

Title

First Crystal Structure of a Non-Structural Hepatitis E Viral Protein Identifies a Putative Novel Zinc-Binding Protein

Running Title

First Structure of an HEV Non-Structural Protein

Byline

Andrew Proudfoot¹, Anastasia Hyrina², Meghan Holdorf², Andreas O. Frank¹, Dirksen Bussiere^{1,#}

Affiliation

¹Structural and Biophysical Chemistry Group, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, 5300 Chiron Way, Emeryville, California 94608, United States

²Infectious Disease Group, Novartis Institutes for BioMedical Research, 5300 Chiron Way, Emeryville, California 94608, United States

Corresponding Author

dirksen.bussiere@novartis.com

Abstract

Hepatitis E virus (HEV) is a 7.2 kb positive-sense, single-stranded RNA virus containing three partially overlapping reading frames, ORF 1-3. All non-structural proteins required for viral replication are encoded by ORF1 and are transcribed as a single transcript. Computational analysis of the complete ORF1 polyprotein identified a previously uncharacterized region of predicted secondary structure bordered by two disordered regions coinciding partially with a region predicted as a putative cysteine protease. Following successful cloning, expression and purification of this region, the crystal structure of the identified protein was determined and identified to have considerable structural homology to a fatty acid binding domain. Further analysis of the structure revealed a metal binding site, shown unambiguously to specifically bind zinc via a non-classical, potentially catalytic zinc-binding motif. We present analysis for the first time of this identified non-structural protein, expanding the knowledge and understanding of the complex mechanisms of HEV biology.

Importance

Hepatitis E virus (HEV) is an emerging virus found predominately in developing countries causing an estimated 20 million infections, which result in approximately 57,000 deaths a year. Although it is known that the non-structural proteins of the HEV ORF1 are expressed as a single transcript, there is debate as to whether ORF1 functions as a single polyprotein or if it is processed into separate domains via a viral or endogenous cellular protease. In the following paper, we present the first structural and biophysical characterization of a HEV non-structural protein using a construct that has partially overlapping boundaries with the predicted putative cysteine protease.

37 Based on the structural homology of the HEV protein with known structures, along with the
38 presence of a catalytic zinc-binding motif, it is possible that the identified protein corresponds to
39 the HEV protease, which could require activation or repression through the binding of a fatty acid.
40 This represents a significant step forward in the characterization and the understanding of the
41 molecular mechanisms of the HEV genome.

Introduction

Hepatitis E virus (HEV) is a non-enveloped, positive single-stranded RNA virus that infects approximately 20 million people annually, with a global mortality rate of 2% (1, 2). This fatality rate dramatically increases up to 30 % for infected pregnant women in their third trimester for unknown reasons (3). The virus is transmitted via the fecal-oral route and is most common in semi-tropical and developing countries with poor sanitation or contaminated water supplies (4), however cases of HEV infection reported in more industrialized countries in Europe, as well as in the USA and in Japan are increasingly more common (5–7). Although currently there is no accepted treatment for HEV infection, the off-label treatments of both interferon and/or ribavirin have been used successfully to treat chronic HEV infection (8, 9). As HEV is a growing worldwide threat, better understandings of the viral molecular mechanisms are necessary to help enable the development of targeted therapeutics.

HEV is a member of the *Hepeviridae* family comprised of four main mammalian genotypes; genotypes 1 and 2 are infection limited to only the human host, while genotypes 3 and 4 are zoonotic and have been identified in humans along with other animal species including swine (10). The HEV genome is approximately 7.2 kb in size and is comprised of three partially overlapping open reading frames (ORF) flanked by 5' and 3' untranslated / non-coding regions (11–13). ORF1 encodes all the non-structural proteins required for viral replication. ORF2 is the only ORF that has been structurally characterized in the HEV genome, and encodes the viral capsid protein that is involved with virion assembly (13). ORF3, the smallest of the three ORFs, encodes a 113 - 115 amino acid protein implicated to either encode an ion channel or aid in HEV virion release (14). In addition, the HEV genome contains a 7-methylguanosine cap at the 5' end, and two *cis*-reactive

elements that are dispersed through the viral genome, all of which are essential for viral replication (15, 16).

The largest of the three ORFs is ORF1, which is transcribed as a single 1693 residue polypeptide (17). In the absence of structural characterization, ORF1 has been compared to homologous viruses and computationally analyzed to identify eight putative domains (18). These include a methyltransferase domain (Met), Y domain (Y), papain-like cysteine protease (PCP), a proline-rich region that contains a hypervariable region (H), X-domain (X), helicase (Hel) and an RNA dependent RNA polymerase (RdRP) (Figure 1). With the exception of the Y-domain and papain-like cysteine protease, all identified domains have now been characterized to some extent, and experiments have been performed to confirm the predicted protein functions (19–21). The predicted protease is more elusive as a distinct protease is not identifiable in the HEV polypeptide sequence, however homology modelling of other alphaviruses homologous to HEV which contain a cysteine protease, predict a putative cysteine protease between residues 433 and 592 in ORF1 (18).

Previous research has been unable to confirm the processing of ORF1 or the presence of a protease. Some reports show that ORF1 is expressed as a single 186 kDa polypeptide in *E. coli*, insect and mammalian cells, and in a cell free system (22). Although other reports support polypeptide processing of ORF1, when expressed in mammalian cells two peptides of approximately 107 and 78 kDa in size are observed (23). Likewise, expression of ORF1 in insect cells results in eight small peptides which can be inhibited in the presence of a cell-permeable cysteine protease inhibitor (24). Current research on the purified putative protease domain *in vitro* has been limited. Protein purified under denaturing conditions has been shown upon refolding to have proteolytic activity against ORF1 and ORF2 products, and mutation analysis of highly conserved cysteine and

87 histidine residues in the predicted protease inhibit proteolytic activity and are essential for viral
 88 replication (25). A three-dimensional model of the predicted protease region has been produced
 89 using consensus-prediction results of the closely related p150 rubella protease (26), however
 90 characterization of the predicted protease domain is required in order to understand better the viral
 91 biology. In this study, we report a comprehensive analysis of the biochemical, biophysical and
 92 structural parameters of the predicted protease domain that may help elucidate the function of the
 93 protein in the HEV lifecycle.

Results

All previous work performed on the HEV non-structural proteins of ORF1 has been done based on the boundaries originally proposed by Koonin *et al* (18) (Figure 1A), with the resulting protein constructs typically produced with low yields or requiring refolding from the insoluble fraction following purification (11, 20, 25). To evaluate if these difficulties were due to distinct differences between the actual and proposed domain boundaries, ORF1 was analyzed using secondary structure prediction to identify regions of order and disorder within the open reading frame (Figure 1B). The HEV X-domain, helicase and RNA-dependent RNA polymerase show high levels of homology to other homologous proteins, so it was not surprising that secondary structure is predicted between residues 780 – 1693 in ORF1. Our analysis also showed that secondary structure elements at the N-terminus of ORF1 are sandwiched between four regions of disorder within ORF1, residues 1 – 8, 270 – 273, 492 – 509 and 692 – 779. The region of predicted disorder between residues 692 – 779 aligns with the region previously identified to be the disordered hinge region (residues 712 – 778), with the three remaining regions of disorder bordering three regions of predicted secondary structure. The two regions of order identified between residues 9 – 269 and 274 – 491 coincide with the previously identified methyltransferase and Y-domain respectively (18), but surprisingly our analysis revealed a previously unidentified predicted domain between residues 510 – 691. Prior analysis of this region had shown residues 433 – 592 contained the putative cysteine protease, however the prediction used to locate the putative protease was not conclusive and was based on two assumptions: HEV contains a cysteine protease and the proposed catalytic residues align with the identified distant homologous proteases (18). In addition, subsequent mutagenesis studies and homology modeling of this region have been performed to support the presence of the protease, but a protein structure and elucidation of the exact catalytic

mechanism is required to fully understand any potential HEV protease (25–27). To confirm the presence of a domain between residues 510 and 691, a construct containing this region was codon-optimized, cloned and expressed in *E. coli* as an MBP-fusion protein (HEV⁵¹⁰⁻⁶⁹¹). Following cleavage of the fusion protein and purification, a final yield of approximately 120 mg of soluble protein per liter of unlabeled medium was achieved, and analysis by 1D NMR confirmed that the protein was folded. This high-level of expression suggests that this protein is extremely stable.

The protein was concentrated to 10 mg/ml and 864 crystallization conditions were screened at two temperatures, 4 °C and 20 °C. Two very similar conditions, which differed only by the type of nitrate present, provided initial hits (Figure 2A). Optimized crystals grew at 4 °C in precipitant consisting of 18 % w/v PEG 3350, 180 mM LiNO₃ and 10 mM NiCl₂ (Figure 2B) and diffracted to 1.8 Å resolution on an in-house X-ray source. Analysis of the RCSB showed that HEV⁵¹⁰⁻⁶⁹¹ did not have sequence homology to any deposited structure, so molecular replacement was not possible. Given this, the structure was solved using a mercury heavy-atom derivative and single-isomorphous replacement with anomalous scattering (SIRAS) (Table 1). The structure consists of ten β-strands and four α-helices, with the β-strands 1 – 4 and 5 – 10 arranged in two antiparallel sheets arranged in a sandwich-like fold. α-helices 1 & 2 are located between β-strands 1 & 2 and α-helices 3 & 4 are positioned at the C-terminus of the protein, with α-helix 4 sitting between the two antiparallel β-sheets (Figure 2C).

The crystal structure was analyzed using DALI to identify if there were any structurally homologous proteins in the PDB (28). The structural alignment showed that although there was only 13 – 15% amino acid conservation with the identified structural homologs, residues 514 – 635 of HEV⁵¹⁰⁻⁶⁹¹ has significant structural homology to multiple fatty acid binding domains, with the homologous regions aligning with a backbone RMSD of ~ 2.5 Å (Figure 3A & B). This

homologous region of the HEV protein corresponded to all β -strands and the first two α -helices, while helices 3 and 4 were not present in any of the structural homologs identified. In our structure, α -helix 4 is located between the anti-parallel β -sheets, and analysis of the identified homologous structures showed fatty acids bound in the same location (Figure 3C). An additional HEV construct from residues 510 – 635 (HEV⁵¹⁰⁻⁶³⁵), which has boundaries homologous to the identified fatty acid domains was produced. However, in this case the removal of the C-terminal residues abolished all protein expression, indicating that the additional residues and resulting secondary structure were likely part of the identified domain and may play an important role in protein folding and/or function.

Analysis of the region around α -helices 3 and 4 identified that there was density corresponding to a bound metal ion, coordinated by two residues, His 671 and Glu 673. In addition, another residue, His 686, was positioned in close proximity to act potentially as the third co-ordination site. However, in our structure, His 686 was in the opposite conformation and the side chain was not oriented to coordinate the bound metal (Figure 4A). Native HEV⁵¹⁰⁻⁶⁹¹ crystals soaked in a 1 mM solution of EDTA showed signs of cracking following 1 hour of soaking and the resulting datasets showed no electron density in the region where we had previously identified the bound metal. This confirmed that the identified density corresponded to a metal or some other entity that can also be coordinated by EDTA. In the HEV⁵¹⁰⁻⁶⁹¹ crystal structure, the density corresponding to the metal atom could not be fully satisfied by a single metal during refinement. This indicated that there were either different metal species present (i.e. the endogenous metal in addition to metal from either the purification or the crystallization conditions), or that the binding site was not fully saturated with the endogenous metal and thus was partially occupied. To identify which endogenous metal the protein had highest affinity for, DSF was used to screen a number of

different possibilities (Table 2) to see which, if any, metal could induce an increase in thermal stability. Of the 17 metals tested, zinc was the only metal that stabilized the protein and induced an increase in thermal stability of 3.0 °C (Figure 4B). Datasets acquired with HEV⁵¹⁰⁻⁶⁹¹ crystals soaked in a 1 mM solution of zinc chloride showed zinc at this position. However, the His 686 sidechain was still not orientated properly to act as the final zinc co-ordination site.

His 686 is located close to the end of the crystalized construct, so it is plausible that additional residues at the C-terminus of the construct could induce the formation of additional secondary structure and change the orientation of the His 686 side chain. To test this, additional constructs were designed selecting C-terminal boundaries based upon the presence of stretches of low complexity amino acid sequences (i.e. poly-glycine, -alanine, -serine or combinations thereof) with the additional consideration of not adding a significant number of disordered residues to the HEV⁵¹⁰⁻⁶⁹¹ construct. A construct from residues 510 – 720 (HEV⁵¹⁰⁻⁷²⁰) was produced that was 29 residues longer than HEV⁵¹⁰⁻⁶⁹¹, and was analyzed by NMR. Comparison of HEV⁵¹⁰⁻⁷²⁰ with the HEV⁵¹⁰⁻⁶⁹¹ showed that the additional residues had amide chemical shifts in the region of 7.5 – 8.5 ppm, which is indicative of disordered amino acids (Figure 5A), and coincided with the performed secondary structure prediction (Figure 1B). Analysis by DSF showed that in the presence of zinc, the HEV⁵¹⁰⁻⁷²⁰ had a greater increase in thermal stability (5.5 °C) compared to the HEV⁵¹⁰⁻⁶⁹¹ (Figure 5C). To optimize the construct further, additional C-terminal truncations were made and biophysically characterized to identify a construct that both added the fewest disordered residues and induced an equivalent increase in protein thermal stability in the presence of zinc. A construct from residues 510 – 696 (HEV⁵¹⁰⁻⁶⁹⁶) was identified to be the optimal construct and exhibited a 6 °C increase in thermal stability in the presence of zinc (Figure 5C), but did not contain an additional

increase in disorder in the NMR spectrum (Figure 5B). HEV⁵¹⁰⁻⁶⁹⁶ was used for crystal screening, but no crystals were observed with this construct in the presence or absence of zinc.

When comparing the [¹H – ¹⁵N] HSQC spectra of HEV⁵¹⁰⁻⁶⁹¹ and HEV⁵¹⁰⁻⁶⁹⁶, although many of the signals in the two spectra overlapped, there were also a number of amide resonances that experienced a significant chemical shift perturbation. Assignments of all amide peaks were obtained, and the residues experiencing a significant chemical shift perturbation were mapped to the three potential zinc co-ordination residues, the residues at the C-terminus and a stretch of amino acids on β -strand 4, between residues 572 – 574 (Figure 5D). Although these constructs did not crystallize, it can be inferred from the NMR data that the additional C-terminal residues form a β -strand, running antiparallel to β -strand 4. This would also allow His 686 to be oriented in such a way as to act as the final zinc co-ordination site. To confirm if His 671, Glu 673 and His 686 were involved with zinc co-ordination, each of the three residues were mutated individually and together to an alanine residue. Mutation of each residue contributed approximately a 2 °C reduction in thermal stability when compared to HEV⁵¹⁰⁻⁶⁹⁶ (approximately 2 °C for each single mutation and 6 °C for the triple mutant). However in comparison to HEV⁵¹⁰⁻⁶⁹⁶, all mutant constructs failed to show any increase in thermal stability in the presence of zinc (Figure 6), indicating that all three residues are involved with zinc co-ordination.

Work characterizing the predicted HEV protease has been performed with constructs that closely matched the boundaries identified by Koonin *et al* (18, 25, 29) (Figure 1A). To investigate the effect of additional residues at the N-terminus of HEV⁵¹⁰⁻⁶⁹¹, additional constructs were produced that extended the N-terminal boundaries to residues 455 (HEV⁴⁵⁵⁻⁶⁹¹), 440 (HEV⁴⁴⁰⁻⁶⁹¹) and 403 (HEV⁴⁰³⁻⁶⁹¹). These boundaries were selected as described above, in regions that had been predicted not to contain regular secondary structure. Previous work identified that *E. coli* BL21

cells expressing residues 440 – 610 experienced cell death upon induction at different temperatures and various concentrations of IPTG (25). Consistent with these observations, when the three N-terminally extended constructs were expressed in unlabeled medium in the absence of trace metals, cell growth was significantly inhibited following induction of the target proteins with IPTG. However, when *E. coli* were grown in unlabeled medium that had been supplemented with 1 x trace metals, normal cell growth and soluble protein expression was observed following induction with IPTG.

As the number of residues at the N-terminus increased, a corresponding reduction in protein expression was observed, with HEV⁴⁴⁰⁻⁶⁹¹ identified to have comparable protein expression to the shorter constructs tested, which coincided with the previously identified putative protease domain (18). Following gel filtration of HEV⁴⁴⁰⁻⁶⁹¹, it was apparent that the purified protein eluted from the gel filtration column at a much lower retention volume (Figure 7A) than expected (at a retention volume corresponding to a protein of Mw 270 kDa) and was brown in color. Mass spectrometry was performed with the protein species under denaturing conditions and a single peak at 27136.33 Da was observed (Figure 7B), which corresponded to the expected molecular weight of the monomeric protein and indicated that the protein eluting from the column was oligomerized. It was likely that the brown hue was coming from a bound metal and to identify if this was the case, the protein was thermally denatured, spun down and the supernatant was analyzed with 5F-BAPTA. 5F-BAPTA is divalent cation chelator with two fluorine atoms positioned at simultaneous positions, which yields a single peak in the ¹⁹F NMR spectrum. This signal can be monitored using ¹⁹F NMR as different bound metal ions induce different characteristic downfield chemical shift perturbations to the free 5F-BAPTA signal (30). Although a strong brown color was observed in the protein pellet following denaturation and centrifugation, the metal that remained in solution

231 induced a 27.8 ppm chemical shift perturbation to the 5F-BAPTA signal (Figure 7C) which is
232 consistent with the observed chemical shift upon 5F-BAPTA interacting with divalent iron (Fe^{2+})
233 (28.1 ppm).

Discussion

Detailed characterization of the non-structural proteins in ORF1 has proven to be a significant impediment in the understanding of the HEV genome. Initial structural analysis found regions in the HEV ORF1 that were predicted to correspond to a methyltransferase, Y-domain, disordered hinge, X-domain, helicase and RNA dependent RNA polymerase. Based on other positive-sense RNA viruses, it was hypothesized that a protease should also be encoded within ORF1. However, structural analysis could not conclusively identify a corresponding region, and the proposed cysteine protease was assigned to a region between residues 433 and 592 (18).

Our analysis of the HEV ORF1 identified that there is a region of ordered secondary structure (residues 510 – 691) sandwiched between two regions of disorder (residues 492 – 509 and 692 – 779). Overexpression of this region resulted in a soluble, stable protein, suitable for crystallization. The resulting protein structure of the HEV⁵¹⁰⁻⁶⁹¹ displayed a high structural homology to fatty acid binding domains, that when compared to homologous fatty acid binding proteins, revealed two additional α -helices present at the C-terminus of the construct. One of the two additional α -helices is located between the two β -sheets, in the same position where homologous fatty acid binding proteins bind their respective fatty acids (Figure 3C). Removal of these two additional α -helices from the HEV construct resulted in no expression of the protein, suggesting that this region was important for structural stability.

Further analysis of the additional α -helices showed the presence of a bound zinc metal ion coordinated by residues His 671, Glu 673 and His 686, which was confirmed by DSF and site-directed mutagenesis studies (Table 2 and Figure 6). Each single mutant had a similar T_m in the absence of zinc (within 2°C) suggesting that each of these mutants abolished zinc coordination but

did not adversely affect the protein fold. Although it is currently unknown if the coordinated zinc performs a catalytic or structural role, coordination of zinc for the purpose of structural integrity is predominantly achieved through a combination of cysteine and histidine residues, whereas acidic amino acids are included in the coordination site when the bound zinc is used for catalytic purposes (31). Analysis of all annotated HEV sequences present in Uniprot show complete conservation of residues Glu 673 and His 686, whereas residue 671 fluctuates between a histidine (78%) and a tyrosine (22%) in the available deposited sequences.

Taking into consideration the high conservation of the coordinating residues along with the observation that the protein has a potential catalytic zinc coordination site, it is possible that the solved structure corresponds to a zinc metalloprotease. All metalloproteases require an acidic residue in close proximity to the coordinated zinc to act as the catalytic residue during the proteolytic reaction. Analysis of the amino acid conservation of other residues in close proximity to the bound zinc identified the highly conserved Glu 583, which is positioned between β -strands 5 and 6, which could potentially serve as the catalytic residue. As expected, mutagenesis of Glu 583 to alanine did not have any significant effect on the induced increase in thermal stability upon addition of zinc as this residue is not directly involved with zinc coordination and thus does not directly stabilize the proteins tertiary structure. In the crystal structure of HEV⁵¹⁰⁻⁶⁹¹, the current geometries of the zinc coordinating residues are not optimal to catalyze a proteolytic reaction, however there remains the possibility that the binding of a fatty acid or some other endogenous ligand between the two β -sheets could displace α -helices 3 and 4, re-orientating the zinc coordinating amino acids into a catalytically active orientation. While we have not identified the endogenous ligand that interacts with the domain, it is well documented that the liver is rich in fatty acids so binding of a specific fatty acid to the protease could be used by the virus as a

regulatory mechanism. In addition, it is known that the biochemical profile of the liver changes throughout pregnancy (32). Thus, if the protein has a specificity for a fatty acid that is at elevated levels in pregnant women, this could also explain why the virus has increased mortality and morbidity in pregnant women in their third trimester (3).

There are no reports in the literature either unambiguously confirming or refuting the existence of a protease in HEV, with assays performed *in vitro* and in different cell lines reporting several distinct cleavage products (22–25, 27, 33, 34). If the HEV protease requires binding of a particular fatty acid to correctly form the catalytic site for ORF1 processing, this could potentially explain these results, as the overall replication rate of the virus would be determined by the concentration of a particular fatty acid in the different cell strains tested.

In an attempt to validate the potential protease activity and if the bound zinc plays a role in viral replication, we attempted to implement mutations to the zinc coordinating residues in a HEV replicon, based off of the well-studied pSK-HEV2 construct (GenBank: AF444002.1) (16, 35). Wild type and mutant replicons, containing the previously described replication-abolishing GDD to GAD mutation in the conserved motif of the viral RNA-dependent-RNA polymerase (35), were synthesized with a NanoLuc-PEST reporter positioned in frame with the ORF2, and sequence verified. Studies with the pSK-HEV2 replicon typically require the use of the Huh7-S10-3 liver cell line (36), which we were unable to secure access to, and unfortunately all other cell lines tested in-house were unable to support HEV replication. Although we could not execute studies with the HEV replicon, these experiments are vital to gain a more detailed understanding of the HEV life cycle and we are more than willing to provide our replicons to any research group who would be able to follow up with these studies.

Region 510 – 696 was identified to sit between two predicted regions of disorder, so it is plausible that the crystal structure constitutes the complete domain. To confirm this, additional constructs were produced that extended the HEV⁵¹⁰⁻⁶⁹¹ at both the N- and C- termini and analyzed. NMR studies determined that additional residues added at the C-terminus were disordered, which was consistent with both the secondary structure prediction and the assessment that these residues were part of the disordered ‘hinge’ region. Incremental addition of residues at the N-terminus caused a significant reduction in protein expression, with no protein expression observed with HEV⁴⁰³⁻⁶⁹¹ or other constructs extended at the N-terminus. Addition of 70 amino acids to the N-terminus (HEV⁴⁴⁰⁻⁶⁹¹) of the HEV⁵¹⁰⁻⁶⁹¹ did not seem to have an adverse effect on protein expression, but resulted in the production of a protein oligomer, which was brown in color due to the presence of bound iron. The 70 additional N-terminal residues added contain six cysteine residues, some of which had previously been implicated as potential catalytic residues for the HEV protease (25). Contrary to this report, we would like to propose that these cysteine residues are involved in the formation of an iron-sulfur cluster. This is consistent with observations of iron-sulfur clusters in other hepatitis viruses (38).

As HEV continues to infect a significant number of individuals each year, a greater understanding of the viral biology is required so dedicated antiviral therapeutics can be developed. Until now, the presence of a protease or other non-structural protein between the predicted Y-Domain and hyper-variable region in ORF1 was based upon comparison to other positive-sense RNA viruses and mutagenesis data of residues predicted to be involved with a proteolytic mechanism. Although published data is ambiguous about the presence of a protease in ORF1, the results published in this manuscript confirm the presence of a non-structural protein in the HEV ORF1 between residues 510 and 691. Having characterized the domain, both structurally and biophysically, it is clear the

324 protein has structural homology to a fatty acid binding domain, contains a non-characteristic
 325 catalytic zinc-binding motif and could potentially act as a zinc metalloprotease. While we have
 326 shown *in vitro* that the identified zinc coordinating residues are essential for metal coordination,
 327 we have been unable to confirm the importance of the zinc-binding motif in viral replication.
 328 Although we were unable to perform these studies, it is our hope that another research group will
 329 follow-up on characterizing the identified domain *in vivo*. It is our belief that the data presented
 330 here has provided a significant advancement in the understanding and characterization of the HEV
 331 non-structural proteins and we hope that it will form the building blocks for a more detailed
 332 understanding of HEV biology.

Materials and Methods

Cloning of HEV Constructs

The complete HEV ORF1 (Uniprot ID: P33424) was synthesized as an *E. coli* codon optimized gene from Synthetic Genomics Inc. Constructs were amplified from the complete optimized ORF1 sequence with primers containing 20 bp of homology (Table 3) to the pET24a plasmid that had been modified to contain a His₈-MBP-TEV sequence between the NdeI and BamHI restriction sites. The PCR product was analyzed by gel electrophoresis using a precast 1 % agarose gel (Invitrogen), and the PCR products of the desired length were purified using a QIA quick Gel Extraction kit (Qiagen). The purified PCR product was assembled into the vector using Gibson Assembly (NEB Gibson Assembly 2x Master Mix), and transformed into *E. coli* 10G chemically competent cells (Lucigen, Middleton, WI). Colonies resistant to kanamycin (Kan) were grown in LB medium, and the plasmid was isolated via Miniprep (Qiagen) prior to being sequence verified using forward and reverse primers. Site directed mutagenesis of the HEV⁵¹⁰⁻⁶⁹⁶ was performed using primers detailed in Table 1 and the Q5 Site-directed Mutagenesis Kit (NEB).

Growth and Expression of HEV MBP-fusion Proteins

E. coli BL21(DE3) chemically competent cells (NEB product # C2530H) were transformed with 100 ng of pET24a His₈-MBP plasmid (Thermo Fisher Scientific) encoding the desired HEV constructs and incubated overnight at 37 °C on LB Agar plates containing kanamycin (50 µg/mL). 50 mL of nutrient rich LB medium containing kanamycin (50 µg/mL) was inoculated with a single colony and grown overnight at 37 °C with agitation at 250 rpm. Using the overnight starter culture, 1 L of Terrific Broth (TB) supplemented with 50 mM 3-(N-morpholino) propanesulfonic acid

(MOPS) pH 7.5, 1 x Trace Metals, and kanamycin (50 µg/mL) was inoculated to a starting optical density at 600 nm (OD₆₀₀) of 0.1. If labeled protein was required for NMR experiments, M9 minimal medium where 1 g/L of ¹⁵N-ammonium chloride and 3 g/L of protonated ¹³C-glucose were used as the principal nitrogen and carbon sources, respectively was used instead of TB medium. Cells were grown at 37 °C with agitation at 250 rpm until an OD₆₀₀ of 0.8 was achieved and at this stage, the temperature of the incubator was reduced to 18 °C and expression of all constructs was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were left growing at 18 °C for 18 hours and were harvested by centrifugation, washed with phosphate buffered saline (PBS) buffer and stored at -20 °C.

Purification of HEV Proteins

Purification of all constructs were performed using the same protocol. *E. coli* cells containing the overexpressed HEV proteins were re-suspended in binding buffer (10 mL/g of cell pellet; 50 mM Tris pH 8.0, 300 mM sodium chloride, 25 mM imidazole), containing Roche Protease Inhibitor without ethylenediamine-tetraacetic acid (EDTA) and passed five times through a microfluidics cell homogenizer (Microfluidics modelM-110P cell homogenizer) at 18,000 psi. Cell lysate was centrifuged at 35,000 g for 60 minutes to remove insoluble cell debris and the soluble lysate was loaded onto a 5 mL His Trap HP Ni column (GE Healthcare) at a rate of 1.5 mL/minute. The Ni resin was washed with 100 mL of binding buffer to remove all non-specifically bound protein before the desired MBP-fusion proteins were eluted from the column using 100 mL elution buffer (50 mM Tris pH 8.0, 300 mM sodium chloride, 500 mM imidazole).

All fractions identified to contain the desired protein were combined and placed in a 30 mL dialysis cassette (Thermo Fisher) with 100 units His tagged TEV protease per 1 mg of MBP-fusion protein,

and dialyzed against 2 L dialysis buffer (20 mM Tris pH 8.0, 300 mM sodium chloride, 10 mM imidazole) overnight at 4 °C. The protein was further purified using reverse IMAC with the cleaved protein being loaded onto a 5 mL His Trap HP Ni column (GE Healthcare) at a rate of 1.5 mL/minute. The Ni resin was washed with binding buffer until all cleaved protein eluted from the column. All fractions identified to contain the cleaved protein were concentrated to a volume of 5 mL and loaded onto a HiLoad 16/60 Superdex 75 gel filtration column (GE Healthcare) that had been pre-equilibrated with either crystallography buffer (20 mM Bis-Tris pH 7.0, 150 mM NaCl, 1 mM Dithiothreitol (DTT)) or NMR buffer (20 mM Sodium Phosphate pH 7.0, 150 mM NaCl, 1 mM deuterated DTT) at a rate of 1.5 mL/minute. Samples were concentrated to a final concentration of approximately 10 mg/mL for crystallography experiments, or 1 mM for NMR experiments, flash frozen in liquid nitrogen and stored at -80 °C until needed.

Crystallography

Crystallization screening was performed using sitting drops, where 400 nL of an equal mixture of protein and precipitant (200 nL each) were incubated against 60 µL of precipitant. Screening was run at two temperatures, 4 and 20 °C, and trays were monitored for crystallization events using robotics. Suitable crystals grew at 4 °C using 18 % w/v PEG 3350, 180 mM LiNO₃ and 10 mM NiCl₂. Crystals were cryopreserved using precipitant supplemented with 20 % v/v of glycerol for data collection. All X-ray data was collected on a Rigaku FR-E+ SuperBright generator using a Pilatus R 300K CCD.

As no suitable homologous search model existed for molecular replacement and there was only a single methionine in the protein sequence, the crystal structure was solved using single isomorphous replacement with anomalous scattering (SIRAS). A suitable mercury derivative was

generated by soaking crystals in precipitant supplemented with 1 mM of thiomersal for 3 hours. The mercury covalently reacts with the protein primarily at the reactive cysteine residue Cys 649. Initial Patterson map interpretation, phasing, and initial structure determination were executed using SOLVE (39). This initial model comprised approximately 2/3rd of the protein sequence and the remainder of the model was manually built into 2Fo-Fc and Fo-Fc maps using COOT (40). All coordinates were refined to convergence using BUSTER and the PHENIX suite of programs (41–43). The statistics of data used for refinement is reported in Table 1.

NMR Experiments

All NMR experiments were conducted with 500 µL of sample containing 5% D₂O in a 5 mm NMR tube and were recorded at 298 K on a Bruker AVANCE III 600 MHz spectrometer equipped with a 5 mm CP-QCI-F z-gradient probe. 2D [¹⁵N, ¹H]-HSQC experiments were acquired with uniform sampling collecting a total of 1024 and 256 points in the direct and indirect dimensions, respectively. 3D HNCA, HNCO, HN(CA)CO, HN(CO)CA and CBCA(CO)NH experiments were acquired for the assignment of backbone atoms using 50 % non-uniform sampling (44), collecting 2048 points in the direct dimension and a total of 3300 points across all indirect dimensions. Non-uniformly sampled datasets were collected using Poisson Gap sampling schemes and processed in Topspin 3.5 using the hmsIST algorithm (45, 46). All data for the assignment of backbone atoms was analyzed in CARA (47) and residues which experienced a change in chemical shift between the different constructs were mapped onto the protein structure in Pymol (48).

Differential Scanning Fluorimetry

423 19 μ l of Apo HEV construct (5 μ M) was incubated with 10X Sypro Orange fluorescent dye in
424 crystallography buffer and 1 μ l of additive. The 20 μ l reaction was incubated at 25 $^{\circ}$ C for 5 min
425 before the temperature was gradually increased to 95 $^{\circ}$ C at a rate of 1.8 $^{\circ}$ C/min with fluorescence
426 readings being taken every 0.35 $^{\circ}$ C. Additives were screened using an 8-point 2-fold dilution series
427 of metals with the top concentration analyzed being 1.25 mM. All measurements were performed
428 in replicates of four, with an average standard deviation of 0.27 $^{\circ}$ C observed between the replicates.

429 **Acknowledgments**

430 We thank Drs. Andreas Lingel and Donald Ganem for their advice and mentorship during the
431 project. We also thank Barbara Leon and Dr. Tobias Flecken for advice in regards to construct
432 design, protein expression and purification as well as their useful comments and discussion.
433 Finally, we thank Drs. Suzanne Emerson and Dianjun Cao for helpful discussions about the HEV
434 replicon. All authors have no conflicts of interest to declare.

435 Tables

Coordinate Set	HEV ⁵¹⁰⁻⁶⁹¹ Native	HEV ⁵¹⁰⁻⁶⁹¹ Thiomersal Derivative	HEV ⁵¹⁰⁻⁶⁹¹ Zinc Co-Structure
PDB ID			6NU9
Data Collection Statistics			
Resolution (Å)	45.56–1.95 (2.02-1.95)	45.54-2.10 (2.18-2.10)	52.62–1.76 (1.82–1.76)
Space group	P6 ₅ 22	P6 ₅ 22	P6 ₅ 22
Cell dimensions a, b, c (Å) α , β , γ (°)	64.49, 64.49, 157.41 90.00, 90.00, 120.00	64.20, 64.20, 158.70 90.00, 90.00, 120.00	64.40, 64.40, 158.90 90.00, 90.00, 120.00
Total reflections	891654 (69081)	378180 (29740)	680486 (19154)
Unique reflections	14844 (1430)	21378 (2114)	20137 (1960)
Multiplicity	60.1 (48.3)	17.7 (14.1)	33.8 (9.8)
Completeness (%)	99.4 (99.2)	99.9 (98.9)	99.9 (99.3)
I/ σ (I)	56.6 (22.5)	20.5 (8.1)	13.4 (1.2)
R _{merge}	0.068 (0.218)	0.101 (0.333)	0.228 (1.028)
R _{pim}	0.009 (0.032)	0.023 (0.090)	0.038 (0.334)
CC1/2	0.999 (0.998)	0.999 (0.975)	0.997 (0.339)
Phasing Statistics			
Resolution (Å)		38.24-2.50	
Heavy atom sites		3 Hg, 2 Ni	
FOM		0.400	
Refinement Statistics			
R _{work}			0.236
R _{free}			0.268
Number of non-H atoms			
protein			1306
ligands			10
solvent			339
Protein residues			
RMS bonds(Å)			0.004
RMS angles (°)			0.66
Ramachandran favored (%)			96.30
Ramachandran allowed (%)			3.70
Ramachandran outliers (%)			0
Clashscore			3.81

Average B-factor (\AA^2)			17.51
-------------------------------------	--	--	-------

436

437 **Table 1:** Data collection and refinement statistics of the HEV⁵¹⁰⁻⁶⁹¹ X-ray structure.

438

Metals that have a positive impact on protein stabilization			
Zinc			
Metals that destabilize of do not impact protein stability			
Barium	Cadmium	Cesium	Calcium
Chromium	Cobalt	Copper	Iron
Lithium	Magnesium	Manganese	Nickel
Potassium	Praseodymium	Strontium	Yttrium

439

440 **Table 2:** Metals screened by DSF for binding to HEV⁵¹⁰⁻⁶⁹¹.

441

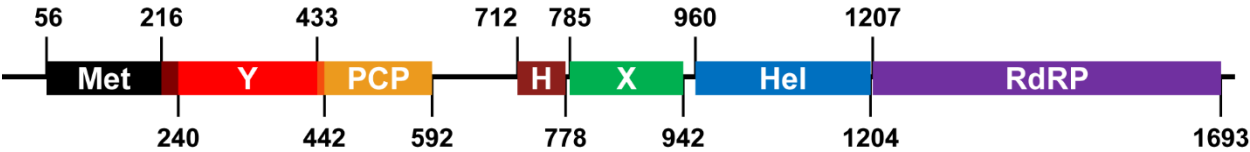
Forward		Reverse	
Residue Number	Primer	Residue Number	Primer
403	acctgtattccaaggatccgcgcagaagttattactcg	635	caatgataatgatgctcgagttaacccgggtgcaagactgcg
440	acctgtattccaaggatccgccggcttcacctggatcc	691	caatgataatgatgctcgagttaggcagattcccacacgtggc
455	acctgtattccaaggatccgctccatgtcattgccgtac	696	caatgataatgatgctcgagttagccacagaacgggttggcag
510	acctgtattccaaggatccgctgagtcgctatcagcga		

442

443 **Table 3:** Primers used to clone all HEV constructs.

Figure Legends

A



B

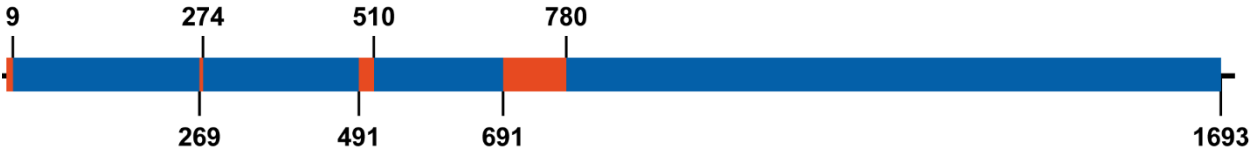


Figure 1: Schematic representations of the HEV ORF1 depicting the boundaries of the A) methyltransferase (Met), Y-domain (Y), putative cysteine protease (PCP), disordered hinge region (H), X-domain (X), helicase (Hel) and RNA dependent RNA polymerase (RdRP) as identified by Koonin *et al.* (18) and B) the predicted disordered (Orange) and secondary structure containing (Blue) regions.

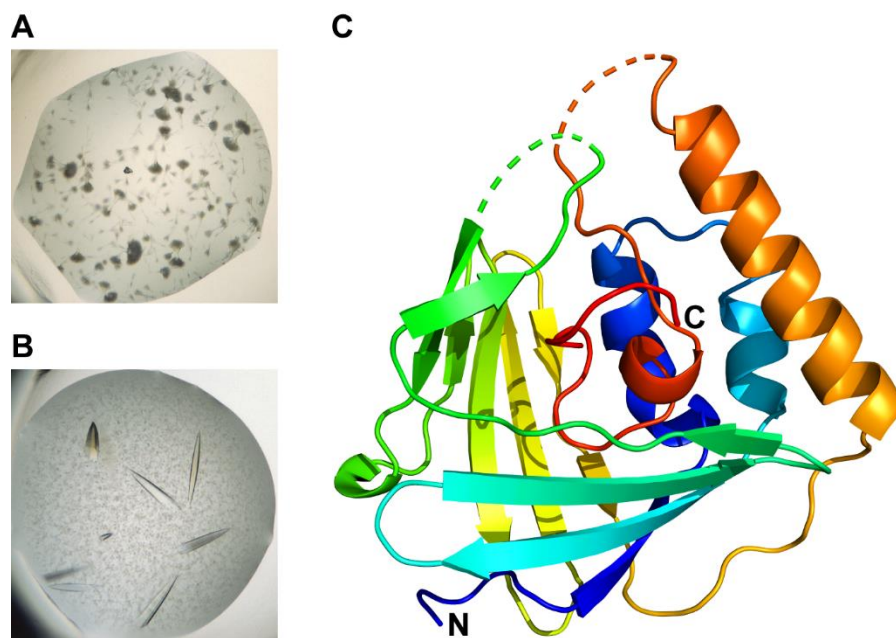


Figure 2: Crystallization and structure solution of HEV⁵¹⁰⁻⁶⁹¹. A) Initial crystal hits obtained in 18 % w/v PEG 3350, 180 mM LiNO₃ as precipitant and grown at 4°C. B) Crystals obtained following optimization of the initial screening hit with 10 mM NiCl₂. C) Structure of HEV⁵¹⁰⁻⁶⁹¹ shown in rainbow color representation with the N-terminus shown in blue and the C-terminus shown in red (PDB 6NU9).

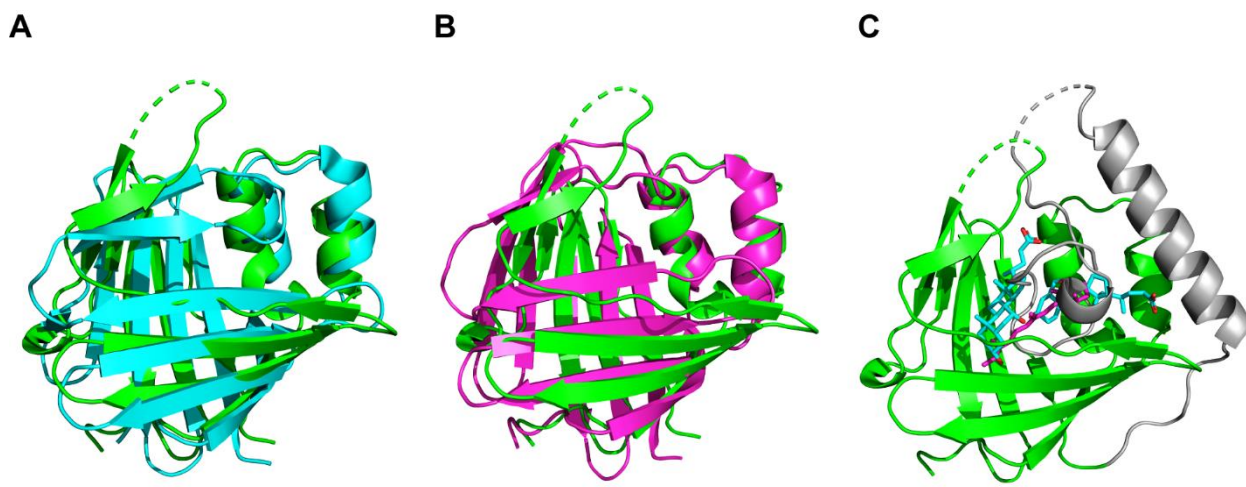


Figure 3: Superposition of HEV⁵¹⁰⁻⁶⁹¹ with the most significant structural homologs identified in DALI. Residues 514 – 635 of HEV⁵¹⁰⁻⁶⁹¹ with A) chicken liver fatty acid binding protein (49) (PDB 1TW4) Z-score 10.6 and B) retinol binding protein (PDB 5FEN) Z-score 11.2. C) Superposition of the bound cholic acid (cyan) present in 1TW4 and retinol (magenta) present in 5FEN with HEV⁵¹⁰⁻⁶⁹¹.

A

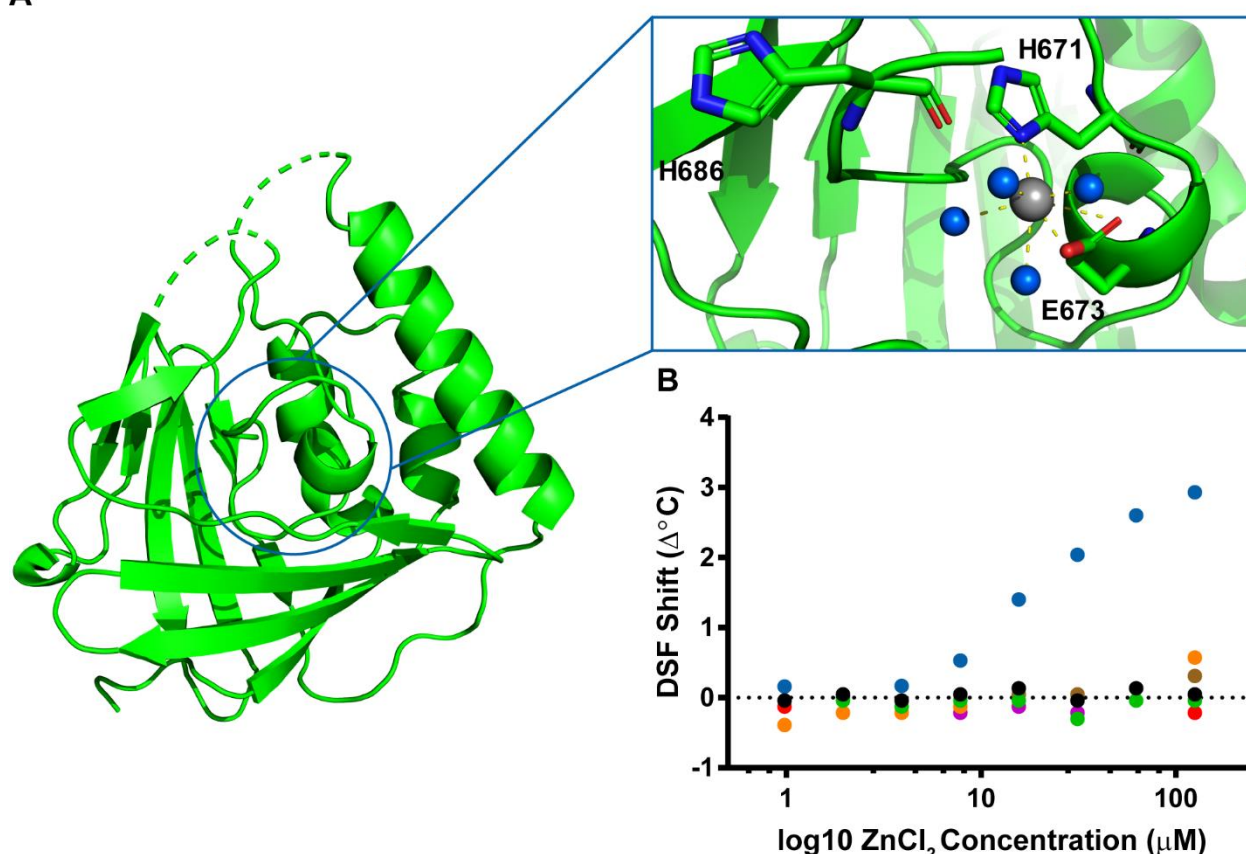


Figure 4: Analysis of metal interactions with HEV⁵¹⁰⁻⁶⁹¹. A) Magnification of region adjacent to the bound metal (grey) showing surrounding waters present (blue) and potential coordinating residues H671, E673 and H686. B) DSF analysis of wild type HEV⁵¹⁰⁻⁶⁹⁶ in the presence of

increasing concentrations of zinc (blue), calcium (green), cobalt (brown), iron (orange), manganese (red) and magnesium (magenta).

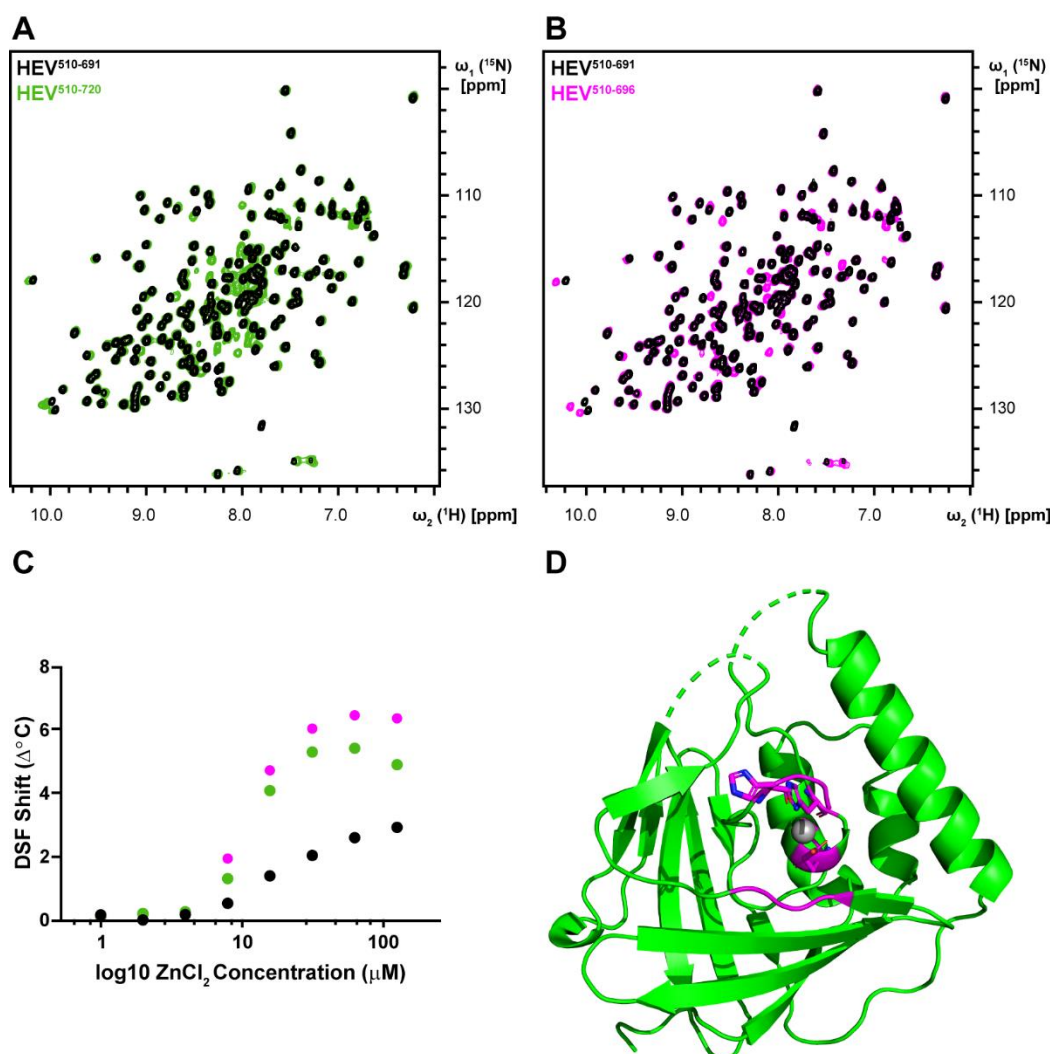
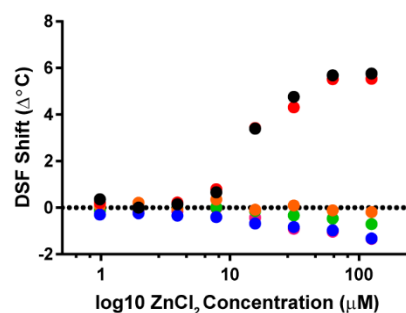


Figure 5: Optimization of the C-terminus of HEV⁵¹⁰⁻⁶⁹¹. Superposition of 2D [^{15}N , ^1H]-HSQC spectra acquired with HEV⁵¹⁰⁻⁶⁹¹ (black) and A) HEV⁵¹⁰⁻⁷²⁰ (green) and B) HEV⁵¹⁰⁻⁶⁹⁶ (magenta). C) DSF analysis of HEV⁵¹⁰⁻⁶⁹¹ (black), HEV⁵¹⁰⁻⁷²⁰ (green) and HEV⁵¹⁰⁻⁶⁹⁶ (magenta) in the presence of increasing concentrations of ZnCl₂. Structure of HEV⁵¹⁰⁻⁶⁹¹ showing all residues (magenta) that experience a significant chemical shift perturbation when compared to HEV⁵¹⁰⁻⁶⁹⁶.

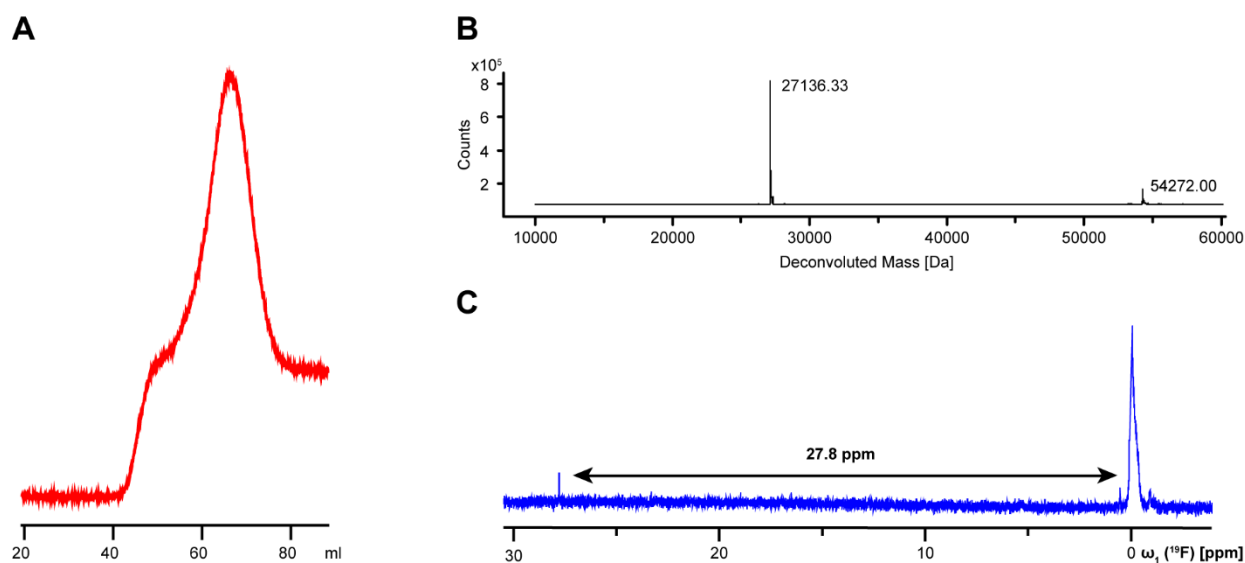
479



480

481 **Figure 6:** DSF analysis of HEV⁵¹⁰⁻⁶⁹⁶ (black), ΔE583A (red), ΔH671A (blue), ΔE673A (orange),
482 ΔH686A (green) and ΔH671A/ΔE673A/ΔH686A (magenta) HEV⁵¹⁰⁻⁶⁹⁶ in the presence of
483 increasing concentrations of ZnCl₂.

484



485

486 **Figure 7:** Purification and analysis of HEV⁴⁴⁰⁻⁶⁹¹. A) Elution profile measured at 280 nm of
487 HEV⁴⁴⁰⁻⁶⁹¹ from a HiLoad 16/60 Superdex 75 gel filtration column. B) Mass spectrometry analysis
488 of the purified HEV⁴⁴⁰⁻⁶⁹¹. C) ¹⁹F 1D analysis of metals co-purified with HEV⁴⁴⁰⁻⁶⁹¹, in complex
489 with 5F-BAPTA.

References

1. Nan Y, Zhang YJ. 2016. Molecular biology and infection of hepatitis E virus. *Front Microbiol* 7:1–21.
2. Abravanel F, Lhomme S, Dubois M, Peron J-M, Alric L, Kamar N, Izopet J. 2013. Hepatitis E virus. *Médecine Mal Infect* 43:263–270.
3. Navaneethan U, Al Mohajer M, Shata MT. 2008. Hepatitis E and pregnancy - Understanding the pathogenesis. *Liver Int* 28:1190–1199.
4. Cao D, Meng XJ. 2012. Molecular biology and replication of hepatitis E virus. *Emerg Microbes Infect* 1:1–10.
5. Mushahwar IK. 2008. Hepatitis E virus: Molecular virology, clinical features, diagnosis, transmission, epidemiology, and prevention. *J Med Virol* 80:646–658.
6. Minuk GY, Sun A, Sun DF, Uhanova J, Nicolle LE, Larke B, Giulivi A. 2007. Serological evidence of hepatitis E virus infection in an indigenous North American population. *Can J Gastroenterol* 21:439–442.
7. Bendall R, Ellis V, Ijaz S, Thurairajah P, Dalton HR. 2008. Serological response to hepatitis E virus genotype 3 infection: IgG quantitation, avidity, and IgM response. *J Med Virol* 80:95–101.
8. Kamar N, Rostaing L, Abravanel F, Garrouste C, Esposito L, Cardeau- Desangles I, Mansuy JM, Selves J, Peron JM, Otal P, Muscari F, Izopet J. 2010. Pegylated interferon- α for treating chronic hepatitis E virus infection after liver transplantation. *Clin Infect Dis* 50:e30–e33.

- 511 9. Kamar N, Rostaing L, Abravanel F, Garrouste C, Lhomme S, Esposito L, Basse G,
512 Cointault O, Ribes D, Nogier MB, Alric L, Peron JM, Izopet J. 2010. Ribavirin therapy
513 inhibits viral replication on patients with chronic hepatitis E virus infection.
514 Gastroenterology 139:1612–1618.
- 515 10. Smith DB, Simmonds P, Jameel S, Emerson SU, Harrison TJ, Meng XJ, Okamoto H, Van
516 der Poel WHM, Purdy MA. 2014. Consensus proposals for classification of the family
517 Hepeviridae. J Gen Virol 95:2223–2232.
- 518 11. Ahmad I, Holla RP, Jameel S. 2011. Molecular virology of hepatitis E virus. Virus Res
519 161:47–58.
- 520 12. Tam AW, Smith MM, Guerra ME, Huang CC, Bradley DW, Fry KE, Reyes GR. 1991.
521 Hepatitis E virus (HEV): Molecular cloning and sequencing of the full-length viral
522 genome. Virology 185:120–131.
- 523 13. Tsarev SA, Emerson SU, Reyes GR, Tsareva TS, Legters LJ, Malik IA, Iqbal M, Purcell
524 RH, Legters LJ, Li IAM, Li MI, Purcell RH. 1992. Characterization of a prototype strain of
525 hepatitis E virus. Proc Natl Acad Sci U S A 89:559–563.
- 526 14. Ding Q, Heller B, Capuccino JM V., Song B, Nimgaonkar I, Hrebikova G, Contreras JE,
527 Ploss A. 2017. Hepatitis E virus ORF3 is a functional ion channel required for release of
528 infectious particles. Proc Natl Acad Sci 114:1147–1152.
- 529 15. Cao D, Huang Y-W, Meng X-J. 2010. The Nucleotides on the Stem-Loop RNA Structure
530 in the Junction Region of the Hepatitis E Virus Genome Are Critical for Virus
531 Replication. J Virol 84:13040–13044.

- 532 16. Emerson SU, Zhang M, Meng X-J, Nguyen H, St. Claire M, Govindarajan S, Huang YK,
533 Purcell RH. 2001. Recombinant hepatitis E virus genomes infectious for primates:
534 Importance of capping and discovery of a cis-reactive element. *Proc Natl Acad Sci*
535 98:15270–15275.
- 536 17. Parvez MK. 2017. The hepatitis E virus nonstructural polyprotein. *Future Microbiol*
537 12:915–924.
- 538 18. Koonin E V., Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW. 1992.
539 Computer-assisted assignment of functional domains in the nonstructural polyprotein of
540 hepatitis E virus: Delineation of an additional group of positive-strand RNA plant and
541 animal viruses. *Proc Natl Acad Sci* 89:8259–8263.
- 542 19. Magden J, Takeda N, Li T, Auvinen P, Ahola T, Miyamura T, Merits A, Kaariainen L.
543 2001. Virus-specific mRNA capping enzyme encoded by hepatitis E virus. *J Virol*
544 75:6249–6255.
- 545 20. Agrawal S, Gupta D, Panda SK. 2001. The 3' end of hepatitis E virus (HEV) genome
546 binds specifically to the viral RNA-dependent RNA polymerase (RdRp). *Virology*
547 282:87–101.
- 548 21. Karpe YA, Lole KS. 2010. NTPase and 5' to 3' RNA Duplex-Unwinding Activities of the
549 Hepatitis E Virus Helicase Domain. *J Virol* 84:3595–3602.
- 550 22. Ansari IH, Nanda SK, Durgapal H, Agrawal S, Mohanty SK, Gupta D, Jameel S, Panda
551 SK. 2000. Cloning, sequencing, and expression of the hepatitis E virus (HEV)
552 nonstructural open reading frame 1 (ORF1). *J Med Virol* 60:275–283.

- 553 23. Ropp SL, Tam AW, Beames B, Purdy M, Frey TK. 2000. Expression of the hepatitis E
554 virus ORF1. *Arch Virol* 145:1321–1337.
- 555 24. Sehgal D, Thomas S, Chakraborty M, Jameel S. 2006. Expression and processing of the
556 hepatitis E virus ORF1 nonstructural polyprotein. *Virol J* 3:1–9.
- 557 25. Paliwal D, Panda SK, Kapur N, Varma SPK, Durgapal H. 2014. Hepatitis E virus (HEV)
558 protease: A chymotrypsin-like enzyme that processes both non-structural (pORF1) and
559 capsid (pORF2) protein. *J Gen Virol* 95:1689–1700.
- 560 26. Parvez MK, Khan AA. 2014. Molecular modeling and analysis of hepatitis E virus (HEV)
561 papain-like cysteine protease. *Virus Res* 179:220–224.
- 562 27. Parvez MK. 2013. Molecular characterization of hepatitis E virus ORF1 gene supports a
563 papain-like cysteine protease (PCP)-domain activity. *Virus Res* 178:553–556.
- 564 28. Holm L, Laakso LM. 2016. Dali server update. *Nucleic Acids Res* 44:W351–W355.
- 565 29. Parvez MK. 2015. The hepatitis E virus ORF1 “X-domain” residues form a putative
566 macrodomain protein/Appr-1”-pase catalytic-site, critical for viral RNA replication. *Gene*
567 566:47–53.
- 568 30. Benters J, Flögel U, Schäfer T, Leibfritz D, Hechtenberg S, Beyersmann D. 1997. Study
569 of the interactions of cadmium and zinc ions with cellular calcium homoeostasis using
570 ¹⁹F-NMR spectroscopy. *Biochem J* 322 (Pt 3:793–9.
- 571 31. McCall KA, Huang C, Fierke CA. 2000. Function and mechanism of zinc
572 metalloenzymes. *J Nutr* 130:1437S–1446S.
- 573 32. Ryan JM, Heneghan MA. 2014. Pregnancy and the liver. *Clin Liver Dis* 4:51–54.

- 574 33. Suppiah S, Zhou Y, Frey TK. 2011. Lack of processing of the expressed ORF1 gene
575 product of hepatitis E virus. *Virol J* 8:245.
- 576 34. Perttila J, Spuul P, Ahola T. 2013. Early secretory pathway localization and lack of
577 processing for hepatitis E virus replication protein pORF1. *J Gen Virol* 94:807–816.
- 578 35. Emerson SU, Nguyen H, Graff J, Stephany D a., Brockington A, Purcell RH. 2004. In
579 vitro replication of hepatitis E virus (HEV) genomes and of an HEV replicon expressing
580 green fluorescent protein. *J Virol* 78:4838–4846.
- 581 36. Graff J, Torian U, Nguyen H, Emerson SU. 2006. A bicistronic subgenomic mRNA
582 encodes both the ORF2 and ORF3 proteins of hepatitis E virus. *J Virol* 80:5919–5926.
- 583 37. Shukla P, Nguyen HT, Torian U, Engle RE, Faulk K, Dalton HR, Bendall RP, Keane FE,
584 Purcell RH, Emerson SU. 2011. Cross-species infections of cultured cells by hepatitis E
585 virus and discovery of an infectious virus-host recombinant. *Proc Natl Acad Sci*
586 108:2438–2443.
- 587 38. Brea RJ, Cole CM, Lyda BR, Ye L, Prosser RS, Sunahara RK, Devaraj NK. 2017.
588 Hepatitis B virus protein has unusual iron-sulfur cluster. *C&EN Glob Enterp* 95:11–11.
- 589 39. Terwilliger T. 2004. SOLVE and RESOLVE: Automated structure solution, density
590 modification, and model building. *J Synchrotron Radiat* 11:49–52.
- 591 40. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot.
592 *Acta Crystallogr Sect D Biol Crystallogr* 66:486–501.
- 593 41. Bricogne G., Blanc E., Brandl M., Flensburg C., Keller P. PW, Roversi P, Sharff A.,
594 Smart O.S., Vonnrhein C. WTO. 2016. BUSTER. 2.11.6. Cambridge, United Kingdom:

Global Phasing Ltd.

42. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66:213–221.
43. Terwilliger TC, Grosse-Kunstleve RW, Afonine P V., Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD. 2007. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr Sect D Biol Crystallogr* 64:61–69.
44. Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY. 2005. High-resolution four-dimensional ¹H-¹³C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. *J Am Chem Soc* 127:2767–2775.
45. Hyberts SG, Milbradt AG, Wagner AB, Arthanari H, Wagner G. 2012. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J Biomol NMR* 52:315–327.
46. Hyberts SG, Takeuchi K, Wagner G. 2010. Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data. *J Am Chem Soc* 132:2145–2147.
47. Keller R. 2004. The computer aided resonance assignment tutorial Goldau, Switzerland: Cantina Verlag. Cantina Verlag, Goldau, Switzerland.

- 616 48. Schrödinger, LLC. 2015. The {PyMOL} Molecular Graphics System, Version~1.8.
- 617 49. Nicesola D, Perduca M, Capaldi S, Carrizo ME, Righetti PG, Monaco HL. 2004. Crystal
618 structure of chicken liver basic fatty acid-binding protein complexed with cholic acid.
619 Biochemistry 43:14072–14079.