Abstract
The hepatitis C virus (HCV) is the main causative agent of non-A, non-B hepatitis in humans and a major cause of mortality and morbidity in the world. Currently there is no effective treatment available for the infection caused by this virus, whose replication depends on an unusual translation-initiation mechanism. The viral RNA contains an internal ribosome-entry site (IRES) that is recognized specifically by the small ribosomal subunit and by eukaryotic initiation factor 3, and these interactions allow cap (7-methylguanine nucleotide)-independent initiation of viral protein synthesis. In this article, we review the structure and mechanism of translation initiation of the HCV IRES, and its potential as a target for novel antivirals.

Introduction
The hepatitis C virus (HCV) is the main causative agent of post-transfusion hepatitis. This human pathogen is a positive (messenger) sense single-stranded RNA virus (Figure 1) that relies on an
The viral genome contains a large single open reading frame encoding a polyprotein precursor of approx. 3000 amino acids, which is processed by cellular and virally encoded proteases into functional proteins. Despite extensive variation among different viral isolates across most of the HCV genome, largely due to the infidelity of the viral RNA polymerase [12], the IRES is highly conserved, indicating that this structure is essential for the virus. ss, single-stranded. Modified from [30] with permission. © (1999) Thieme Medical Publishers.

Figure 1
HCV life cycle

The viral genome contains a large single open reading frame encoding a polyprotein precursor of approx. 3000 amino acids, which is processed by cellular and virally encoded proteases into functional proteins. Despite extensive variation among different viral isolates across most of the HCV genome, largely due to the infidelity of the viral RNA polymerase [12], the IRES is highly conserved, indicating that this structure is essential for the virus. ss, single-stranded. Modified from [30] with permission. © (1999) Thieme Medical Publishers.

Figure 1
HCV life cycle

The viral genome contains a large single open reading frame encoding a polyprotein precursor of approx. 3000 amino acids, which is processed by cellular and virally encoded proteases into functional proteins. Despite extensive variation among different viral isolates across most of the HCV genome, largely due to the infidelity of the viral RNA polymerase [12], the IRES is highly conserved, indicating that this structure is essential for the virus. ss, single-stranded. Modified from [30] with permission. © (1999) Thieme Medical Publishers.

Figure 1
HCV life cycle

The viral genome contains a large single open reading frame encoding a polyprotein precursor of approx. 3000 amino acids, which is processed by cellular and virally encoded proteases into functional proteins. Despite extensive variation among different viral isolates across most of the HCV genome, largely due to the infidelity of the viral RNA polymerase [12], the IRES is highly conserved, indicating that this structure is essential for the virus. ss, single-stranded. Modified from [30] with permission. © (1999) Thieme Medical Publishers.

The viral genome contains a large single open reading frame encoding a polyprotein precursor of approx. 3000 amino acids, which is processed by cellular and virally encoded proteases into functional proteins. Despite extensive variation among different viral isolates across most of the HCV genome, largely due to the infidelity of the viral RNA polymerase [12], the IRES is highly conserved, indicating that this structure is essential for the virus. ss, single-stranded. Modified from [30] with permission. © (1999) Thieme Medical Publishers.

alternative mechanism for initiating the translation of its polyprotein. Initiation of HCV protein synthesis does not depend on recognition of a 7-methyl-guanine nucleotide (cap) present at the 5' terminus of most human mRNAs. Instead, the HCV RNA contains a well-defined structure of about 340 nucleotides in its 5'-untranslated region (UTR), called the internal ribosome-entry site (IRES; Figure 2a) [1]. The IRES of HCV is bound specifically by the host-cell small (40 S) ribosomal subunit and by eukaryotic initiation factor (eIF) 3. These interactions allow the viral start codon to be recognized directly in a cap- and poly(A)-independent way, and protein synthesis to be initiated by a mechanism which is fundamentally distinct from most host mRNAs (Figure 2b).

At this time, there is no effective vaccination or cure for hepatitis C, an infection affecting at least 170 million people worldwide. This slow-progressing disease is transmitted through contaminated blood transfusions and needle sharing, and frequently leads to liver cirrhosis and cancer [2,3]. The IRES-dependent mechanism for translation initiation is shared by other viruses and by some eukaryotic RNAs. However, the structure of the HCV IRES is likely to be different from the IRES of human mRNAs, making this RNA motif and its complexes with 40 S and eIF3 attractive targets for the discovery of new antiviral agents because of the potential for selectivity. As a consequence, both the structure and the mechanism of translation initiation by the IRES of HCV have been the subject of intense research in recent years [4–6]. In this article, we will provide a brief overview on the structure and function of the HCV IRES.
Figure 2
(a) Structure and (b) mechanism of translation initiation of HCV IRES

(a) Sequence and secondary structure of HCV IRES (nucleotides 1–383 of genotype 1a). The binding sites of eIF3 and the 40 S ribosomal subunit are indicated with boxes, the location of the start codon is marked with an arrow, and the subdomains whose structure has been determined by NMR spectroscopy are marked with asterisks. (b) In cap-mediated initiation of translation (top), the 43 S particle (composed of the 40 S subunit, eIF3, eIF2, and initiator tRNA) binds to the eIF4F-cap complex. The 5'-UTR (shown as an undulating line) is then scanned until the start codon (AUG) is encountered and the ribosome is assembled. In HCV IRES-mediated initiation of translation (bottom), the 43 S particle recognizes the AUG codon directly through interaction with the IRES. The eIFs are identified by their numbers. (a) is modified from [15] with permission. © (1999) American Society for Microbiology.
Mechanism of translation initiation by the HCV IRES

Cap-dependent translation involves the recognition of a methylated guanine nucleotide (cap) at the 5' terminus of the mRNA by translation-initiation factor eIF4F, a protein complex that includes eIF4A, eIF4G and the cap-binding protein eIF4E (Figure 2b). The 43 S particle (composed of the 40 S subunit, eIF3 and a ternary Met-tRNA-eIF2-GTP complex) then binds to the mRNA-eIF4F-eIF4B complex, probably through interactions between eIF3 and eIF4G, and the 5'-UTR region is subsequently scanned until the first AUG initiation codon is encountered. Protein synthesis is then initiated after release of the initiation factors and binding of the 60 S large ribosomal subunit (for a review, see [7]). In contrast, HCV IRES-mediated translation initiation only requires interaction between the IRES and two components of the 43 S particle, the 40 S subunit and eIF3 [8,9]. This interaction results in the direct recognition of the viral start codon and the initiation of protein synthesis (Figure 2b). How this occurs mechanistically is not known.

IRES elements were first described in picornaviruses, later in HCV, and more recently in a growing number of eukaryotic mRNAs [1,10]. IRESs of different viral families or cellular RNAs present no obvious similarities in secondary structure, and do not necessarily share the same mechanism for translation initiation. The IRES of picornaviruses, for example, require several of the canonical factors necessary for cap-dependent initiation of translation [11]. However, all IRESs allow cap-independent initiation of protein synthesis, providing the virus with a selective advantage for the translation of its own proteins by by-passing key regulatory steps involving the eIF4F complex [7] (Figure 2).

The structure of the HCV IRES

The IRES encompasses most of the 5'-UTR of the HCV RNA and is highly conserved compared with the rest of the viral genome. This conservation indicates that it plays an essential role in the viral life cycle (Figure 1), and allows use of the viral 5'-UTR for diagnostics and genotype classification [3,12]. The HCV IRES contains four conserved secondary structure domains, which have been identified on the basis of thermodynamics, phylogeny, mutational analyses and ribonuclease mapping [13–15] (Figure 2a). It was recently demonstrated that these domains form a unique but not highly compact tertiary structure in the presence of magnesium ions [16].

Efforts to elucidate the complete atomic structure of the HCV IRES are currently under way in a number of laboratories. Crystallization of the entire IRES is proving difficult, but recently published electron microscopy (EM) studies have advanced our understanding of the overall structural organization of the HCV IRES [17,18]. In addition, NMR studies from several groups have provided high-resolution structures of three isolated subdomains, comprising approx. 20% of the IRES structure (Figure 2a) [6,19,20].

The 20 Å resolution map of the HCV IRES bound to the 40 S subunit by EM reveals a single elongated shape. Domains IIId–IIIe and II (Figure 2a) extend in opposite directions from a small central domain that includes stem loops and junctions IIle–IIIf [17]. The same structural organization was observed in isolated IRES particles by other EM and small-angle X-ray scattering studies [16,18]. The three regions of the HCV IRES that have been characterized structurally by NMR spectroscopy are subdomains IIId, IIIe [19,20] and more recently an internal loop within subdomain IIIb [6] (Figures 2a and 3). The hairpin loop of the small IIIe subdomain

Secondary structure of the HCV IRES subdomains IIIb internal loop recognized by eIF3

The two mismatched cytosine bases and the base pairs separating them from the internal loop are conserved in all HCV isolates, but a significant degree of variability is observed for the nucleotides forming the internal loop. This figure shows the most common sequence (found in HCV genotype 1, which is predominant worldwide), and the sequences of genotypes 2 (top right) and 3 (bottom right). Nucleotide variations are shown in italics.

3' 5'
| G : C | G : C |
| G : C | C : G |
| A : U | C : 211 |
| 186 | 210 |

3' 5'
| G : C |
| A |
| 182 |

3' 5'
| G : C |
| A |
| 182 |

3' 5'
| G : C |
| A |
| 182 |

3' 5'
| G : C |
| A |
| 182 |

3' 5'
| G : C |
| A |
| 182 |
forms a novel tetraloop fold with three exposed Watson–Crick faces that may be involved in 40 S binding [20]. The stem of subdomain IIId forms a loop E motif similar to those observed in prokaryotic and eukaryotic ribosomal RNA, and a six-nucleotide hairpin loop containing an S-turn motif [19,20]. The sequences of the hairpin loops of subdomains IIIe and IIId are conserved among all HCV isolates and play an important role in translation initiation [19,20]. In contrast, the internal loop of subdomain IIIb is less conserved (Figure 3), but it is also essential for translation activity [6].

Interaction of the HCV IRES with the 40 S subunit and eIF3

Enzymic and chemical footprinting and domain-deletion experiments have identified the ribosome- and eIF3-binding sites within the HCV IRES. The ribosome 40 S subunit binds to subdomains IIb, IIIa, IIIc, IIId, IIIe, IIIf and IV and junctions IIIabc and IIIef [8,20–22], while eIF3 binding is localized to subdomains IIIa and IIIb and junction IIIabc [8,9,22–24] (see Figure 2a). Thus there is an extensive association of the IRES with the 40 S subunit that encompasses nearly all of the IRES, while the interaction with eIF3 is much less extensive. Indeed, formation of the IRES–40 S–eIF3 ternary complex is driven by high-affinity interactions between 40 S and IRES and between 40 S and eIF3 [22]. Thus the IRES probably recognizes pre-assembled 40 S–eIF3 complexes in vivo, and the binding of these two translation components to specific sites on the HCV IRES probably enhances the direct placement of the initiation codon in the ribosomal decoding site [9].

So far, there is no atomic-level information concerning how the IRES–40 S–eIF3 complex associates. In a cryo-EM reconstruction of the IRES–40 S complex [17], the IRES binds the solvent side of the 40 S subunit in a position that is consistent with the proposed path of the mRNA through the subunit, and the observed intermolecular contacts are in agreement with the protection sites described above. Specifically, the apical half of domain II (stem-loop IIb, Figure 2a) contacts the 40 S subunit near the E site (where deacylated tRNA binds before it exits the ribosome). The binding of subdomain IIb induces a conformational change in the 40 S subunit that may help the 3’ end of the coding RNA to thread into the mRNA entry channel [17]. Thus, different HCV IRES domains may be playing the role of canonical initiation factors until the ribosome is assembled and elongation begins.

The eIF3-binding site
eIF3 is the largest of the eIFs. It is a multimeric complex of approx. 650 kDa comprising at least 11 protein subunits and shaped like a flat triangular prism [25,26]. While the X-ray structure of both prokaryotic ribosome subunits is now available [27,28], the atomic structure of eIF3 is unknown. Six of its subunits (p170, p116, p110, p66, p47 and p44) have been proposed to have RNA-binding activity, and four (p170, p116, p66 and p47) have been implicated in HCV IRES binding by UV cross-linking [8,23]. During cap-dependent initiation, eIF3 binds to the small ribosomal subunit and eIF-4G and stabilizes the binding of the Met-tRNA-eIF2–GTP ternary complex to the 40 S subunit [26]. Although eIF3 is not required for the assembly of HCV IRES and 40 S subunits, this factor is absolutely necessary for HCV translation initiation: the IRES interacts with a 40 S–eIF3 complex in vivo, and eIF3 is probably required for the assembly of active 80 S ribosomal complexes [8].

As stated above, it is well established that eIF3 binds to subdomains IIIb and IIIa and junction IIIabc of the HCV IRES [8,9,22–24] (Figure 2a), an area much smaller than that bound by the 40 S subunit. In the EM representation of the IRES–40 S complex, the extended stem-loop IIIb is not involved in 40 S binding and protrudes away from the ribosome subunit [17]. Chemical and enzymic footprinting experiments show particularly strong footprints at and near an internal loop below the apical hairpin loop of subdomain IIIb [8,22] (Figures 2a and 3). Subdomain IIIb is less well conserved than other domains within the HCV IRES. In fact, the sequence of the IIIb internal loop footprinted by eIF3 is different in the three most common HCV genotype variants (Figure 3; see e.g. [29]). This observation led to erroneous suggestions that this element would not be important for IRES function. However, based on a combination of domain deletion, mutagenesis and NMR spectroscopy, we have shown that the IIIb internal loop, together with an adjacent C:C mismatch (Figures 2a and 3), are essential for IRES-mediated initiation of translation [6]. In spite of its sequence variability, the IIIb internal loop folds into a three-dimensional structure that is
conserved among three major HCV genotypes, possibly among all HCV isolates. This structural element and the adjacent mismatched helix are required for HCV IRES-mediated protein synthesis [6].

The potential of HCV IRES as a drug target

The sequence and secondary structure of the HCV IRES are highly conserved compared with the rest of the genome and the proteins that it encodes. Furthermore, the mechanism by which the IRES mediates initiation of protein synthesis is distinct from the common eukaryotic pathway. Both arguments indicate that the HCV IRES and its complexes are attractive drug targets. Specific sites for interfering with IRES function include subdomain IIb, which on the basis of EM and mutational analyses interacts with the 40 S ribosomal subunit and may be required for correct placement of the coding RNA strand. Another potential target is the interaction between conserved subdomains IIIe or IIId and the 40 S subunit, which is required for IRES-mediated translational activity and is very likely to be recognized by eIF3. Since this factor is essential for the formation of active ribosomal complexes, the conserved IIIb internal loop structure also provides the opportunity to apply structure-based methods for the discovery of new inhibitors of HCV protein synthesis and replication. These efforts are underway in several biotechnology and pharmaceutical companies and will hopefully result in new lead compounds for the development of effective anti-HCV drugs.

This work has been supported by an EU Marie Curie Fellowship to J.G. (QLK2-CT-1999-51436), by an EU FP5 grant (QLK2-2000-01535) and by the Medical Research Council.

References


Received 20 November 2001