SCIENTIFIC REPORTS

OPEN

Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging

H. Stewart[1], R.J. Bingham[2], S. J. White[3], E. C. Dykeman[2], C. Zothner[1], A. K. Tuplin[1,3], P. G. Stockley[3], R. Twarock[2] & M. Harris[1,3]

The specific packaging of the hepatitis C virus (HCV) genome is hypothesised to be driven by Core-RNA interactions. To identify the regions of the viral genome involved in this process, we used SELEX (systematic evolution of ligands by exponential enrichment) to identify RNA aptamers which bind specifically to Core *in vitro*. Comparison of these aptamers to multiple HCV genomes revealed the presence of a conserved terminal loop motif within short RNA stem-loop structures. We postulated that interactions of these motifs, as well as sub-motifs which were present in HCV genomes at statistically significant levels, with the Core protein may drive virion assembly. We mutated 8 of these predicted motifs within the HCV infectious molecular clone JFH-1, thereby producing a range of mutant viruses predicted to possess altered RNA secondary structures. RNA replication and viral titre were unaltered in viruses possessing only one mutated structure. However, infectivity titres were decreased in viruses possessing a higher number of mutated regions. This work thus identified multiple novel RNA motifs which appear to contribute to genome packaging. We suggest that these structures act as cooperative packaging signals to drive specific RNA encapsidation during HCV assembly.

Hepatitis C virus (HCV) is the leading cause of chronic liver disease, hepatocellular carcinoma and liver transplants in the developed world. It is estimated that 170 million people are chronically infected worldwide, approximately 25% of whom will eventually present with liver cirrhosis. HCV therefore represents a significant health and economic burden. Direct-acting antivirals (DAAs) targeting 3 viral non-structural proteins are now available for treatment of HCV infection. The rates of viral clearance, however, differ significantly between genotypes and a low genetic barrier to resistance for certain DAAs means that viral escape is highly likely. Additionally, the pricing of the most recently available DAA (Sofosbuvir; a polymerase inhibitor) makes it unfeasible to treat the majority of patients in many countries. It is therefore apparent that further research into the basic virology of HCV is warranted.

HCV is of the family *Flaviviridae*, genus *Hepaciviridae* and possesses a ~9.6 kb single-stranded positive-sense RNA genome, encompassing highly structured untranslated flanking regions (UTRs). The 5′ UTR includes an internal ribosome entry site (IRES), which directs cap-independent translation of a single polyprotein that is subsequently cleaved into 10 constituent proteins: Core, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B. Although the non-structural (NS) proteins contribute to some degree to the assembly and release of the virion, Core, E1 and E2 play mostly structural roles[1]. The five C terminal NS proteins (NS3, NS4A, NS4B, NS5A and NS5B) were shown in 1999 to represent the minimal essential requirement for RNA replication of a subgenomic replicon[2]. These proteins and the ensuing RNA replication process have been available for targeted drug design

[1]School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, United Kingdom. [2]York Centre for Complex Systems Analysis, Departments of Mathematics and Biology, University of York, York, YO10 5DD, United Kingdom. [3]Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, United Kingdom. Correspondence and requests for materials should be addressed to M.H. (email: M.Harris@leeds.ac.uk)

much longer than the processes of virion assembly, RNA packaging and viral egress. The availability of a fully infectious viral molecular clone (JFH-1) means that these later stages can also be investigated in detail[3]. The role(s) of genomic RNA structures during genome encapsidation has not, however, been the subject of extensive research.

Genomes are packaged during formation of the nucleocapsid by the Core protein. Core is encoded by the 5′ region of the ORF and is released from the initially-translated polyprotein by host signal peptidase cleavage at the C terminus. Core is then further processed by signal peptide peptidase, resulting in a mature 21 kDa protein which forms homodimers and localises to cytosolic lipid droplets (for review see[1]). The NS proteins forming the RNA replication complex remain membrane-bound within an ER-derived membranous web where RNA replication occurs. There is accumulating evidence that the delivery of a nascent RNA by its cognate replication complex to the site of virion assembly is required for assembly initiation (for review see 4,5). However whether nucleocapsid formation and genome packaging occur at the lipid droplet, or whether this organelle merely acts as a storage site for Core prior to its relocalisation back to the cytoplasmic face of the ER (where assembly may occur), is a controversial topic[6]. Following assembly the nascent nucleocapsid buds into the ER, acquiring the lipid envelope and the associated E1-E2 viral glycoproteins prior to cellular egress (for review see 7).

It is apparent that the RNA replication and virion assembly processes are spatially distinct within the cellular microenvironment and interactions between the replication complex and the Core protein are required for efficient packaging[8]. Consequently, the packaging of replication-defective genomes (which are not presented by a replication complex) is notoriously inefficient in HCV. It is this inability to study assembly without concurrent replication which has hindered the identification of RNA structures contributing solely to assembly. The requirement for RNA presentation by the replication complex may prevent encapsidation of defective and/or partial HCV genome fragments by Core. The fact that cellular RNAs are excluded from this process indicates additional selective factors must contribute. For many positive-sense RNA viruses, the nucleocapsid assembly and genome packaging events occur simultaneously through recognition of a large stable RNA structure (the packaging signal, ψ) by their capsid protein. The essential role of these structural motifs within the viral RNA in directing genome encapsidation has long been recognised[9]. The fact that pre-formed "empty" HCV capsid-like particles have not been identified suggests that HCV utilises a similar mechanism of RNA recognition to initiate assembly[1], although RNA-free capsid-like structures have been isolated from *in vitro* translation systems[10]. However, unlike prototypic packaging signals such as those found in the retroviruses, alphaviruses and coronaviruses, a single RNA structure which is both essential and sufficient to target non-viral RNAs to a nascent HCV nucleocapsid particle has not been identified.

It is possible that HCV utilises a recently discovered novel mechanism of genome encapsidation during nucleocapsid assembly, utilising multiple, relatively low-affinity structures termed Packaging Signals (PSs)[11,12]. We wished to investigate the possibility that similar specific secondary structures are present within HCV RNAs destined for packaging, and whether their interactions with Core cooperatively drive the RNA encapsidation and nucleocapsid assembly processes.

Results

Systematic evolution of ligands by exponential enrichment (SELEX) of RNA aptamers against Core.
Mature Core consists of two distinct domains: the N terminal domain 1 (D1, 124 residues), which is highly basic and binds RNA, and the hydrophobic C terminal domain 2 (D2, 50 residues) which possesses a membrane binding region. D2 stabilises Core on the surface of lipid droplets and exposes the hydrophilic D1 to the viral RNA, therefore it is essential for nucleocapsid formation in mammalian cells[13]. However, *in vitro* D1 is sufficient to form stable proteinase-resistant nucleocapsids when stabilised by interactions with nucleic acids[14–16]. We therefore prepared D1 as a SELEX target by recombinant protein expression. D1 HCV Core (JFH-1) was expressed in *E. coli* and purified by nickel affinity chromatography against an N-terminal hexahistidine tag. The identity of the his-tagged protein was confirmed by LC-MS/MS of Coomassie stained bands excised from an SDS-PAGE gel. The antigenicity of the final purified protein was also confirmed by immunoblotting for Core D1 and the hexahistidine tag (Fig. 1).

This protein was used for SELEX with a 2′OH RNA aptamer library encompassing a N35 random region. SELEX is an established process for the *in vitro* isolation of high-affinity aptamers: *in vitro* selected nucleic acid molecules which bind with high affinity and specificity to their target[17,18]. Each aptamer within a randomised library possesses a unique tertiary structure, depending on the series of stems, pseudoknots, kinks and/or bulges which are present in its most stable conformation. Each aptamer will bind with varying affinity to the ligand of interest. SELEX is therefore likely to enrich for aptamers with conserved conformation rather than a unique primary sequence.

Locations and consensus motif of putative packaging signals (pPSs).
The sequences present in the final round of SELEX were determined by Next Generation Sequencing (NGS) of the pool and the individual aptamer sequences were aligned to the JFH-1 genome (GenBank ID AB047639.1). It should be noted that known PSs in other ssRNA viruses are sequence/structure degenerate and their capsid/coat protein recognition sequences are discontinuous and minimal[12,19,20]. A consequence is that the pool of binders from SELEX should include a majority of oligonucleotides that, although matching essential features of HCV PSs, are unlikely to match extended regions of the genome. Despite this constraint, there are aptamers that match multiple sites within the genome and occur with statistical significance (a Bernoulli score of 12 or more; red peaks, Fig. 2) compared to the unselected naïve library (grey peaks, Fig. 2). This outcome suggests that the HCV genome does contain multiple PS sites. In order to test this hypothesis, the same procedure was applied to an additional 14 genotype 2 HCV genomes, i.e. they were compared to the aptamers selected against JFH-1, and a similar picture emerged. This suggests that these PS sites are evolutionarily conserved. Such peaks, within 10 nucleotides of the
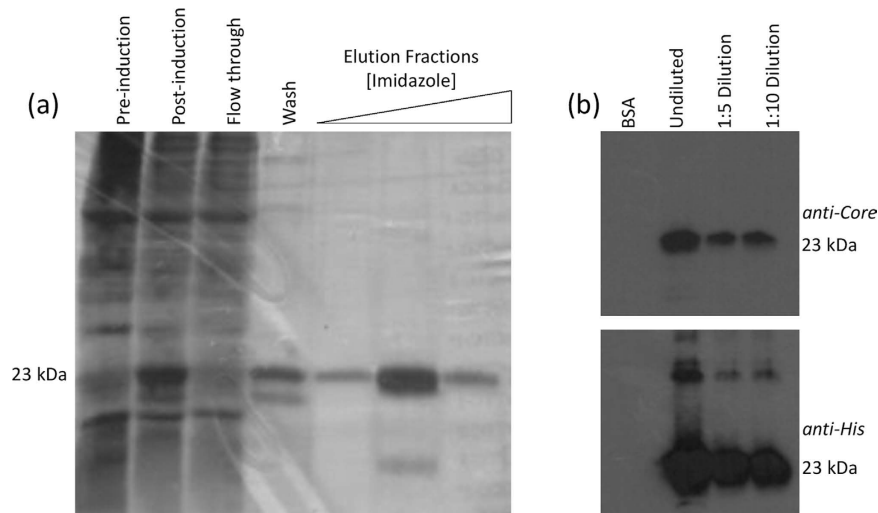
Figure 1. Expression and purification of Core D1. Core D1 was expressed in *E. coli* and purified via nickel affinity chromatography. The purity of the protein was assessed by silver staining SDS-PAGE of cell lysates and imidazole elution fractions (**a**) Immunoblots against the purified protein, compared to bovine serum albumin as a loading control, were performed using either an in-house polyclonal rabbit anti-Core serum (upper right panel) or a commercial anti-His murine monoclonal antibody (Invitrogen) (**b**).
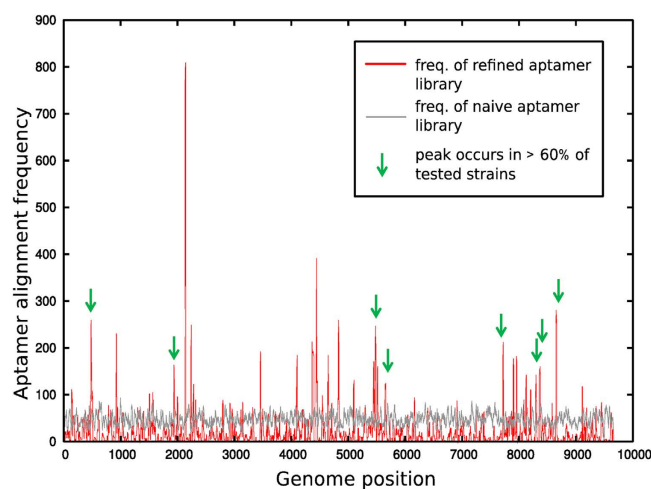

Figure 2. Packaging signals in HCV. Statistically significant alignments of the aptamer pool with the JFH-1 genome correspond to peaks in the red curve that are above the corresponding alignment of the naïve library (grey curve). Peaks occurring within 10 nucleotides of statistically significant peaks in at least 60% of the representative set of HCV genomes analysed here are marked by a green arrow.

peaks in Fig. 2, which were conserved for at least 60% of these additional strains, are highlighted by green arrows (Fig. 2).

For each of these conserved peaks the 30 nucleotides 5′ and 3′ to the peak nucleotide in the JFH-1 genome were extracted, and a range of possible secondary structure folds were considered via mFold[21]. A similarity analysis revealed the potential formation of stem-loops with similar loop motifs in each of these fragments. An alignment of the loop portions of these stem-loops is displayed in Fig. 3a. This analysis reveals a bias towards a GRRGR loop motif, R denoting a purine, suggesting that PSs should exhibit this motif or close variations thereof. In order to establish the statistical significance of this motif, we compared its occurrence in the loop portion of stem-loops in the wild-type genome with that in randomised versions of the genome, both for JFH-1 and for the other strain variants considered here. This confirmed a significant statistical bias in the wild-type genomes for this motif. We then interrogated sub-motifs of GRRGR in order to identify those with an even higher statistical significance, which revealed a GGRGG motif. The positions of stem-loops with this motif across the JFH-1 genome are displayed in Fig. 3b.

Modeling indicates that an ensemble of packaging signals with different degrees of affinity for capsid protein is required to promote efficient assembly[11,22]. Lower-affinity motifs may not necessarily be easily identifiable through a SELEX screen, yet their presence at statistically significant numbers within the wild-type genome
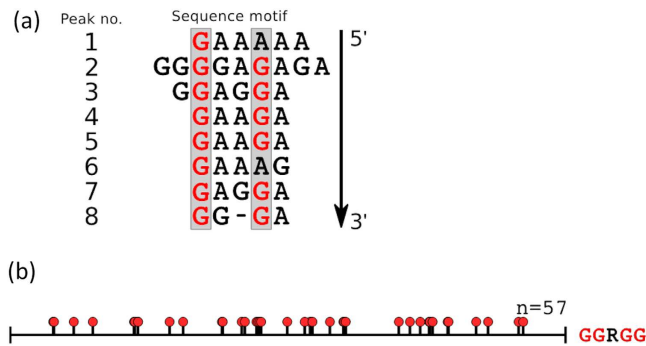
Figure 3. Identification and location of pPS recognition motifs. (**a**) Alignment of the loop portions of the stem-loops identified as a representative for each of the 8 conserved peaks in Fig. 2 (marked by a green arrow), revealing a dominant GRRGR motif, R denoting a purine. (**b**) Positions of stem-loops containing the GGRGG motif within the JFH-1 genome are shown as a suite of red lollipops.

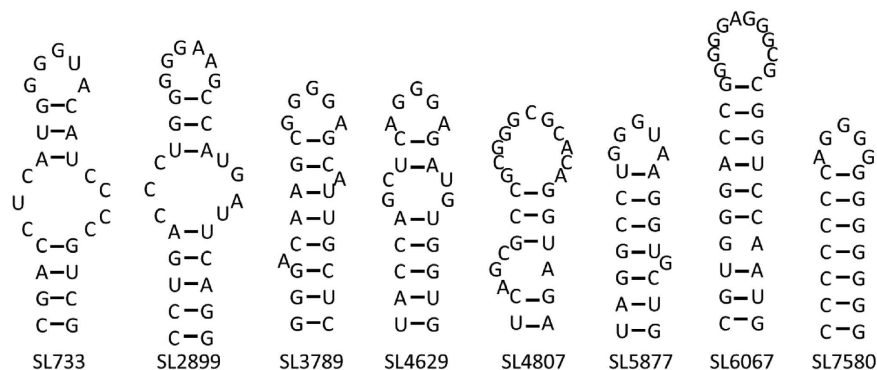**Figure 4. Predicted putative PS structures within the HCV genome.** The local RNA structure of each of the regions of interest was modelled by mFold[21]. Motifs are named according to the first paired nucleotide within the JFH-1 genome, according to convention.

would be indicative of a potential role in virus assembly. Based on previously-published models, the statistical bias observed for the GGRGG motif would be predicted to encompass at most 25% of the putative packaging signals (pPSs)[19]. Variations of the motif with statistical bias are the hallmark for PSs with intermediate or weaker affinity for the capsid protein. The same statistical analysis was therefore used to interrogate other variations of the GGRGR motif, in an effort to identify sub-motifs that were more frequent in the wild-type genome than in the randomised version. The motif GXRXR fulfilled these criteria (see Methods for details).

Mutagenesis to disrupt pPS.   Since modeling suggests that pPSs act collectively to promote virus assembly, an ensemble of 8 pPSs was selected for mutagenesis. The region encoding NS5B was avoided as numerous studies have reported the existence of essential structures required for RNA replication; mutations in this region would render the virus replication-defective[23,24]. Modeling moreover implies that a combination of high and low affinity binders is important, i.e., a combination of pPSs following the GGRGG motif and those that exhibit other variations of the GRRGR motif: e.g. GXRXR[19]. Eight stem-loops were selected for mutagenesis, three of which contained the predicted highly conserved motif (GGRGG). The remaining five stem-loops contained the lower-affinity GXRXR motif identified above (Fig. 4). The stem-loops are numbered according to the first paired nucleotide within the context of the JFH-1 genome, according to standard nomenclature.

A range of silent mutations were introduced into the ORF of HCV, using the infectious molecular clone pJFH-1. These were designed to disrupt the base pairing within the stems of the pPSs (Fig. 4). It was not possible in the majority of cases to completely ablate favourable folding energies for the formation of stem-loops and even mutant sites with unfavourable folding free energy might still form in the presence of a protein ligand such as Core. Eight mutant PSs were created, each possessing between 5 and 11 silent single nucleotide polymorphisms, collectively targeting an individual pPS (Fig. 5). These are referred to as the "single" pPS mutants, ΔSL733, ΔSL2899, ΔSL3789, ΔSL4629, ΔSL4807, ΔSL5877, ΔSL6067 and ΔSL7580. All 57 silent mutations were also combined into a single viral genome, termed Δ8XSL, in order to examine co-operative effects between the altered pPSs.

Replication and translation are unaffected in pPS mutants.   RNAs encoding all 8 single pPS mutant viruses, the Δ8XSL mutant, JFH-1 (wild-type) and JFH-1-GND (a polymerase-deficient control that is unable to undergo viral genome replication) were electroporated into Huh-7 cells. The intracellular RNA was quantified by

LOCATION	ΔG	MOTIF	SEQUENCE
Core	-9.7	SL733	GC**C** GAC CTC ATG GGG TAC ATC CCC GTC **G**TA A D L M G Y I P V V
	7.3	ΔSL733	GC<u>G</u> GA<u>T</u> CTC ATG GG<u>T</u> TA<u>T</u> AT<u>A</u> CCC GT<u>A</u> GTA
NS2	-12.0	SL2899	CT**C** CTG ACC CTG GGG GAA GCC ATG ATT CAG **G**AG L L L L G E A M I Q E
	-3.0	ΔSL2899	CT<u>T</u> <u>T</u>T<u>A</u> ACC CT<u>C</u> GG<u>C</u> GA<u>G</u> GCC ATG ATT CAG GAG
NS3	-9.4	SL3789	CGC G**G**G GAC AAG CGG GGA GCA TTG CT**C** TCC R G D K R G A L L S
	0.5	ΔSL3789	<u>A</u>GG GG<u>A</u> GAC AA<u>A</u> CG<u>T</u> GGA GCA TT<u>A</u> CT<u>G</u> AG<u>C</u>
NS3	-4.3	SL4629	A**T**A CCA GCT CAG GGA GAT GTG GT**G** I P A Q G D V V
	-1.7	ΔSL4629	ATA CC<u>C</u> GCT CA<u>A</u> GGA GA<u>C</u> GT<u>T</u> GTG
NS3	-8.7	SL4807	AG**T** CAG CGC CGC GGG CGC ACA GGT AG**A** S Q R R G R T G R
	1.8	ΔSL4807	<u>TCA</u> CA<u>A</u> A<u>G</u>A CGC GG<u>T</u> CGC ACA GGT AGA
NS4B	-7.5	SL5877	AGC A**T**A GGC CTG GGT AAG GTG CT**G** GTG S I G L G K V L V
	-0.4	ΔSL5877	<u>TCC</u> AT<u>T</u> GG<u>A</u> <u>TTA</u> GGT AA<u>A</u> GT<u>A</u> <u>TTA</u> GT<u>T</u>
NS4B	-12.0	SL6067	CA**C** GTG GGA CCG GGG GAG GGC GCG GTC CAA T**G**G ATG AAC H V G P G E G A V Q W M N
	-0.5	ΔSL6067	CAC GT<u>C</u> GGA CC<u>A</u> GG<u>A</u> GA<u>A</u> GG<u>A</u> GC<u>A</u> GT<u>A</u> CAA TGG ATG AA<u>T</u>
NS5A	-16.0	SL7580	**C**CC CCC CAG GGG GGG GGG **G**TA P P Q G G G V
	-7.4	ΔSL7580	CC<u>T</u> CC<u>G</u> CA<u>A</u> GG<u>C</u> GG<u>T</u> GG<u>C</u> GTA

Figure 5. Silent mutations designed to disrupt the secondary structures of the pPSs within JFH-1. Multiple silent mutations (red, underlined) were incorporated into the pJFH-1 plasmid to disrupt the pPSs. The 5′ and 3′ nucleotides of the structures displayed from Fig. 4 are underlined in bold in the wildtype (upper) sequences; on occasion additional nucleotides immediately flanking these regions were mutated to maximise the disruption potential. Mutant genomes were named according to the disrupted stem-loop (ΔSL). The peptides encoded within each motif are described between the nucleotide sequences. ΔG refers to the free energy of the most stable predicted RNA secondary structure (kcal/mol) as calculated by mFold[21].

qRT-PCR, at both 4 and 48 hours post electroporation (h.p.e) (Fig. 6a). Impaired replication was not observed in mutant viruses containing impaired pPSs, although ΔSL4807 possessed a slightly decreased replication rate which was not statistically significant. This was further confirmed by detection of the viral NS5A protein compared to cellular GAPDH expression via Western blot (Fig. 6b).

Infectious viral titres are decreased in the Δ8XSL pPS mutant. Both intracellular and released virus was titrated upon naïve Huh-7 cells and the infectious titre calculated in Infectious Units per millilitre (IU/mL) as previously described[25]. This reflects the number of infectious RNA-containing virions which had formed and/or been released from within the electroporated cells. Non-RNA containing particles, or aberrantly-assembled defective particles, therefore do not contribute to the viral titre as they are non-transmissible. At 48 h.p.e., only the Δ8XSL mutant virus displayed significantly decreased titres compared to the JFH-1 WT control, although ΔSL5877 and ΔSL7580 also appeared non-significantly decreased (Fig. 7a). The reduction in Δ8XSL infectious titre was consistent across the entire 48 hours, as evidenced by time course assays during which virus was harvested every 6 hours post-electroporation to assess the rate of virion formation and release (Fig. 7b). This provides evidence that although mutating each pPS independently has no apparent effect, the combined disruption of all 8 pPSs affects the late stages of the viral life cycle (assembly and/or egress) whilst not affecting RNA replication.

RNA secondary structures are altered within the packaging signal mutants. To confirm that the introduced mutations caused structural alterations, *in vitro* transcripts from a section of the NS4B-coding region from JFH-1 (WT) and Δ8XSL mutant genomes were synthesised. RNA was folded and selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) mapping was performed as previously described[24]. Briefly, folded RNA was treated with a compound which reacts preferentially with single-stranded, flexible regions. This reaction forms an irreversible adduct which causes premature termination during the subsequent reverse transcription of the RNA[26]. Primer extensions using radiolabelled primers were conducted and the resulting fragments were resolved on a 7% polyacrylamide denaturing gel. The terminations, indicative of a flexible nucleotide, were
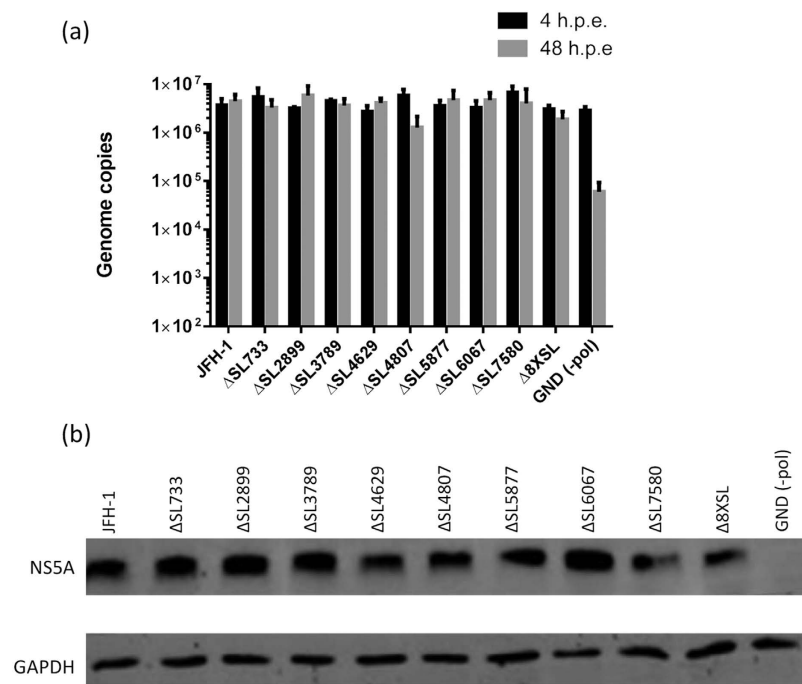
Figure 6. RNA replication and translation are unaffected in viral genomes containing disrupted pPSs. *In vitro* transcribed RNA was electroporated into naïve Huh-7 cells. At 4 and 48 h.p.e., RNA was extracted and quantified through qRT-PCR (**a**). Data represents the mean + SEM from three independent experiments. At 48 h.p.e., cells were lysed in Glasgow lysis buffer and 30 μg of each lysate was resolved by SDS-PAGE, followed by immunoblotting with NS5A-specific sheep polyclonal antiserum and a GAPDH commercial monoclonal antibody (**b**).
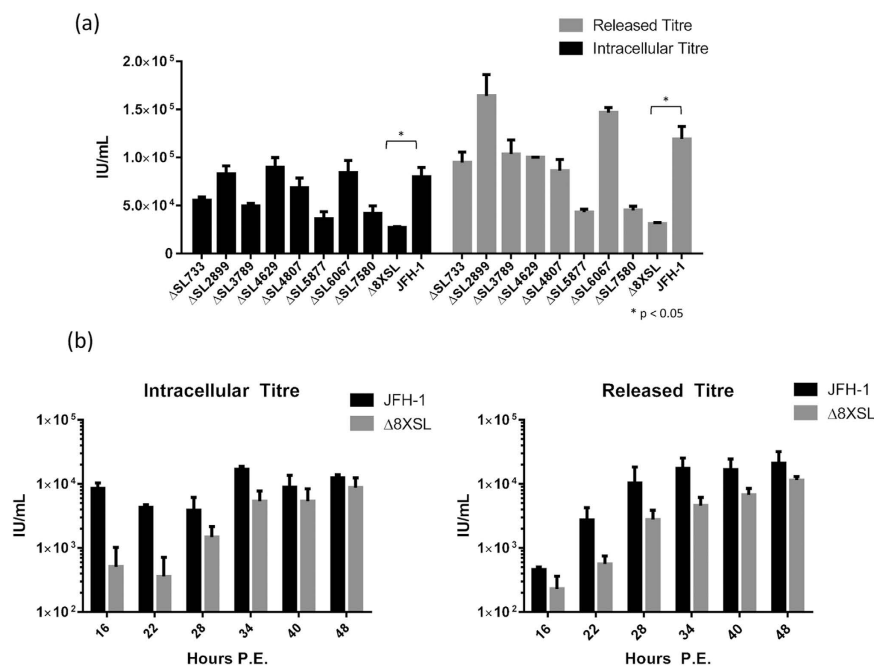

Figure 7. Viral titres are decreased in the multiple packaging signal mutant virus, Δ8XSL. 48 h.p.e., virus-containing supernatant and PBS lysates from electroporated cells were titrated upon naïve cells to determine the released and intracellular viral titres, respectively, of all single pPS mutant viruses, the multiple mutant virus Δ8XSL and wildtype JFH-1 (**a**). To assess the rate of virion formation and release, intracellular and released virus titres were also calculated for Δ8XSL and JFH-1 at multiple time points following electroporation (**b**). Data represents the mean + SEM from three independent experiments.
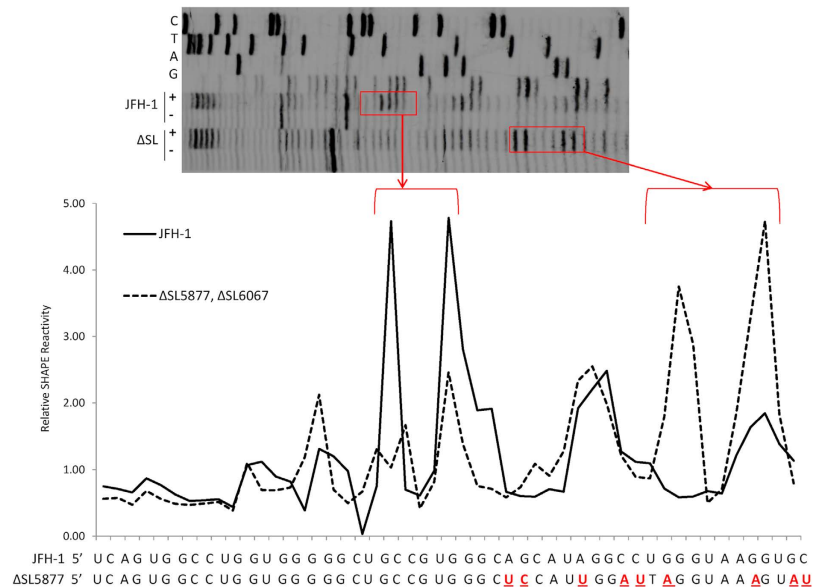
Figure 8. RNA secondary structure appears altered in subgenomic fragments containing disrupted pPSs. *In vitro* transcripts of the ~1.5 kb region encompassing SL5877 and SL6067 were synthesised from both JFH-1 and mutant templates. RNA was folded and subjected to structural analysis via SHAPE mapping. Following NMIA (+) or DMSO (−) treatment and reverse transcription, a primer designed to bind approximately 20 nucleotides downstream of SL5877 was used for primer extension. Products were resolved on a 7% denaturing polyacrylamide gel (upper panel) and the relative reactivity was quantified using SAFA[58]. Sequencing ladders were included to identify individual nucleotides, based upon the wildtype genome sequence.

visualised and their reactivity normalised to that of a DMSO-treated control RNA. The specific location of such terminations is determined by comparison to ddNTP sequencing ladders.

A high SHAPE reactivity is indicative of a flexible nucleotide, available for acylation by the compound; hence only the terminal loop of a prototype stem-loop structure would appear as "reactive" nucleotides. The HCV genome is dynamic and multiple RNA conformations are potentially able to form, including kissing-loops, pseudoknots, G-quadruplexes and other long-range interactions[24]. Additionally the active folds of pPSs may only form in the presence of an RNA chaperone such as Core or NS5A, which is absent during SHAPE mapping. The SHAPE data therefore indicates whether the engineered mutations altered the RNA flexibility, rather than confirming the mFold-predicted structure of a particular motif. The reactivity profile across the region encompassing SL5877 provided evidence that the silent mutations had altered the flexibility of the mutant RNA compared to wildtype (Fig. 8).

Together our data indicates the 8 putative packaging signals which we identified may form within the HCV genome, presumably during assembly. A mutant genome unable to form such structures displays impaired viral infectivity whilst RNA replication is unaffected, indicating these motifs may interact specifically with Core. The synergistic effect of these individual interactions may ensure encapsidation of a single viral genome occurs during virion assembly, thereby preventing non-specific packaging of cellular RNAs.

Discussion

The mechanism by which the HCV Core protein selectively binds and encapsidates viral genomic RNA during virion assembly is currently unknown. In contrast to other positive-strand RNA viruses[27,28] there is no evidence for a single *cis*-acting RNA structure capable of directing packaging of both viral and non-viral RNAs to a nascent capsid particle. Therefore it is plausible that multiple RNA structures contribute to RNA-Core binding and act cooperatively to drive encapsidation and assembly. The presence of multiple, weak-affinity packaging signals in other RNA virus genomes has recently been reported[19,29], which represents a novel mechanism for viral packaging in viruses which do not possess a readily identifiable prototypic packaging signal. Here, we provide evidence that the abrogation of short motifs (pPSs) located across the HCV genome has a significant effect on RNA encapsidation, manifested by a decrease in infectious titre. This is only apparent in the multiple pPS mutant (Δ8XSL), consistent with RNA packaging being a cooperative process in the *Hepaciviridae*. The fact that depletion of all 8 pPSs was required before a significant packaging phenotype was observed may explain why previous mutagenic screens have been unsuccessful in identifying these novel motifs[30] – if the pPSs act synergistically, a threshold of ablation must be reached before significant phenotypic changes are observed. Additional low-affinity pPSs presumably also exist which were not mutated in this analysis. There are multiple occurrences of our described motifs within the JFH-1 genome which were not identified by our original aptamer screen; it may be that a large number of functionally redundant pPSs exist and hence are able to easily compensate for many mutations which may occur during the error-prone RNA replication process *in vivo*.

Previous models of assembly which utilise multiple contributing packaging signals predict that highly efficient capsid assembly occurs in the presence of one higher-affinity RNA structure; initial binding of this motif to the structural protein then instigates additional multimerisation and concurrent interactions with the lower-affinity packaging signals[22]. The relative affinities of each of the pPSs described in this study for the Core protein would therefore be of significant interest.

Prior to the development of the infectious JFH-1 cell culture system, models of HCV virion assembly relied upon either self-assembled capsid-like particles generated from purified recombinant protein[15,31,32], or by expression of Core in yeast[31], insect cells[33] or bacterial systems[34] followed by purification of the intracellular capsid-like particles. The formation of such regular capsid-like particles in these protocols requires the presence of structured RNA, proving that protein-protein interactions are insufficient to drive Core multimerisation. It is now recognised that these models do not reflect the assembly process in mammalian cells, therefore although investigating the binding kinetics of each pPS and its respective mutant would be of significant interest, future work may have to utilise alternatives to such *in vitro* assembly models. Additionally, the fact that intact HCV genomes cannot be packaged in *in vitro* assembly systems[15] supports our model wherein the structures of interest may be formed temporarily and potentially only in certain scenarios, such as during presentation of nascent RNA within a replication complex. The NS5A protein (an essential component of the viral replication complex) possesses a broad RNA binding spectrum[35] and interacts specifically with Core protein[36]; this combined with the fact that Core exhibits extensive RNA chaperone activity[37,38] suggests a highly plausible model of RNA conformational alterations occurring during RNA transfer and subsequent assembly.

It is well established in the literature that Core binds to the viral 5′ UTR[15,39–42], although this interaction is not sufficient to drive genome packaging[43]. We did not focus upon this region as its interactions with Core are already well-explored; the preferred Core-RNA binding motif is the IIId loop and this interaction inhibits IRES-mediated translation[40,44]. This motif is the only region of the IRES which possesses similar features to the stem-loops selected for mutagenesis; specifically, the internal bulge and a G-rich stretch within the unpaired terminal loop. This suggests that multiple structures, each containing this motif in their terminal loop, may be required to interact with Core during RNA encapsidation. As this IRES subdomain also forms long-range interactions with CREs in the NS5B coding region[45,46], a model may be envisaged in which packaging, RNA replication and translation are mutually exclusive events, partially dictated by the ligand bound to the IRES IIId loop (Core, CRE RNA or the ribosomal complex).

The majority of our mutated pPSs are within the non-structural protein coding regions, with the exception of SL733. *Trans*-packaging of viral subgenomic replicons has been extensively documented (for review see)[47] and it has been noted by other research groups that the functionality of these systems reflects the lack of essential *cis*-acting RNA elements within the Core to NS2 genomic region[48]. This is supported by our data wherein only 2 of the 8 mutated pPSs are located in this area (SL733 and SL2899); their removal may not reach a threshold level to affect packaging efficiency. *Trans*-packaging studies utilising baculovirus-mediated structural protein expression also found that although the presence of the NS2 protein improves the production of replicon-containing particles, this was most apparent when NS2 is expressed in *cis* (within the replication-cassette), rather than in *trans* (within the structural protein construct)[49]. Although this may be due to the protein-protein interactions required for virion assembly[50], the fact that the pPS SL2899 is present upon this particular replicon may increase the efficiency of packaging, resulting in higher titres of virus-like particles in this system. In a parallel scenario, naturally occurring mutants with deletions spanning E1-p7 have been found in multiple patient samples; as these subgenomic mutants may be packaged and released[51,52] an essential packaging signal is definitively not located in those coding regions.

Given the well-recognised roles of RNA structure in the HCV life cycle, it is important to highlight the novelty of the mutated structures in this report as many previously-annotated structures were not altered by our silent mutations, including the NS5B-located CREs[24,30,53]. In addition, the genome of HCV possesses genome-scale ordered RNA structure (GORS); a phenomenon amongst particular RNA viruses which exhibit extensive RNA structure across the entire genome[54]. Consequently the genome is constrained in its plasticity and evolutionary potential, in contrast to the predicted behaviour of a RNA virus. It has been suggested that GORS contributes to viral persistence or modulates host innate defence mechanisms. It therefore must be considered that these pPS structures may additionally contribute to the overall architecture of the HCV genome and may not be solely involved in the Core-RNA interactions during packaging – they may have multiple functions, requiring distinct structural conformations to be adopted at precise points during the viral replication cycle. The Δ8XSL mutant may represent a virus which has reached a threshold level of GORS interruption. However this is unlikely; an equivalent effect upon replication and assembly would be predicted if this were the case. It is therefore apparent that our analysis does not merely reflect RNA structures already annotated within the HCV genome, but rather the identification of novel structures utilised for Core-RNA binding.

Materials and Methods

Recombinant protein expression.
Core (D1) of JFH-1 was cloned into the pET28B(+) construct (Novagen), in-frame with a N-terminal hexahistidine tag. Recombinant protein was expressed in BL21[DE3] cells (F– *ompT hsd*SB(rB– mB– *gal dcm* (DE3)). Luria broth media (1% glucose) was inoculated with a 1:1000 dilution of starter culture and maintained at 37 °C until an OD_{600} of 0.6–0.8 was reached. Protein expression was then induced by the addition of 100 mM isopropyl β-D-1-thiogalactopyranoside and cultures were maintained at 20 °C for 6 hours whilst shaking, before being harvested by centrifugation.

Protein purification.
Bacterial pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 7.4, 200 mM NaCl, 1% Triton-X-100, 10% glycerol, 0.06% β-mercaptoethanol, 20 mM imidazole, protease inhibitors, DNase1, RNase A, pH 7.5) and lysed by sonication. The soluble fraction was filtered (0.45 μm) before being loaded on

to an equilibrated Ni^{2+} Sepharose resin column (GE Healthcare). Resin was washed (50 mM Tris-HCl pH 7.4, 500 mM NaCl, 10% glycerol, 0.06% β-mercaptoethanol, 100 mM imidazole) and the protein was eluted into wash-based buffer with 500 mM imidazole. Protein was dialysed overnight into PBS and quantified by spectrometry. SDS-PAGE was conducted upon both bacterial lysates and eluted protein fractions according to previously described methods[55] using 12% gels. Silver staining was conducted according to manufacturer's instructions (SilverQuest kit, Invitrogen). Immunoblots were conducted with an in-house rabbit anti-core polyclonal sera (1:2000), commercial anti-His monoclonal antibody (1:1000, Invitrogen) and HRP-labelled secondary antibodies. For protein identification, Coomassie stained bands were excised, reduced, alkylated and digested with trypsin. The recovered peptides were analysed by LC-MS/MS and subsequent searching of the tandem MS data against the Uniprot sequence database.

Systematic evolution of ligands by exponential enrichment (SELEX).   Recombinant D1 (1 mg) was immobilised via the His-tag to 12 mg of Dynabeads His-Tag Isolation and Pulldown magnetic beads (ThermoFisher Scientific) following the manufacturers protocol. Excess protein was removed with 3 washes of 100 mM sodium phosphate pH 8, 1 M NaCl, 0.02% (v/v) Tween-20. Bead immobilised protein was then washed 3 times with selection buffer (PBS, 0.05% (v/v) Tween-20, 1× Roche complete protease inhibitor per 50 mL). A Biomek 2000 laboratory automation work station (Beckman Coulter) was used to perform 10 rounds of *in vitro* SELEX as described previously[29], using a N35 2′OH RNA library. This starting library was synthesised according to an experimentally and iteratively optimised protocol. It was purchased from AptaIT (Munich, Germany) who performed three rounds of synthesis, Next Generation Sequencing (NGS) and statistical analysis of NGS data from the library using the COMPAS software. In doing so the random region of the library has been optimised in respect of an equal distribution of nucleotide building blocks (25% of G, C, A & T, at each position), as well as in respect of a Gaussian distribution of motifs of length 4–8 nucleotides. An equal distribution of nucleotides increases the sequence space on the level of shorter motifs, which finally mediate binding to the target of interest. Consequently, the chance for a successful SELEX experiment increases with a homogeneous random region. Here the library was initially transcribed with a HiScribe T7 High Yield RNA transcription kit (New England Biolabs). Negative selections were carried out at each round of SELEX using bare His-Tag isolation and pulldown beads. The stringency of the aptamer-target interaction was raised by increasing the number of washes from 10 to 15 (slower off-rate) and by decreasing the amount of bead-immobilised protein by half (faster on-rate). Selections were carried out at 37 °C. Following the final round of SELEX, aptamers were analysed by NGS using a MiSeq Desktop Sequencer (Illumina). The SELEX library was prepared using the MiSeq reagent kit following the manufacturer's protocol. Raw sequencing reads were processed using in-house scripts to identify sequences that contained correct 5′ and 3′ primers. Processed sequences were used for packaging signal identification (see below).

Potential packaging signal identification.   The following HCV genotype 2 genome sequences were extracted from GenBank: accession numbers AY232741; AY232730; AY746460; D00944; D50409; DQ364460; AB030907.1; AB031663.1; HM777359; AB047640; AF238481; AF238485; AF238486; AY232749; AB047640.1. The final selected aptamer library contained 112,117 unique sequences, each 35 nucleotides in length that were aligned against the JFH-1 genome as follows: each aptamer sequence was slid along the genome in increments of 1 nucleotide. For each such position of the reference frame, the subset of the aptamer sequence with the best alignment to the genome was identified according to the Bernoulli score B, which benchmarks the probability of a non-contiguous alignment to that of a contiguous alignment of B nucleotides. The Bernoulli scores for all reference frames of a given aptamer sequence in the library were rank-ordered, starting from the largest score, and all matches with the genome up to a Bernoulli score of 12 counted. The procedure was then repeated for the other aptamer sequences, and corresponding matches added, resulting in the peaks in Fig. 2.

Identification of a consensus motif.   This analysis was repeated for all genomes listed above. Peaks which occurred for at least 60% of the genomes within <10 nucleotides of each other were marked by a green arrow to indicate their conservation. These were used as a basis to identify the highly conserved pPS recognition motif. For this, sequences of 60 nucleotides, centred around the peak nucleotide, were extracted and all possible mFold predicted folds were considered. A similarity analysis of these stem-loops was performed, comparing both sequence and structure elements. The foldings of each peak area which had the highest degree of similarity with secondary structure elements in the other peaks was identified. This returned a stem-loop for each peak; an alignment of the corresponding terminal loop sequences is displayed in Fig. 2. In order to identify which features of the consensus motif are statistically significant, the number of occurrence of stem-loops with different types of sub-motifs in the genome was calculated and benchmarked against their occurrence in randomised versions of the genome. The consensus motif was GRRGR; this motif was 2.05 times more likely to occur in the wildtype than randomised genomes. In order to refine this motif, we performed similar statistical tests for submotifs of this consensus motif; we identified GGRGG and GXRXR motifs as being more likely to occur in the wildtype genome than in a randomised version, by a factor of 2.36 and 1.14, respectively, suggesting that pPSs with the former motif could correspond to high affinity PSs, and those with the latter motif could correspond to lower affinity PSs.

Mutagenesis of pJFH-1.   The DNA construct containing the JFH-1 viral genome (pJFH-1) and a replication-defective control mutant of this plasmid (termed pJFH-1-GND, possessing a GDD > GND mutation within the NS5B RNA-dependent-RNA-polymerase active site) have been described previously[3]. Mutagenesis of pJFH-1 to introduce silent mutations was performed by either site-directed mutagenesis (QuikChange, Agilent) or through overlap PCR utilising mutagenic primers (primer sequences available upon request). The entire genome of all mutant viruses was confirmed through Sanger sequencing to ensure additional mutations were not present.

***In vitro* transcription.** DNA constructs were linearised with *Xba*I, briefly treated with mungbean nuclease to degrade 3′ overhangs (New England Biolabs) and purified by phenol-chloroform extraction. Linearised DNA was used as a template for *in vitro* transcription to produce full-length HCV genomic RNA (RiboMAX Express; Promega, as per the manufacturer's instructions). Following DNAse digestion, RNA transcripts were purified using silica-gel columns (RNEasy, Qiagen) and quantified by absorbance at 260 nm prior to transfection into mammalian cells.

Mammalian cell culture. Huh-7 cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Sigma) supplemented with 10% foetal bovine serum (FBS), 100 IU/mL penicillin, 100 μg/mL streptomycin, 25 mM HEPES and 1% (v/v) non-essential amino acids in a humidified incubator at 37 °C in 5% CO_2[56].

Infectious HCV assays. 6.0×10^6 Huh-7 cells were electroporated with 5 μg of viral RNA at 975 μF and 260 V for 25 ms. Cells were resuspended in complete medium and seeded into multiple 6 well plates (6.0×10^5 cells per well) (Corning). 4 hours post electroporation (h.p.e.), cells were harvested with TRIZol (Invitrogen Life Technologies) for RNA quantification. 24 h.p.e., all remaining wells were washed with PBS to remove excessive extracellular input RNA and the media replaced. 48 h.p.e., cells were harvested in TRIZol, protein lysis buffer or PBS, for RNA quantification, protein detection and intracellular virus titration, respectively. Virus supernatant was harvested at 48 h.p.e. for RNA quantification and released infectious virus titrations. Intracellular virus was isolated by 5 repeated freeze-thaw cycles of PBS cell lysates followed by clarification. Virus titres were calculated according to previously reported methods[25]. Briefly, either viral supernatant or cellular lysates were serially diluted two-fold upon naïve Huh-7 cells in 96 well plate format, seeded 6 hours prior to infection. At 48 hours post-infection, cells were washed, fixed with 4% paraformaldehyde and the viral NS5A protein was detected via indirect immunofluorescence. Automated counting was performed using the IncuCyte ZOOM (Essen Bioscience) using previously-described parameters[25].

Quantitative reverse-transcriptase PCR. RNA was purified by TRIZol extraction and diluted to 100 ng/uL. 200 ng was used in a Taqman probe-based one-step qRT-PCR according to previously-reported methods[57]. To quantify RNA, cycle threshold (Ct) values were converted to genome copies per 200ng of input RNA by comparison to a standard curve of serially-diluted pJFH-1 (10^2 to 10^9 copies).

SDS-PAGE and Western blots. Following lysis in Glasgow lysis buffer (1% Triton-X-100, 120 mM KCl, 30 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 10 mM PIPES-NaOH, pH 7.2 with protease inhibitors), lysates were clarified at $2000 \times g$ for 2 minutes before protein concentration was determined by BCA assay (Pierce). 30 μg of cellular protein was separated by electrophoresis on 12% polyacrylamide gels and analysed by Western blot to detect NS5A and GAPDH as described previously[55].

RNA mapping by selective 2′ hydroxy acylation analysed by primer extension (SHAPE). A 1527 nucleotide region across NS4B, flanked by the *Nsi*I and *San*DI restriction enzyme sites, was cloned into pCR-Blunt after PCR amplification from both pJFH-1 (WT) and Δ8XSL plasmid templates. The mutant insert therefore contained 19 silent point mutations ablating the predicted structures of SL5877 and SL6067. Subgenomic *in vitro* transcripts were then synthesised from a T7 promoter located upstream of the insert. RNA was heated to 95 °C for 3 minutes, before being refolded at 37 °C for 30 minutes in folding buffer (330 mM HEPES (pH 8.0), 20 mM $MgCl_2$ and 330 mM NaCl). Folded RNA was modified with either 100 mM N-methylisatoic anhydride or DMSO at 37 °C for 45 minutes. Primer extension reactions were conducted as described previously[24]. Primers were labelled with γ-[33p]-ATP (PerkinElmer) and purified by ethanol precipitation prior to primer extension. cDNA extension products were separated by denaturing electrophoresis with a 7% polyacrylamide TBE gel (7 M urea), at 60 W for between 1.5 to 2.5 hours. Gels were visualised with a phosphoimager and the final images were processed to obtain normalised SHAPE reactivity using SAFA v1.1 software[58].

References

1. Lindenbach, B. D. Virion assembly and release. *Current topics in microbiology and immunology* **369,** 199–218, doi: 10.1007/978-3-642-27340-7_8 (2013).
2. Lohmann, V. *et al.* Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* **285,** 110–113 (1999).
3. Wakita, T. *et al.* Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature medicine* **11,** 791–796, doi: 10.1038/nm1268 (2005).
4. Chatel-Chaix, L. & Bartenschlager, R. Dengue virus- and hepatitis C virus-induced replication and assembly compartments: the enemy inside–caught in the web. *Journal of virology* **88,** 5907–5911, doi: 10.1128/JVI.03404-13 (2014).
5. Paul, D., Madan, V. & Bartenschlager, R. Hepatitis C virus RNA replication and assembly: living on the fat of the land. *Cell host & microbe* **16,** 569–579, doi: 10.1016/j.chom.2014.10.008 (2014).
6. Roingeard, P., Hourioux, C., Blanchard, E. & Prensier, G. Hepatitis C virus budding at lipid droplet-associated ER membrane visualized by 3D electron microscopy. *Histochemistry and cell biology* **130,** 561–566, doi: 10.1007/s00418-008-0447-2 (2008).
7. Jones, D. M. & McLauchlan, J. Hepatitis C virus: assembly and release of virus particles. *The Journal of biological chemistry* **285,** 22733–22739, doi: 10.1074/jbc.R110.133017 (2010).
8. Miyanari, Y. *et al.* The lipid droplet is an important organelle for hepatitis C virus production. *Nature cell biology* **9,** 1089–1097, doi: 10.1038/ncb1631 (2007).
9. Schlesinger, S., Makino, S. & Linial, M. L. Cis-Acting Genomic Elements and Trans-Acting Proteins Involved in the Assembly of Rna Viruses. *Semin Virol* **5,** 39–49, doi: 10.1006/smvy.1994.1005 (1994).
10. Klein, K. C., Polyak, S. J. & Lingappa, J. R. Unique features of hepatitis C virus capsid formation revealed by de novo cell-free assembly. *Journal of virology* **78,** 9257–9269, doi: 10.1128/JVI.78.17.9257-9269.2004 (2004).
11. Dykeman, E. C., Stockley, P. G. & Twarock, R. Solving a Levinthal's paradox for virus assembly identifies a unique antiviral strategy. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 5361–5366, doi: 10.1073/pnas.1319479111 (2014).

12. Patel, N. *et al.* Revealing the density of encoded functions in a viral RNA. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 2227–2232, doi: 10.1073/pnas.1420812112 (2015).

13. Kopp, M., Murray, C. L., Jones, C. T. & Rice, C. M. Genetic analysis of the carboxy-terminal region of the hepatitis C virus core protein. *Journal of virology* **84,** 1666–1673, doi: 10.1128/JVI.02043-09 (2010).

14. Boulant, S., Vanbelle, C., Ebel, C., Penin, F. & Lavergne, J. P. Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *Journal of virology* **79,** 11353–11365, doi: 10.1128/JVI.79.17.11353-11365.2005 (2005).

15. Kunkel, M., Lorinczi, M., Rijnbrand, R., Lemon, S. M. & Watowich, S. J. Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. *Journal of virology* **75,** 2119–2129, doi: 10.1128/JVI.75.5.2119-2129.2001 (2001).

16. Kunkel, M. & Watowich, S. J. Conformational changes accompanying self-assembly of the hepatitis C virus core protein. *Virology* **294,** 239–245, doi: 10.1006/viro.2001.1325 (2002).

17. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–510 (1990).

18. Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–822, doi: 10.1038/346818a0 (1990).

19. Dykeman, E. C., Stockley, P. G. & Twarock, R. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *Journal of molecular biology* **425,** 3235–3249, doi: 10.1016/j.jmb.2013.06.005 (2013).

20. Rolfsson, O. *et al.* Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. *Journal of molecular biology*, doi: 10.1016/j.jmb.2015.11.014 (2015).

21. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* **31,** 3406–3415 (2003).

22. Dykeman, E. C., Stockley, P. G. & Twarock, R. Building a viral capsid in the presence of genomic RNA. *Physical review. E, Statistical, nonlinear, and soft matter physics* **87,** 022717 (2013).

23. You, S., Stump, D. D., Branch, A. D. & Rice, C. M. A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for hepatitis C virus RNA replication. *Journal of virology* **78,** 1352–1366 (2004).

24. Tuplin, A., Struthers, M., Simmonds, P. & Evans, D. J. A twist in the tail: SHAPE mapping of long-range interactions and structural rearrangements of RNA elements involved in HCV replication. *Nucleic acids research* **40,** 6908–6921, doi: 10.1093/nar/gks370 (2012).

25. Stewart, H. *et al.* A novel method for the measurement of hepatitis C virus infectious titres using the IncuCyte ZOOM and its application to antiviral screening. *Journal of virological methods* **218,** 59–65, doi: 10.1016/j.jviromet.2015.03.009 (2015).

26. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *Journal of the American Chemical Society* **127,** 4223–4231, doi: 10.1021/ja043822v (2005).

27. Kuo, L. & Masters, P. S. Functional analysis of the murine coronavirus genomic RNA packaging signal. *Journal of virology* **87,** 5182–5192, doi: 10.1128/JVI.00100-13 (2013).

28. Kim, D. Y., Firth, A. E., Atasheva, S., Frolova, E. I. & Frolov, I. Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. *Journal of virology* **85,** 8022–8036, doi: 10.1128/JVI.00644-11 (2011).

29. Bunka, D. H. *et al.* Degenerate RNA packaging signals in the genome of Satellite Tobacco Necrosis Virus: implications for the assembly of a T = 1 capsid. *Journal of molecular biology* **413,** 51–65, doi: 10.1016/j.jmb.2011.07.063 (2011).

30. Chu, D. *et al.* Systematic analysis of enhancer and critical cis-acting RNA elements in the protein-encoding region of the hepatitis C virus genome. *Journal of virology* **87,** 5678–5696, doi: 10.1128/JVI.00840-12 (2013).

31. Majeau, N. *et al.* The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation. *The Journal of general virology* **85,** 971–981 (2004).

32. Fromentin, R. *et al.* A method for *in vitro* assembly of hepatitis C virus core protein and for screening of inhibitors. *Analytical biochemistry* **366,** 37–45, doi: 10.1016/j.ab.2007.03.033 (2007).

33. Baumert, T. F., Ito, S., Wong, D. T. & Liang, T. J. Hepatitis C virus structural proteins assemble into viruslike particles in insect cells. *Journal of virology* **72,** 3827–3836 (1998).

34. Lorenzo, L. J. *et al.* Assembly of truncated HCV core antigen into virus-like particles in Escherichia coli. *Biochemical and biophysical research communications* **281,** 962–965, doi: 10.1006/bbrc.2001.4449 (2001).

35. Foster, T. L., Belyaeva, T., Stonehouse, N. J., Pearson, A. R. & Harris, M. All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. *Journal of virology* **84,** 9267–9277, doi: 10.1128/JVI.00616-10 (2010).

36. Gawlik, K. *et al.* HCV core residues critical for infectivity are also involved in core-NS5A complex formation. *PloS one* **9,** e88866, doi: 10.1371/journal.pone.0088866 (2014).

37. Cristofari, G. *et al.* The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA *in vitro*. *Nucleic acids research* **32,** 2623–2631, doi: 10.1093/nar/gkh579 (2004).

38. Ivanyi-Nagy, R., Lavergne, J. P., Gabus, C., Ficheux, D. & Darlix, J. L. RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. *Nucleic acids research* **36,** 712–725, doi: 10.1093/nar/gkm1051 (2008).

39. Shimoike, T., Mimori, S., Tani, H., Matsuura, Y. & Miyamura, T. Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *Journal of virology* **73,** 9718–9725 (1999).

40. Tanaka, Y. *et al.* Selective binding of hepatitis C virus core protein to synthetic oligonucleotides corresponding to the 5′ untranslated region of the viral genome. *Virology* **270,** 229–236, doi: 10.1006/viro.2000.0252 (2000).

41. Fan, Z., Yang, Q. R., Twu, J. S. & Sherker, A. H. Specific *in vitro* association between the hepatitis C viral genome and core protein. *Journal of medical virology* **59,** 131–134 (1999).

42. Hwang, S. B., Lo, S. Y., Ou, J. H. & Lai, M. M. Detection of Cellular Proteins and Viral Core Protein Interacting with the 5′ Untranslated Region of Hepatitis C Virus RNA. *Journal of biomedical science* **2,** 227–236 (1995).

43. Friebe, P. & Bartenschlager, R. Role of RNA structures in genome terminal sequences of the hepatitis C virus for replication and assembly. *Journal of virology* **83,** 11989–11995, doi: 10.1128/JVI.01508-09 (2009).

44. Shimoike, T. *et al.* Down-regulation of the internal ribosome entry site (IRES)-mediated translation of the hepatitis C virus: critical role of binding of the stem-loop IIId domain of IRES and the viral core protein. *Virology* **345,** 434–445, doi: 10.1016/j.virol.2005.10.013 (2006).

45. Shetty, S., Stefanovic, S. & Mihailescu, M. R. Hepatitis C virus RNA: molecular switches mediated by long-range RNA-RNA interactions? *Nucleic acids research* **41,** 2526–2540, doi: 10.1093/nar/gks1318 (2013).

46. Romero-Lopez, C. & Berzal-Herranz, A. A long-range RNA-RNA interaction between the 5′ and 3′ ends of the HCV genome. *Rna* **15,** 1740–1752, doi: 10.1261/rna.1680809 (2009).

47. Fournier, C., Duverlie, G. & Castelain, S. Are trans-complementation systems suitable for hepatitis C virus life cycle studies? *Journal of viral hepatitis* **20,** 225–233, doi: 10.1111/jvh.12069 (2013).

48. Steinmann, E., Brohm, C., Kallis, S., Bartenschlager, R. & Pietschmann, T. Efficient trans-encapsidation of hepatitis C virus RNAs into infectious virus-like particles. *Journal of virology* **82,** 7034–7046, doi: 10.1128/JVI.00118-08 (2008).

49. Adair, R. *et al.* Expression of hepatitis C virus (HCV) structural proteins in trans facilitates encapsidation and transmission of HCV subgenomic RNA. *The Journal of general virology* **90,** 833–842, doi: 10.1099/vir.2008.006049-0 (2009).

50. Jones, C. T., Murray, C. L., Eastman, D. K., Tassello, J. & Rice, C. M. Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. *Journal of virology* **81,** 8374–8383, doi: 10.1128/JVI.00690-07 (2007).

51. Sugiyama, K. *et al.* Genetic analysis of hepatitis C virus with defective genome and its infectivity *in vitro*. *Journal of virology* **83,** 6922–6928, doi: 10.1128/JVI.02674-08 (2009).
52. Pacini, L., Graziani, R., Bartholomew, L., De Francesco, R. & Paonessa, G. Naturally occurring hepatitis C virus subgenomic deletion mutants replicate efficiently in Huh-7 cells and are trans-packaged *in vitro* to generate infectious defective particles. *Journal of virology* **83,** 9079–9093, doi: 10.1128/JVI.00308-09 (2009).
53. Tuplin, A., Evans, D. J. & Simmonds, P. 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. *The Journal of general virology* **85,** 3037–3047, doi: 10.1099/vir.0.80141-0 (2004).
54. Simmonds, P., Tuplin, A. & Evans, D. J. 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, doi: 10.1261/rna.7640104 (2004).
55. Ross-Thriepland, D., Mankouri, J. & Harris, M. Serine phosphorylation of the hepatitis C virus NS5A protein controls the establishment of replication complexes. *Journal of virology*, doi: 10.1128/JVI.02995-14 (2014).
56. Nakabayashi, H., Taketa, K., Miyano, K., Yamane, T. & Sato, J. Growth of human hepatoma cells lines with differentiated functions in chemically defined medium. *Cancer research* **42,** 3858–3863 (1982).
57. Takeuchi, T. *et al.* Real-time detection system for quantification of hepatitis C virus genome. *Gastroenterology* **116,** 636–642 (1999).
58. Laederach, A. *et al.* Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nature protocols* **3,** 1395–1401, doi: 10.1038/nprot.2008.134 (2008).

Acknowledgements

Author Contributions

M.H conceived the study and designed experiments. H.S. expressed and purified recombinant protein. S.J.W. and P.G.S. designed and performed the SELEX reactions. E.C.D., R.J.B. and R.T. analysed the aptamer dataset. C.Z. and H.S. performed virus assays. A.K.T. and H.S. conducted SHAPE analysis. H.S. conducted all remaining experiments. H.S., M.H., P.G.S. and R.T. wrote the manuscript. All authors approved the manuscript prior to publication.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Stewart, H. *et al.* Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Sci. Rep.* **6**, 22952; doi: 10.1038/srep22952 (2016).