, as the number of neutral sites or the limitations on sequence change at variable sites is not known, so there is no denominator with which to calculate and correct for multiple substitutions. The constriction of sequence space of viruses such as HCV with GORS implies that many of the branches that are evident on phylogenetic analysis of contemporary sequences that define virus species, genotypes or genera occurred at remote times in the past. In making the molecular clock-based estimates above of 350–1000 years for the time of divergence of genotypes, we are therefore in danger of telescoping a much longer period of virus evolution into an unrealistically short time-frame.

A much longer time perspective on HCV evolution, provided by our understanding of GORS constraints, fits much better with the globally distributed nature of HCV and the concentration of specific genotypes with historically relatively isolated populations in sub-Saharan Africa and south Asia. As a potential comparison, GORS in the widely distributed human virus hepatitis G virus/GB virus C appears to have restricted sequence drift to 11–13% nucleotide sequence divergence over the course of evolution of modern humans over the last 100 000–150 000 years (González-Pérez *et al.*, 1997; Pavesi, 2001; Simmonds, 2001). The greater sequence diversity between HCV genotypes implies times of origin that occurred even further back in the evolution of humans.

Biological differences

The major features of HCV structure, replication, transmission and ability to establish persistent infection are shared between all known variants. Indeed, viewed purely as a survival machine, the widespread distribution of genotypes 1–6 in human populations indicates that that each is equally successful in maintaining infections in human populations. Despite this obvious evidence for phenotypic similarity, there is growing evidence for genotype-specific differences in persistence and interactions with innate cell defences and the immune system that have important repercussions for current and probable future therapy.

Treatment response. Beginning with observational data,

the clearest difference between genotypes is in their susceptibility to treatment with IFN monotherapy or IFN/ribavirin (RBV) combination therapy. Typically, only 10–20 and 40–50% of individuals infected chronically with genotype 1 HCV on monotherapy and combination therapy, respectively, exhibit complete and permanent clearance of virus infection. This long-term response rate is much lower than the rates of 50 and 70–80% that are observed on treatment of HCV genotype 2 or 3 infections (reviewed by Pawlotsky, 2003a; Zeuzem, 2004). This difference has proved to be highly significant in patient management and has led to the use of higher doses and longer durations of treatment for type 1 (and type 4) infections, in order to achieve acceptable efficacy. In numerous multivariate analyses, genotype-specific differences in treatment response have been shown to be independent of host variables, such as stage of disease progression, age, duration of infection, sex and HIV and other virus co-infections. It is similarly independent of virus-specific factors, such as pre-treatment viral load, although this also correlates independently (inversely) with response.

Despite this wealth of observational data, we still lack basic understanding of the mechanism of these differences, mainly because the *in vivo* mechanism of action of exogenous IFN or RBV remains largely unknown. Insights into mechanisms of treatment resistance might be obtained through investigation of the inhibitory effect of IFN or IFN/RBV on the *in vitro* replication of subgenomic or full-length genomic replicons of HCV (Lohmann *et al.*, 1999; Ikeda *et al.*, 2002; Pietschmann *et al.*, 2002; Blight *et al.*, 2003). Replication of the replicon can be inhibited by the addition of exogenous IFN (Blight *et al.*, 2000; Frese *et al.*, 2001; Lanford *et al.*, 2003), at least in part through inhibition of translation (Wang *et al.*, 2003). This model has, however, provided only very limited information on treatment resistance, mainly as a result of poorly understood current limitations of the model system. The range of HCV variants that can be cultured is extremely restricted (limited to genotypes 1a and 1b), which are both equivalently IFN-resistant clinically, although a full-length replicon of the more clinically sensitive genotype 2a has recently been described (Kato *et al.*, 2003). Secondly, their *in vitro* replication requires or is enhanced by 'adaptive' amino acid changes in NS5A and NS3 (Bartenschlager *et al.*, 2003), even though these play no role in natural infections and actually attenuate replication in experimentally infected chimpanzees (Bukh *et al.*, 2002). Mutations in NS5A are particularly problematic, as they cluster in a region of the protein that is associated clinically with resistance to IFN therapy and that interacts with the dsRNA-dependent protein kinase (PKR) and other host-cell defences as part of an evasion strategy. It is therefore unclear whether IFN treatment responses can be modelled realistically in this artificial, *in vitro* system.

In the future, the replicon model will be of great value in

the development and assessment of antiviral activity of newly developed protease and RNA polymerase inhibitors for HCV therapy (reviewed by De Francesco *et al.*, 2003) and for investigating the development of antiviral resistance (Lu *et al.*, 2004). The model is, at present, again limited by the lack of availability of replicons from other genotypes, particularly as there are concerns that antiviral agents modelled specifically on the active sites of genotype 1b protease or RNA polymerase may not be as active against corresponding sites of other subtypes or genotypes (Holland-Staley *et al.*, 2002). Very recently, it was indeed found that non-genotype 1-infected individuals were non-responsive or only weakly responsive to short-term treatment with the BILN 2061 protease inhibitor (Reiser, 2004), in contrast to its efficacy in genotype 1-infected individuals (Lamarre *et al.*, 2003). This is consistent with biochemical evidence for a nearly 100-fold reduction in binding affinity of BILN 2061 to genotype 2 and 3 proteases (Thibeault *et al.*, 2004). Genotype-specific differences in response to the new generation of antiviral agents will be a major research priority in the future.

Natural history. In contrast to the clear-cut differences between genotypes in their response to antiviral therapy, it has been much more difficult to obtain data on the differences in natural history and pathogenicity between HCV genotypes. Part of the problem with these investigations lies in the nature of the patient cohorts that have been studied to date and the exceptionally long time over which complications of HCV infection present clinically. With a few exceptions, severity of disease has typically been assessed in cross-sectional cohorts recruited from patients who were referred because of overt liver disease (such as portal hypertension, cirrhosis or abnormalities in liver-function tests, e.g. elevated alanine aminotransferase levels). This biased recruitment towards the minority with clinically apparent disease creates study cohorts that lack the community denominator and information on durations of infection with which to estimate the time-course of disease development. More importantly for this discussion, cross-sectional recruitment of 'hepatitis' patients cannot resolve whether some genotypes are more likely to present clinically than others.

Longitudinal studies, where the course of HCV disease over time in individuals with known times of infection is monitored prospectively, are few in number and frequently limited to patients who are infected with a single genotype. For example, natural history studies of the Irish and East German anti-D cohorts considered individuals who were infected only with genotype 1b (Power *et al.*, 1994; Takaki *et al.*, 2000). Similarly, a prospective study in the USA of individuals who were exposed to HCV by blood transfusion in the 1970s was limited to predominantly genotype 1a or 1b infections (Seeff *et al.*, 2001). However, more genotype diversity is found in several European cohorts in which an early diagnosis of infection was possible through specific risk factors, such as haemophilia, or in community-based

case-control studies. In these studies, genotype 1 appeared invariably to be more likely to establish persistence and, in carriers, to be associated with more severe liver disease, compared with genotypes 2 and 3 (Yee *et al.*, 2000; Franchini *et al.*, 2001; Mazzeo *et al.*, 2003; Resti *et al.*, 2003).

Surprisingly, and in contrast to the probable greater long-term pathogenicity of genotype 1, infections with genotype 3 are associated with a higher incidence of steatosis (Rubbia-Brandt *et al.*, 2000; Adinolfi *et al.*, 2001), which is thought to result from direct cytopathic damage to hepatocytes from a block in lipoprotein secretion (Serfaty *et al.*, 2001). As with the many other manifestations of biological differences between genotypes (including the vexed question of whether genotype 1 is more likely to cause hepatocellular carcinoma; Di Bisceglie, 1997), the availability of an *in vitro* system for investigating differences in the replication of different genotypes would be of considerable value in dissecting out the differences in virus-host cell interactions that underlie these clinical observations.

Interaction with the host cell. HCV is unusual for an RNA virus in being able to establish persistent infections in humans. This has attracted considerable efforts to understand the immunological basis for this phenomenon. As described in the section entitled 'Sequence variability within genotypes', there is substantial evidence that HCV is subjected to a variety of immune-mediated pressures that shape its evolution, driving, for example, sequence variability in the envelope region to evade antibody-mediated neutralization and, potentially, of T-cell epitopes to evade cytotoxic T-cell responses. However, it is now increasingly recognized that the most significant and evolutionarily ancient component of host defences against virus infections lies within the cell. Intracellular defences are generally triggered through the recognition of dsRNA replication intermediates through interaction with members of a large family of structurally related dsRNA-binding proteins (DRBPs; Girardin *et al.*, 2002; Saunders & Barber, 2003). DRBPs are coupled to a range of antiviral effector pathways that, in vertebrates, include PKR-mediated induction of apoptosis and modulation of the IFN-response pathways (Clemens & Elia, 1997) and activation of oligoadenylate synthetase, resulting in RNase L production and consequent cytoplasmic RNA cleavage (Player & Torrence, 1998). Mammalian cells may additionally control RNA virus infections through the actions of the mammalian homologues of Dicer and the siRNA-armed RNA-induced silencing complex (Waterhouse *et al.*, 2001; Gitlin & Andino, 2003).

In common with other RNA and DNA viruses, HCV has developed a range of cell-defence evasion mechanisms that are centred around the activities of NS5A (Tan & Katze, 2001; Reyes, 2002; Macdonald & Harris, 2004). Whilst NS5A is a necessary part of the virus replication complex, it shows additional activities in binding to and inactivating

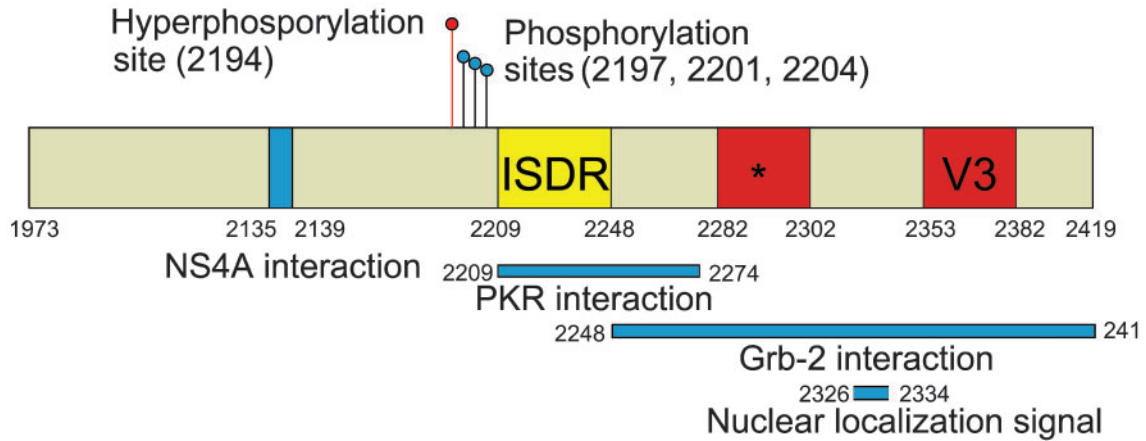


Fig. 5. Diagram of the NS5A gene, showing the ISDR and regions of the protein that are known to interact with PKR and other cellular pathways associated with innate cell defences against viruses. Sequence changes in the ISDR correlate with sensitivity of the virus to IFN therapy, potentially representing a trade-off between functional activity and immune selection. Sequences of other regions under apparent selection during therapy are shown in red.

PKR (Gale *et al.*, 1997), blocking apoptotic pathways through sequestration of p53, modulation of intracellular calcium levels and binding to growth factor receptor-bound protein 2 (Tan *et al.*, 1999; Gong *et al.*, 2001; Majumder *et al.*, 2001) and induction of anti-inflammatory interleukin 8 secretion (Polyak *et al.*, 2001; Fig. 5). It has also recently been shown that the HCV NS3/4A protease blocks the phosphorylation and signalling function of the antiviral IFN regulatory factor 3 (Foy *et al.*, 2003). The E2 protein, when expressed as a non-glycosylated, cytosolic protein (Pavio *et al.*, 2002), also appears to bind to and inhibit PKR as a result of sequence similarity to the (auto)phosphorylation domains of PKR and to e1F2 α (Taylor *et al.*, 1999). Interestingly, the degree of similarity to this ‘homology’ domain was greatest for genotype 1 variants and it was proposed that this contributed to the greater resistance of this genotype to IFN therapy. Finally, the association of GORS with virus persistence (Simmonds *et al.*, 2004) suggests that the formation of extensive RNA secondary structure in the genomic strand of HCV plays a role in the evasion of cell defences, potentially by facilitating escape from innate responses that are induced by certain structured RNAs. Each of these complex cell interactions potentially contributes to host persistence and to the inhibition of secondary T-cell responses to the virus during chronic infection.

One possible explanation for the differences in the outcome of infection between variants and genotypes of HCV is that they interact differently with host cells and achieve varying degrees of effectiveness in counteracting cell defences. Most obviously, the greater similarity of the E2 protein of genotype 1 to the phosphorylation domains of PKR and e1F2 α has been suggested to explain its greater clinical resistance to treatment (see above). However, further studies have generally not confirmed this hypothesis,

with little correlation between the E2 sequence and response between genotypes or subtypes 1b, 2a, 2b, 2c, 3a and 4c/d (4a) (Saito *et al.*, 2003; Watanabe *et al.*, 2003b; Quer *et al.*, 2004).

More promising evidence for a relationship between virus sequence and persistence/treatment resistance was demonstrated in the region of NS5A that interacts with PKR. Long before its function was known, it was observed that there was a clustering of amino acid changes in NS5A during IFN treatment. An association was also found between treatment response and possession of the so-called ‘prototype’ NS5A sequence in the region where mutations occurred (Enomoto *et al.*, 1995). Prototype ‘IFN-sensitivity determining region’ (ISDR) sequences were also associated with higher circulating virus loads in untreated patients (Watanabe *et al.*, 2003a). As the ISDR colocalizes with the part of NS5A that interacts with PKR (Fig. 5), it was suggested that PKR evasion was a key determinant in the persistence of HCV and, potentially, other aspects of virus–host interaction.

Since the original study, several groups have sought to reproduce the findings of a dependence on ISDR sequence of treatment response in other patient cohorts. Despite highly variable results between studies, a recent meta-analysis of all the available data has demonstrated a clear correlation between the prototype ISDR sequence and treatment resistance and, as a corollary, a large number of diverse amino acid changes in non-responders (Witherell & Beineke, 2001). It has also been shown that the same differential response exists in HCV genotype 2a and 2b infections (Murakami *et al.*, 1999). In trying to unravel the mechanism of this interaction, it remains curious that whilst the ‘prototype’ ISDR sequence of NS5A is found specifically in individuals who resist IFN therapy, there is

no evident selection for this sequence in viruses with non-‘prototype’ sequences that are treatment-sensitive.

One theory is that the sequence in NS5A is under strong immune selection and shows varying degrees of freedom to mutate towards the most biologically active (‘prototype’) sequence for each genotype. NS5A is indeed known to contain a high concentration of T- and B-cell epitopes (Zhang *et al.*, 1994; Rodríguez-López *et al.*, 1999; Lee *et al.*, 2000; Dou *et al.*, 2002) and it is possible that immune selection in many individuals drives the ISDR or neighbouring sequence away from the prototype in individuals with certain HLA types that target epitopes in this region. A poorly functioning NS5A protein may make the infecting virus more sensitive to intracellular antiviral responses and, thus, to a greater likelihood of spontaneous viral clearance, as well as increased susceptibility to IFN therapy in those who remain viraemic. Similar immune-mediated selection may underlie the observation of treatment-induced amino acid changes in other functional regions of NS5A, such as V3 and a second region at positions 2282–2302 (marked with an asterisk in Fig. 5) (Nousbaum *et al.*, 2000; Sarrazin *et al.*, 2002).

The balance in this ‘trade-off’ between NS5A function and immunological recognition may differ between genotypes of HCV. For example, the reason that infections by HCV genotypes 2 and 3 are generally much more responsive to IFN treatment may be because a far greater proportion of individuals recognize the prototype NS5A protein immunologically. Subsequent evolution of the infecting virus with a functionally impaired NS5A protein makes it less able to resist the further assault of exogenously administered IFN used for therapy. Human population-specific differences in the frequencies of HLA types in different study groups may also go some way to explaining why the association of ‘prototype’ ISDR (and potentially sequences in other NS5A regions) with treatment resistance varies so much between studies in Japan and Europe (Witherell & Beineke, 2001).

Summary

In summary, the evolution of HCV is a highly dynamic process. It occurs both through multiple processes of adaptive selection that drive sequence change (such as those resulting from the host immune response and potentially from antiviral treatment) and through drift, in which phenotypically neutral sequence changes accumulate over time without altering the phenotype or behaviour of the virus. However, despite its potential to change rapidly, the longer-term evolution of HCV appears to be remarkably conservative. Whilst the differences in treatment response between genotypes are important clinically, there has been little fundamental change in the relationship between HCV genotypes and their human hosts (such as their ability to persist and transmit) over the extremely long periods over which they have probably evolved. HCV thus appears to have successfully filled a very specific ecological niche in

human populations. Knowing more about the intimate host–parasite relationship that balances innate and acquired immune-defence mechanisms in the host with the development of complex evasion mechanisms in the virus is the key to understanding its pathogenesis and for developing future treatment intervention strategies.

References

- Adinolfi, L. E., Gambardella, M., Andreana, A., Tripodi, M.-F., Utili, R. & Ruggiero, G. (2001). Steatosis accelerates the progression of liver damage of chronic hepatitis C patients and correlates with specific HCV genotype and visceral obesity. *Hepatology* **33**, 1358–1364.
- Bartenschlager, R., Kaul, A. & Sparacio, S. (2003). Replication of the hepatitis C virus in cell culture. *Antiviral Res* **60**, 91–102.
- Blight, K. J., Kolykhalov, A. A. & Rice, C. M. (2000). Efficient initiation of HCV RNA replication in cell culture. *Science* **290**, 1972–1974.
- Blight, K. J., McKeating, J. A., Marcotrigiano, J. & Rice, C. M. (2003). Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture. *J Virol* **77**, 3181–3190.
- Booth, J. C., Kumar, U., Webster, D., Monjardino, J. & Thomas, H. C. (1998). Comparison of the rate of sequence variation in the hypervariable region of E2/NS1 region of hepatitis C virus in normal and hypogammaglobulinemic patients. *Hepatology* **27**, 223–227.
- Bukh, J., Purcell, R. H. & Miller, R. H. (1993). At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. *Proc Natl Acad Sci U S A* **90**, 8234–8238.
- Bukh, J., Pietschmann, T., Lohmann, V. & 7 other authors (2002). Mutations that permit efficient replication of hepatitis C virus RNA in Huh-7 cells prevent productive replication in chimpanzees. *Proc Natl Acad Sci U S A* **99**, 14416–14421.
- Candotti, D., Temple, J., Sarkodie, F. & Allain, J.-P. (2003). Frequent recovery and broad genotype 2 diversity characterize hepatitis C virus infection in Ghana, west Africa. *J Virol* **77**, 7914–7923.
- Cantaloube, J.-F., Biagini, P., Attoui, H., Gallian, P., de Micco, P. & De Lamballerie, X. (2003). Evolution of hepatitis C virus in blood donors and their respective recipients. *J Gen Virol* **84**, 441–446.
- Chang, K.-M., Rehmann, B., McHutchison, J. G., Pasquinelli, C., Southwood, S., Sette, A. & Chisari, F. V. (1997). Immunological significance of cytotoxic T lymphocyte epitope variants in patients chronically infected by the hepatitis C virus. *J Clin Invest* **100**, 2376–2385.
- Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). Isolation of a cDNA derived from a blood-borne non-A, non-B hepatitis genome. *Science* **244**, 359–362.
- Clemens, M. J. & Elia, A. (1997). The double-stranded RNA-dependent protein kinase PKR: structure and function. *J Interferon Cytokine Res* **17**, 503–524.
- Cochrane, A., Searle, B., Hardie, A. & 7 other authors (2002). A genetic analysis of hepatitis C virus transmission between injection drug users. *J Infect Dis* **186**, 1212–1221.
- Colina, R., Casane, D., Vasquez, S., García-Aguirre, L., Chunga, A., Romero, H., Khan, B. & Cristina, J. (2004). Evidence of intratypic recombination in natural populations of hepatitis C virus. *J Gen Virol* **85**, 31–37.
- Cooper, S., Erickson, A. L., Adams, E. J., Kansopon, J., Weiner, A. J., Chien, D. Y., Houghton, M., Parham, P. & Walker, C. M. (1999).

- Analysis of a successful immune response against hepatitis C virus. *Immunity* **10**, 439–449.
- De Francesco, R., Tomei, L., Altamura, S., Summa, V. & Migliaccio, G. (2003).** Approaching a new era for hepatitis C virus therapy: inhibitors of the NS3-4A serine protease and the NS5B RNA-dependent RNA polymerase. *Antiviral Res* **58**, 1–16.
- de Lamballerie, X., Charrel, R. N., Attoui, H. & De Micco, P. (1997).** Classification of hepatitis C virus variants in six major types based on analysis of the envelope 1 and nonstructural 5B genome regions and complete polyprotein sequences. *J Gen Virol* **78**, 45–51.
- Di Bisceglie, A. M. (1997).** Hepatitis C and hepatocellular carcinoma. *Hepatology* **26**, 34S–38S.
- Domingo, E., Escarmis, C., Sevilla, N., Moya, A., Elena, S. F., Quer, J., Novella, I. S. & Holland, J. J. (1996).** Basic concepts in RNA virus evolution. *FASEB J* **10**, 859–864.
- Dou, X.-G., Talekar, G., Chang, J., Dai, X., Li, L., Bonafonte, M. T., Holloway, B., Fields, H. A. & Khudyakov, Y. E. (2002).** Antigenic heterogeneity of the hepatitis C virus NS5A protein. *J Clin Microbiol* **40**, 61–67.
- Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998).** Rates of spontaneous mutation. *Genetics* **148**, 1667–1686.
- 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. *J Clin Invest* **96**, 224–230.
- Erickson, A. L., Kimura, Y., Igarashi, S. & 7 other authors (2001).** The outcome of hepatitis C virus infection is predicted by escape mutations in epitopes targeted by cytotoxic T lymphocytes. *Immunity* **15**, 883–895.
- Farci, P., Alter, H. J., Govindarajan, S. & 8 other authors (1992).** Lack of protective immunity against reinfection with hepatitis C virus. *Science* **258**, 135–140.
- Farci, P., Shimoda, A., Coiana, A. & 9 other authors (2000).** The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344.
- Foy, E., Li, K., Wang, C., Sumpter, R., Jr, Ikeda, M., Lemon, S. M. & Gale, M., Jr (2003).** Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. *Science* **300**, 1145–1148.
- Franchini, M., Rossetti, G., Tagliaferri, A. & 7 other authors (2001).** The natural history of chronic hepatitis C in a cohort of HIV-negative Italian patients with hereditary bleeding disorders. *Blood* **98**, 1836–1841.
- Frank, C., Mohamed, M. K., Strickland, G. T. & 8 other authors (2000).** The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet* **355**, 887–891.
- Frese, M., Pietschmann, T., Moradpour, D., Haller, O. & Bartenschlager, R. (2001).** Interferon- α inhibits hepatitis C virus subgenomic RNA replication by an MxA-independent pathway. *J Gen Virol* **82**, 723–733.
- Fretz, C., Jeannel, D., Stuyver, L., Hervé, V., Lunel, F., Boudifa, A., Mathiot, C., de Thé, G. & Fournel, J. J. (1995).** HCV infection in a rural population of the Central African Republic (CAR): evidence for three additional subtypes of genotype 4. *J Med Virol* **47**, 435–437.
- Gale, M. J., Jr, 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.
- Gao, F., Yue, L., White, A. T. & 7 other authors (1992).** Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature* **358**, 495–499.
- Gao, F., Bailes, E., Robertson, D. L. & 9 other authors (1999).** Origins of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397**, 436–441.
- Gates, A. T., Sarisky, R. T. & Gu, B. (2004).** Sequence requirements for the development of a chimeric HCV replicon system. *Virus Res* **100**, 213–222.
- Girardin, S. E., Sansonetti, P. J. & Philpott, D. J. (2002).** Intracellular vs extracellular recognition of pathogens – common concepts in mammals and flies. *Trends Microbiol* **10**, 193–199.
- Gitlin, L. & Andino, R. (2003).** Nucleic acid-based immune system: the antiviral potential of mammalian RNA silencing. *J Virol* **77**, 7159–7165.
- Gong, G., Waris, G., Tanveer, R. & Siddiqui, A. (2001).** Human hepatitis C virus NS5A protein alters intracellular calcium levels, induces oxidative stress, and activates STAT-3 and NF- κ B. *Proc Natl Acad Sci U S A* **98**, 9599–9604.
- González-Pérez, M. A., Norder, H., Bergström, Å., Lopéz, E., Visoná, K. A. & Magnius, L. O. (1997).** High prevalence of GB virus C strains genetically related to strains with Asian origin in Nicaraguan hemophiliacs. *J Med Virol* **52**, 149–155.
- Harvala, H., Kalimo, H., Dahllund, L., Santti, J., Hughes, P., Hyytiä, T. & Stanway, G. (2002).** Mapping of tissue tropism determinants in coxsackievirus genomes. *J Gen Virol* **83**, 1697–1706.
- Heinz, F. X., Collett, M. S., Purcell, R. H., Gould, E. A., Howard, C. R., Houghton, M., Moorman, R. J. M., Rice, C. M. & Thiel, H.-J. (2000).** *Flaviviridae*. In *Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 859–878. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. San Diego, CA: Academic Press.
- Holland-Staley, C. A., Kovari, L. C., Golenberg, E. M., Pobursky, K. J. & Mayers, D. L. (2002).** Genetic diversity and response to IFN of the NS3 protease gene from clinical strains of the hepatitis C virus. *Arch Virol* **147**, 1385–1406.
- Hoofnagle, J. H. (2002).** Course and outcome of hepatitis C. *Hepatology* **36**, S21–S29.
- Ikeda, M., Yi, M., Li, K. & Lemon, S. M. (2002).** Selectable subgenomic and genome-length dicistronic RNAs derived from an infectious molecular clone of the HCV-N strain of hepatitis C virus replicate efficiently in cultured Huh7 cells. *J Virol* **76**, 2997–3006.
- Jarvis, L. M., Watson, H. G., McOmish, F., Peutherer, J. F., Ludlam, C. A. & Simmonds, P. (1994).** Frequent reinfection and reactivation of hepatitis C virus genotypes in multitransfused hemophiliacs. *J Infect Dis* **170**, 1018–1022.
- Jeannel, D., Fretz, C., Traore, Y. & 9 other authors (1998).** Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. *J Med Virol* **55**, 92–97.
- Kalinina, O., Norder, H., Mukomolov, S. & Magnius, L. O. (2002).** A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. *J Virol* **76**, 4034–4043.
- Kalinina, O., Norder, H. & Magnius, L. O. (2004).** Full-length open reading frame of a recombinant hepatitis C virus strain from St Petersburg: proposed mechanism for its formation. *J Gen Virol* **85**, 1853–1857.
- Kantzanou, M., Lucas, M., Barnes, E., Komatsu, H., Dusheiko, G., Ward, S., Harcourt, G. & Klennerman, P. (2003).** Viral escape and T cell exhaustion in hepatitis C virus infection analysed using class I peptide tetramers. *Immunol Lett* **85**, 165–171.
- Kao, J. H., Chen, P. J., Lai, M. Y. & Chen, D. S. (1993).** Superinfection of heterologous hepatitis C virus in a patient with chronic type C hepatitis. *Gastroenterology* **105**, 583–587.

- Kato, T., Date, T., Miyamoto, M., Furusaka, A., Tokushige, K., Mizokami, M. & Wakita, T. (2003). Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology* **125**, 1808–1817.
- Kelleher, A. D., Long, C., Holmes, E. C. & 15 other authors (2001). Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-restricted cytotoxic T lymphocyte responses. *J Exp Med* **193**, 375–386.
- Kimura, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet Res* **11**, 247–269.
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- King, J. L. & Jukes, T. H. (1969). Non-Darwinian evolution. *Science* **164**, 788–798.
- Kumar, U., Brown, J., Monjardino, J. & Thomas, H. C. (1993). Sequence variation in the large envelope glycoprotein (E2/NS1) of hepatitis C virus during chronic infection. *J Infect Dis* **167**, 726–730.
- Kuo, G., Choo, Q. L., Alter, H. J. & 17 other authors (1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* **244**, 362–364.
- Lai, M. E., Mazzoleni, A. P., Balestriesi, A., Argioli, F., De Virgili, S., Cao, A., Purcell, R. H. & Farci, P. (1994). Hepatitis C virus in multiple episodes of acute hepatitis in polytransfused thalassaemic children. *Lancet* **343**, 388–390.
- Lamarre, D., Anderson, P. C., Bailey, M. & 25 other authors (2003). An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**, 186–189.
- Lanford, R. E., Guerra, B., Lee, H., Averett, D. R., Pfeiffer, B., Chavez, D., Notvall, L. & Bigger, C. (2003). Antiviral effect and virus-host interactions in response to alpha interferon, gamma interferon, poly(I)-poly(C), tumor necrosis factor alpha, and ribavirin in hepatitis C virus subgenomic replicons. *J Virol* **77**, 1092–1104.
- Lechner, F., Wong, D. K. H., Dunbar, P. R. & 7 other authors (2000). Analysis of successful immune responses in persons infected with hepatitis C virus. *J Exp Med* **191**, 1499–1512.
- Lee, A. Y., Polakos, N. K., Otten, G. R., Ulmer, J. B., Houghton, M. & Paliard, X. (2000). Quantification of the number of cytotoxic T cells specific for an immunodominant HCV-specific CTL epitope primed by DNA immunization. *Vaccine* **18**, 1962–1968.
- Leslie, A. J., Pfafferoth, K. J., Chetty, P. & 26 other authors (2004). HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med* **10**, 282–289.
- Lohmann, V., Körner, F., Koch, J.-O., Herian, U., Theilmann, L. & Bartenschlager, R. (1999). Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285**, 110–113.
- Lu, L., Pilot-Matias, T. J., Stewart, K. D. & 7 other authors (2004). Mutations conferring resistance to a potent hepatitis C virus serine protease inhibitor in vitro. *Antimicrob Agents Chemother* **48**, 2260–2266.
- Macdonald, A. & Harris, M. (2004). Hepatitis C virus NS5A: tales of a promiscuous protein. *J Gen Virol* **85**, 2485–2502.
- Majumder, M., Ghosh, A. K., Steele, R., Ray, R. & Ray, R. B. (2001). Hepatitis C virus NS5A physically associates with p53 and regulates p21/waf1 gene expression in a p53-dependent manner. *J Virol* **75**, 1401–1407.
- Makuwa, M., Souquière, S., Telfer, P. & 7 other authors (2003). Occurrence of hepatitis viruses in wild-born non-human primates: a 3 year (1998–2001) epidemiological survey in Gabon. *J Med Primatol* **32**, 307–314.
- Mazzeo, C., Azzaroli, F., Giovanelli, S. & 9 other authors (2003). Ten year incidence of HCV infection in northern Italy and frequency of spontaneous viral clearance. *Gut* **52**, 1030–1034.
- Mellor, J., Holmes, E. C., Jarvis, L. M., Yap, P. L., Simmonds, P. & The International HCV Collaborative Study Group (1995). Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. *J Gen Virol* **76**, 2493–2507.
- Menéndez, C., Sánchez-Tapias, J.-M., Alonso, P.-L. & 8 other authors (1999). Molecular evidence of mother-to-infant transmission of hepatitis G virus among women without known risk factors for parenteral infections. *J Clin Microbiol* **37**, 2333–2336.
- Mizokami, M., Gojobori, T., Ohba, K., Ikeo, K., Ge, X.-M., Ohno, T., Orito, E. & Lau, J. Y. N. (1996). Hepatitis C virus types 7, 8 and 9 should be classified as type 6 subtypes. *J Hepatol* **24**, 622–624.
- Mortimer, P. P. (1995). Arsphenamine jaundice and the recognition of instrument-borne virus infection. *Genitourin Med* **71**, 109–119.
- Murakami, T., Enomoto, N., Kurosaki, M., Izumi, N., Marumo, F. & Sato, C. (1999). Mutations in nonstructural protein 5A gene and response to interferon in hepatitis C virus genotype 2 infection. *Hepatology* **30**, 1045–1053.
- Ndjomou, J., Pybus, O. G. & Matz, B. (2003). Phylogenetic analysis of hepatitis C virus isolates indicates a unique pattern of endemic infection in Cameroon. *J Gen Virol* **84**, 2333–2341.
- Neumann, A. U., Lam, N. P., Dahari, H., Gretch, D. R., Wiley, T. E., Layden, T. J. & Perelson, A. S. (1998). Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon- α therapy. *Science* **282**, 103–107.
- Nousbaum, J.-B., Polyak, S. J., Ray, S. C., Sullivan, D. G., Larson, A. M., Carithers, R. L., Jr & Gretch, D. R. (2000). Prospective characterization of full-length hepatitis C virus NS5A quasispecies during induction and combination antiviral therapy. *J Virol* **74**, 9028–9038.
- Okamoto, H., Kojima, M., Okada, S. & 7 other authors (1992). Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability. *Virology* **190**, 894–899.
- Pause, A., Kukolj, G., Bailey, M. & 11 other authors (2003). An NS3 serine protease inhibitor abrogates replication of subgenomic hepatitis C virus RNA. *J Biol Chem* **278**, 20374–20380.
- Pavesi, A. (2001). Origin and evolution of GBV-C/hepatitis G virus and relationships with ancient human migrations. *J Mol Evol* **53**, 104–113.
- Pavio, N., Taylor, D. R. & Lai, M. M. C. (2002). Detection of a novel unglycosylated form of hepatitis C virus E2 envelope protein that is located in the cytosol and interacts with PKR. *J Virol* **76**, 1265–1272.
- Pawlotsky, J.-M. (2003a). Mechanisms of antiviral treatment efficacy and failure in chronic hepatitis C. *Antiviral Res* **59**, 1–11.
- Pawlotsky, J.-M. (2003b). The nature of interferon- α resistance in hepatitis C virus infection. *Curr Opin Infect Dis* **16**, 587–592.
- Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. T. (1998). A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev* **12**, 67–83.
- Pietschmann, T., Lohmann, V., Kaul, A., Krieger, N., Rinck, G., Rutter, G., Strand, D. & Bartenschlager, R. (2002). Persistent and transient replication of full-length hepatitis C virus genomes in cell culture. *J Virol* **76**, 4008–4021.
- Player, M. R. & Torrence, P. F. (1998). The 2–5A system: modulation of viral and cellular processes through acceleration of RNA degradation. *Pharmacol Ther* **78**, 55–113.

- Polyak, S. J., Khabar, K. S. A., Paschal, D. M., Ezelle, H. J., Duverlie, G., Barber, G. N., Levy, D. E., Mukaida, N. & Gretsch, D. R. (2001). Hepatitis C virus nonstructural 5A protein induces interleukin-8, leading to partial inhibition of the interferon-induced antiviral response. *J Virol* **75**, 6095–6106.
- Power, J. P., Lawlor, E., Davidson, F., Yap, P. L., Kenny-Walsh, E., Whelton, M. J. & Walsh, T. J. (1994). Hepatitis C viraemia in recipients of Irish intravenous anti-D immunoglobulin. *Lancet* **344**, 1166–1167.
- Pradat, P. & Trepo, C. (2000). HCV: epidemiology, modes of transmission and prevention of spread. *Baillière's Best Pract Res Clin Gastroenterol* **14**, 201–210.
- Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C. & Harvey, P. H. (2001). The epidemic behavior of the hepatitis C virus. *Science* **292**, 2323–2325.
- Pybus, O. G., Drummond, A. J., Nakano, T., Robertson, B. H. & Rambaut, A. (2003). The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol Biol Evol* **20**, 381–387.
- Quer, J., Murillo, P., Martell, M., Gómez, J., Esteban, J. I., Esteban, R. & Guardia, J. (2004). Subtype mutations in the envelope 2 region including phosphorylation homology domain of hepatitis C virus do not predict effectiveness of antiviral therapy. *J Viral Hepat* **11**, 45–54.
- Ray, S. C., Arthur, R. R., Carella, A., Bukh, J. & Thomas, D. L. (2000). Genetic epidemiology of hepatitis C virus throughout Egypt. *J Infect Dis* **182**, 698–707.
- Reiser, M. (2004). Antiviral effect of BILN 2061, a novel HCV serine protease inhibitor, after oral treatment over 2 days in patients with chronic hepatitis C, non-genotype 1. In *Program and Abstracts of the 54th Annual Meeting of the American Association for the Study of Liver Diseases*, 24–28 October 2003, Boston, MA, USA.
- Resti, M., Jara, P., Hierro, L., Azzari, C., Giacchino, R., Zuin, G., Zancan, L., Pedditzi, S. & Bortolotti, F. (2003). Clinical features and progression of perinatally acquired hepatitis C virus infection. *J Med Virol* **70**, 373–377.
- Reyes, G. R. (2002). The nonstructural NS5A protein of hepatitis C virus: an expanding, multifunctional role in enhancing hepatitis C virus pathogenesis. *J Biomed Sci* **9**, 187–197.
- Rodríguez-López, M., Riezu-Boj, J. I., Ruiz, M., Berasain, C., Civeira, M. P., Prieto, J. & Borrás-Cuesta, F. (1999). Immunogenicity of variable regions of hepatitis C virus proteins: selection and modification of peptide epitopes to assess hepatitis C virus genotypes by ELISA. *J Gen Virol* **80**, 727–738.
- Rubbia-Brandt, L., Quadri, R., Abid, K. & 8 other authors (2000). Hepatocyte steatosis is a cytopathic effect of hepatitis C virus genotype 3. *J Hepatol* **33**, 106–115.
- Ruggieri, A., Argentini, C., Kouruma, F., Chionne, P., D'Ugo, E., Spada, E., Dettori, S., Sabbatani, S. & Rapicetta, M. (1996). Heterogeneity of hepatitis C virus genotype 2 variants in West Central Africa (Guinea Conakry). *J Gen Virol* **77**, 2073–2076.
- Saito, T., Ito, T., Ishiko, H., Yonaha, M., Morikawa, K., Miyokawa, A. & Mitamura, K. (2003). Sequence analysis of PePHD within HCV E2 region and correlation with resistance of interferon therapy in Japanese patients infected with HCV genotypes 2a and 2b. *Am J Gastroenterol* **98**, 1377–1383.
- Santti, J., Hyypiä, T., Kinnunen, L. & Salminen, M. (1999). Evidence of recombination among enteroviruses. *J Virol* **73**, 8741–8749.
- Sarisky, R. T. (2004). Non-nucleoside inhibitors of the HCV polymerase. *J Antimicrob Chemother* **54**, 14–16.
- Sarrazin, C., Herrmann, E., Bruch, K. & Zeuzem, S. (2002). Hepatitis C virus nonstructural 5A protein and interferon resistance: a new model for testing the reliability of mutational analyses. *J Virol* **76**, 11079–11090.
- Saunders, L. R. & Barber, G. N. (2003). The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB J* **17**, 961–983.
- Seeff, L. B. (2002). Natural history of chronic hepatitis C. *Hepatology* **36**, S35–S46.
- Seeff, L. B., Hollinger, F. B., Alter, H. J. & 21 other authors (2001). Long-term mortality and morbidity of transfusion-associated non-A, non-B, and type C hepatitis: a National Heart, Lung, and Blood Institute collaborative study. *Hepatology* **33**, 455–463.
- Serfaty, L., Andreani, T., Giral, P., Carbonell, N., Chazouillères, O. & Poupon, R. (2001). Hepatitis C virus induced hypobetalipoproteinemia: a possible mechanism for steatosis in chronic hepatitis C. *J Hepatol* **34**, 428–434.
- Sheridan, I., Pybus, O. G., Holmes, E. C. & Klenerman, P. (2004). High-resolution phylogenetic analysis of hepatitis C virus adaptation and its relationship to disease progression. *J Virol* **78**, 3447–3454.
- Simmonds, P. (2001). The origin and evolution of hepatitis viruses in humans. *J Gen Virol* **82**, 693–712.
- Simmonds, P. & Smith, D. B. (1999). Structural constraints on RNA virus evolution. *J Virol* **73**, 5787–5794.
- Simmonds, P., Holmes, E. C., Cha, T.-A. & 7 other authors (1993). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *J Gen Virol* **74**, 2391–2399.
- Simmonds, P., Tuplin, A. & Evans, D. J. (2004). Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: implications for virus evolution and host persistence. *RNA* **10**, 1337–1351.
- Simons, J. N., Pilot-Matias, T. J., Leary, T. P. & 9 other authors (1995). Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci U S A* **92**, 3401–3405.
- Smith, D. B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P. L. & Simmonds, P. (1997). The origin of hepatitis C virus genotypes. *J Gen Virol* **78**, 321–328.
- Stuyver, L., Rossau, R., Wyseur, A., Duhamel, M., Vanderborght, B., Van Heuverswyn, H. & Maertens, G. (1993). Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *J Gen Virol* **74**, 1093–1102.
- Takaki, A., Wiese, M., Maertens, G., Depla, E., Seifert, U., Liebetrau, A., Miller, J. L., Manns, M. P. & Rehermann, B. (2000). Cellular immune responses persist and humoral responses decrease two decades after recovery from a single-source outbreak of hepatitis C. *Nat Med* **6**, 578–582.
- Tan, S.-L. & Katze, M. G. (2001). How hepatitis C virus counteracts the interferon response: the jury is still out on NS5A. *Virology* **284**, 1–12.
- Tan, S.-L., Nakao, H., He, Y., Vijaysri, S., Neddermann, P., Jacobs, B. L., Mayer, B. J. & Katze, M. G. (1999). NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. *Proc Natl Acad Sci U S A* **96**, 5533–5538.
- Taniguchi, S., Okamoto, H., Sakamoto, M. & 7 other authors (1993). A structurally flexible and antigenically variable N-terminal domain of the hepatitis C virus E2/NS1 protein: implication for an escape from antibody. *Virology* **195**, 297–301.
- Taylor, D. R., Shi, S. T., Romano, P. R., Barber, G. N. & Lai, M. M. C. (1999). Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science* **285**, 107–110.

- Thibeault, D., Bousquet, C., Gingras, R., Lagacé, L., Maurice, R., White, P. W. & Lamarre, D. (2004). Sensitivity of NS3 serine proteases from hepatitis C virus genotypes 2 and 3 to the inhibitor BILN 2061. *J Virol* **78**, 7352–7359.
- Thimme, R., Oldach, D., Chang, K. M., Steiger, C., Ray, S. C. & Chisari, F. V. (2001). Determinants of viral clearance and persistence during acute hepatitis C virus infection. *J Exp Med* **194**, 1395–1406.
- Thomas, D. L. (2000). Hepatitis C epidemiology. *Curr Top Microbiol Immunol* **242**, 25–41.
- Tokita, H., Okamoto, H., Tsuda, F. & 7 other authors (1994a). Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. *Proc Natl Acad Sci U S A* **91**, 11022–11026.
- Tokita, H., Shrestha, S. M., Okamoto, H., Sakamoto, M., Hirokita, M., Iizuka, H., Shrestha, S., Miyakawa, Y. & Mayumi, M. (1994b). Hepatitis C virus variants from Nepal with novel genotypes and their classification into the third major group. *J Gen Virol* **75**, 931–936.
- Tokita, H., Okamoto, H., Luengrojanakul, P., Vareesangthip, K., Chainuvati, T., Iizuka, H., Tsuda, F., Miyakawa, Y. & Mayumi, M. (1995). Hepatitis C virus variants from Thailand classifiable into five novel genotypes in the sixth (6b), seventh (7c, 7d) and ninth (9b, 9c) major genetic groups. *J Gen Virol* **76**, 2329–2335.
- Trozzi, C., Bartholomew, L., Ceccacci, A. & 10 other authors (2003). In vitro selection and characterization of hepatitis C virus serine protease variants resistant to an active-site peptide inhibitor. *J Virol* **77**, 3669–3679.
- Tuplin, A., Wood, J., Evans, D. J., Patel, A. H. & Simmonds, P. (2002). Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* **8**, 824–841.
- Tuplin, A., Evans, D. J. & Simmonds, P. (2004). Detailed mapping of RNA secondary structures in core and NS5B-encoding region sequences of hepatitis C virus by RNase cleavage and novel bioinformatic prediction methods. *J Gen Virol* **85**, 3037–3047.
- Varaklioti, A., Vassilaki, N., Georgopoulou, U. & Mavromara, P. (2002). Alternate translation occurs within the core coding region of the hepatitis C viral genome. *J Biol Chem* **277**, 17713–17721.
- Walewski, J. L., Keller, T. R., Stump, D. D. & Branch, A. D. (2001). Evidence for a new hepatitis C virus antigen encoded in an overlapping reading frame. *RNA* **7**, 710–721.
- Wang, C., Pflugheber, J., Sumpter, R., Jr, Sodora, D. L., Hui, D., Sen, G. C. & Gale, M., Jr (2003). Alpha interferon induces distinct translational control programs to suppress hepatitis C virus RNA replication. *J Virol* **77**, 3898–3912.
- Wansbrough-Jones, M. H., Frimpong, E., Cant, B., Harris, K., Evans, M. R. W. & Teo, C. G. (1998). Prevalence and genotype of hepatitis C virus infection in pregnant women and blood donors in Ghana. *Trans R Soc Trop Med Hyg* **92**, 496–499.
- Wasley, A. & Alter, M. J. (2000). Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin Liver Dis* **20**, 1–16.
- Watanabe, H., Nagayama, K., Enomoto, N., Itakura, J., Tanabe, Y., Hamano, K., Izumi, N., Sato, C. & Watanabe, M. (2003a). Sequence elements correlating with circulating viral load in genotype 1b hepatitis C virus infection. *Virology* **311**, 376–383.
- Watanabe, H., Nagayama, K., Enomoto, N., Itakura, J., Tanabe, Y., Sato, C., Izumi, N. & Watanabe, M. (2003b). Amino acid substitutions in PKR-eIF2 phosphorylation homology domain (PePHD) of hepatitis C virus E2 protein in genotype 2a/2b and 1b in Japan and interferon efficacy. *Hepatol Res* **26**, 268–274.
- Waterhouse, P. M., Wang, M.-B. & Lough, T. (2001). Gene silencing as an adaptive defence against viruses. *Nature* **411**, 834–842.
- Weiner, A. J., Geysen, H. M., Christopherson, C. & 12 other authors (1992). Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc Natl Acad Sci U S A* **89**, 3468–3472.
- Witherell, G. W. & Beineke, P. (2001). Statistical analysis of combined substitutions in nonstructural 5A region of hepatitis C virus and interferon response. *J Med Virol* **63**, 8–16.
- Xu, L.-Z., Larzul, D., Delaporte, E., Bréchet, C. & Kremsdorf, D. (1994). Hepatitis C virus genotype 4 is highly prevalent in central Africa (Gabon). *J Gen Virol* **75**, 2393–2398.
- Xu, Z., Choi, J., Yen, T. S. B., Lu, W., Strohecker, A., Govindarajan, S., Chien, D., Selby, M. J. & Ou, J. (2001). Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. *EMBO J* **20**, 3840–3848.
- Yee, T. T., Griffioen, A., Sabin, C. A., Dusheiko, G. & Lee, C. A. (2000). The natural history of HCV in a cohort of haemophilic patients infected between 1961 and 1985. *Gut* **47**, 845–851.
- You, S., Stump, D. D., Branch, A. D. & Rice, C. M. (2004). A *cis*-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. *J Virol* **78**, 1352–1366.
- Zeuzem, S. (2004). Heterogeneous virologic response rates to interferon-based therapy in patients with chronic hepatitis C: who responds less well? *Ann Intern Med* **140**, 370–381.
- Zhang, Z.-X., Chen, M., Sönnernborg, A. & Sällberg, M. (1994). Antigenic structure of the complete nonstructural (NS) 2 and 5 proteins of hepatitis C virus (HCV): anti-HCV NS2 and NS5 antibody reactivities in relation to HCV serotype, presence of HCV RNA, and acute HCV infection. *Clin Diagn Lab Immunol* **1**, 290–294.