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

Andrew Proudfoot,^a Anastasia Hyrina,^b Meghan Holdorf,^b Andreas O. Frank,^a Dirksen Bussiere^a

^aStructural and Biophysical Chemistry Group, Global Discovery Chemistry, Novartis Institutes for BioMedical Research, Emeryville, California, USA

^bInfectious Disease Group, Novartis Institutes for BioMedical Research, Emeryville, California, USA

ABSTRACT Hepatitis E virus (HEV) is a 7.2-kb positive-sense, single-stranded RNA virus containing three partially overlapping reading frames, ORF1 to ORF3. 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 nonclassical, potentially catalytic zinc-binding motif. Based on the structural homology of the HEV protein with known structures, along with the presence of a catalytic zinc-binding motif, it is possible that the identified protein corresponds to the HEV protease, which could require activation or repression through the binding of a fatty acid. This represents a significant step forward in the characterization and the understanding of the molecular mechanisms of the HEV genome. We present analysis for the first time of this identified nonstructural 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; it causes an estimated 20 million infections, which result in approximately 57,000 deaths a year. Although it is known that the nonstructural proteins of 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. Here we present the first structural and biophysical characterization of an HEV nonstructural protein using a construct that has partially overlapping boundaries with the predicted putative cysteine protease.

KEYWORDS biophysics, crystal structure, fatty-acid binding, nuclear magnetic resonance, protease

Hepatitis E virus (HEV) is a nonenveloped, positive-sense 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 semitropical 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 becoming more common (5–7). Although currently there is no accepted treatment for HEV infection, off-label treatments with both interferon and ribavirin have been

Citation Proudfoot A, Hyrina A, Holdorf M, Frank AO, Bussiere D. 2019. First crystal structure of a nonstructural hepatitis E viral protein identifies a putative novel zinc-binding protein. *J Virol* 93:e00170-19. <https://doi.org/10.1128/JVI.00170-19>.

Editor J.-H. James Ou, University of Southern California

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Dirksen Bussiere, dirksen.bussiere@novartis.com.

Received 31 January 2019

Accepted 17 April 2019

Accepted manuscript posted online 24 April 2019

Published 14 June 2019

FIG 1 Schematic representations of the HEV ORF1 depicting the boundaries of the 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) (A) and the predicted disordered (orange) and secondary-structure-containing (blue) regions (B).

used successfully to treat chronic HEV infection (8, 9). As HEV is a growing worldwide threat, better understanding of the viral molecular mechanisms is 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 (ORFs) flanked by 5' and 3' untranslated/noncoding regions (11–13). ORF1 encodes all the nonstructural proteins required for viral replication. ORF2 is the only ORF that has been structurally characterized in the HEV genome; it encodes the viral capsid protein that is involved with virion assembly (13). ORF3, the smallest of the three ORFs, encodes a 113- to 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 1,693-residue polyprotein (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) (Fig. 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 modeling of other alphaviruses homologous to HEV which contain a cysteine protease predicts a putative cysteine protease between residues 433 and 592 in ORF1 (18).

Previous research has been unable to confirm the proteolytic processing of ORF1 or the presence of a protease. Some reports show that ORF1 is expressed as a single 186-kDa polyprotein in *Escherichia coli*, in insect and mammalian cells, and in a cell-free system (22). Other reports support polyprotein 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 resulted in the production of eight smaller fragments. Although the authors were unable to identify if these fragments corresponded to the individual expected protein domains, they were able to show that their production was inhibited by a cell-permeable cysteine protease inhibitor (24). Current research on the purified putative protease domain *in vitro* has been limited. Protein predicted to correspond to the protease has been purified under denaturing

conditions and was shown upon refolding to have proteolytic activity against ORF1 and ORF2 products. It was also shown that the mutation of highly conserved cysteine and histidine residues in this region appear to inhibit proteolytic activity (25). A three-dimensional model of the predicted protease region has been produced using the p150 rubella protease (26). Although the sequence homology between the predicted HEV protease and the p150 rubella protease is less than 20%, the authors used consensus prediction results and the alignment of conserved cysteine and histidine residues between the two proteins to produce the model. However, no characterization of the predicted protease domain was performed, which is required in order to understand better the viral biology. Here we report a comprehensive analysis of the biochemical, biophysical, and structural parameters of the predicted protease domain that may help elucidate the function of the protein in the HEV life cycle.

(This article was submitted to an online preprint archive [27].)

RESULTS

All previous work performed on the HEV nonstructural proteins of ORF1 has been done based on the boundaries originally proposed by Koonin et al. (18) (Fig. 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 the online protein fold recognition server PHYRE2 (28). ORF1 was analyzed in its entirety using the "Normal" modeling mode, and using the regular secondary structure and disorder predictions, regions of order (containing regular secondary structure) and disorder were predicted within the open reading frame (Fig. 1B). The C-terminal region of ORF1 contains conserved motifs characteristic of a superfamily 1 helicase and the eight conserved motifs indicative of a positive-stranded RNA viral RNA polymerase (18), so it was not surprising that regions of secondary structure are predicted between residues 780 to 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 to 8, 270 to 273, 492 to 509, and 692 to 779. The region of predicted disorder between residues 692 and 779 aligns with the region previously identified to be the disordered hinge region (residues 712 to 778), with the three remaining regions of disorder bordering three regions of predicted secondary structure. The two regions of order identified between residues 9 and 269 and residues 274 and 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 and 691. Prior analysis of this region had shown that residues 433 to 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 are required to fully understand any potential HEV protease (25, 26, 29). To confirm the presence of a domain between residues 510 and 691, a construct containing this region was codon optimized (Fig. 2), cloned, and expressed in *E. coli* as a maltose binding protein (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 one-dimensional nuclear magnetic resonance (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 (Fig. 3A). Optimized crystals grew at 4°C in precipitant consisting of 18% (wt/vol) polyethylene glycol 3350

```

403      408      413      418      423
A   Q   K   F   I   T   R   L   Y   S   W   L   F   E   K   S   G   R   D   Y   I   P   G   R   Q
gcg cag aag ttt att aca cgt ctt tat agc tgg ctg ttt gaa aaa tcc ggc cgc gat tac att cca ggc cgt cag

428      433      438      443      448
L   E   F   Y   A   Q   C   R   R   W   L   S   A   G   F   H   L   D   P   R   V   L   V   F   D
ctt gag ttc tac gct cag tgc cgt cgc tgg tta agc gcc ggc ttt cac tta gat ccg cgt gta tta gtt ttc gac

453      458      463      468      473
E   S   A   P   C   H   C   R   T   A   I   R   K   A   V   S   K   F   C   C   F   M   K   W   L
gag agc gct cca tgt cat tgc cgt acc gcc att cgc aaa gca gta tcc aag ttt tgc tgc ttc atg aag tgg ctt

478      483      488      493      498
G   Q   E   C   T   C   F   L   Q   P   A   E   G   V   V   G   D   Q   G   H   D   N   E   A   Y
ggc cag gaa tgt acc tgt ttt ttg caa cct gca gaa ggt gtt gtg ggc gat cag ggt cat gat aat gaa gcc tat

503      508      513      518      523
E   G   S   D   V   D   P   A   E   S   A   I   S   D   I   S   G   S   Y   V   V   P   G   T   A
gaa ggc tca gac gtg gac cct gct gag tcc gct atc tcg gac att agt ggg tca tat gtc gtc cct ggg aca gca

528      533      538      543      548
L   Q   P   L   Y   Q   A   L   D   L   P   A   E   I   V   A   R   A   G   R   L   T   A   T   V
ctt caa cca ctt tac cag gct ttg gat ctt ccg gct gag atc gta gct cgc gcc ggc cgc ctt acc gcc acc gta

553      558      563      568      573
K   V   S   Q   V   D   G   R   I   D   C   E   T   L   L   G   N   K   T   F   R   T   S   F   V
aaa gta tcg caa gtg gac ggc cgt att gac tgt gag acc ctg tta ggt aac aag aca ttc cgt acg tcg ttt gta

578      583      588      593      598
D   G   A   V   L   E   T   N   G   P   E   R   H   N   L   S   F   D   A   S   Q   S   T   M   A
gat ggt gcg gta ctt gaa aca aac ggg ccg gaa cgc cat aac ctt agt ttc gat gcg tcc cag tct acc atg gcc

603      608      613      618      623
A   G   P   F   S   L   T   Y   A   A   S   A   A   G   L   E   V   R   Y   V   A   A   G   L   D
gcg ggt ccg ttt agc ctt acg tac gct gct tcc gct gcc ggt tta gag gta cgc tat gtt gca gct ggc ctt gat

628      633      638      643      648
H   R   A   V   F   A   P   G   V   S   P   R   S   A   P   G   E   V   T   A   F   C   S   A   L
cat cgc gca gtc ttc gca ccg ggt gtc tcc ccg cgc tca gcg cct ggc gaa gtg act gct ttc tgt tcg gcc ttg

653      658      663      668      673
Y   R   F   N   R   E   A   Q   R   L   S   L   T   G   N   F   W   F   H   P   E   G   L   L   G
tat cgc ttc aac cgt gag gct cag cgc ctg tcg ctt acc ggt aac ttc tgg ttc cat cct gaa ggg ctt ctt ggt

678      683      688      693
P   F   A   P   F   S   P   G   H   V   W   E   S   A   N   P   F   C   G
ccg ttc gcg cca ttt agt cca ggg cac gtg tgg gaa tcg gcc aac ccg ttc tgt ggc

```

FIG 2 Codon-optimized sequence of HEV ORF1 for residues 403 to 696.

(PEG 3350), 180 mM LiNO₃, and 10 mM NiCl₂ (Fig. 3B) and diffracted to a 1.8-Å resolution on an in-house X-ray source. An amino acid sequence search of the entries in the RCSB Protein Data Bank (PDB) showed that HEV⁵¹⁰⁻⁶⁹¹ did not have sequence homology to any deposited structure, so molecular replacement was considered 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 10 β -strands and 4 α -helices, with β -strands 1 to 4 and 5 to 10 arranged in two antiparallel sheets arranged in a sandwich-like fold. α -Helices 1 and 2 are located between β -strands 1 and 2, and α -helices 3 and 4 are positioned at the C terminus of the protein, with α -helix 4 sitting between the two antiparallel β -sheets (Fig. 3C).

The crystal structure was analyzed using DALI to identify if there were any structurally homologous proteins in the PDB (30). The structural alignment showed that although there was only 13 to 15% amino acid similarity between the identified structural homologs, residues 514 to 635 of HEV⁵¹⁰⁻⁶⁹¹ have significant structural homology to multiple fatty acid binding domains, with the homologous regions aligning with a backbone RMSD of ~2.5 Å (Fig. 4A and B). This homologous region of

FIG 3 Crystallization and structure solution of HEV⁵¹⁰⁻⁶⁹¹. (A) Initial crystal hits obtained with 18% (wt/vol) PEG 3350 and 180 mM LiNO₃ as precipitant and growth 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 code 6NU9).

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 (Fig. 4C). An additional HEV construct from residues 510 to 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 coordination site. However, in our structure, the side chain of His 686 adopted a rotamer that did not facilitate coordination of the bound metal (Fig. 5A). Native HEV⁵¹⁰⁻⁶⁹¹ crystals soaked in a 1 mM solution of EDTA showed signs of cracking following 1 h of soaking. However, data sets could still be collected with the crystals, and the processed data showed no electron density in the regions where we had previously identified the bound metal. This confirmed that incubation with EDTA had resulted in the removal of the bound metal ion and that the original density identified in the apo structure corresponded to a metal that could 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 either that there were 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 if HEV⁵¹⁰⁻⁶⁹¹ had an affinity for a specific metal, differential scanning fluorimetry (DSF) was used to screen a number of different metals to see if any induced an increase in thermal stability. Of the 17 metals tested (barium, cadmium, calcium, cesium, chromium, cobalt, copper, iron, lithium, magnesium, manganese, nickel, potassium, praseodymium, strontium, yttrium, and zinc), zinc was the only metal that stabilized the protein and induced an increase in thermal stability of 3.0°C (Fig. 5B). Data sets acquired with HEV⁵¹⁰⁻⁶⁹¹ crystals soaked in a 1 mM solution of zinc chloride once again showed density corresponding to a coordinated metal ion, which could be fully satisfied with a zinc ion during data reduction. However, as we had previously observed, the His 686 side chain still adopted a rotamer that could not act as the final zinc coordination site.

TABLE 1 Data collection and refinement statistics of the HEV⁵¹⁰⁻⁶⁹¹ X-ray structure

Parameter	Value for:		
	HEV ⁵¹⁰⁻⁶⁹¹ native	HEV ⁵¹⁰⁻⁶⁹¹ thimerosal derivative	HEV ⁵¹⁰⁻⁶⁹¹ zinc costructure (PDB code 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	P6522	P6522	P6522
Cell dimensions			
a, b, c (Å)	64.49, 64.49, 157.41	64.20, 64.20, 158.70	64.40, 64.40, 158.90
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 90.00, 120.00
Total no. of reflections	891,654 (69,081)	378,180 (29,740)	680,486 (19,154)
No. of unique reflections	14,844 (1,430)	21,378 (2,114)	20,137 (1,960)
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)
Rmerge	0.068 (0.218)	0.101 (0.333)	0.228 (1.028)
Rpim	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
No. of non-H atoms			
Protein			1,306
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
Avg B-factor (Å ²)			17.51

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., polygly-

FIG 4 Superposition of HEV⁵¹⁰⁻⁶⁹¹ with the most significant structural homologs identified in DALI. (A and B) Residues 514 to 635 of HEV⁵¹⁰⁻⁶⁹¹ with chicken liver fatty acid binding protein (50) (PDB code 1TW4) (Z-score, 10.6) (A) and retinol binding protein (PDB code 5FEN) (Z-score, 11.2) (B). (C) Superposition of the bound cholic acid (cyan) present in the sequence corresponding to PDB code 1TW4 and retinol (magenta) present in the sequence corresponding to PDB code 5FEN with HEV⁵¹⁰⁻⁶⁹¹. Additional residues present in the HEV protein (636 to 691) are colored white.

FIG 5 Analysis of metal interactions with HEV⁵¹⁰⁻⁶⁹¹. (A) Magnification of region adjacent to the bound metal (gray) 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).

cine, -alanine, or -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 to 720 (HEV⁵¹⁰⁻⁷²⁰) was produced that was 29 residues longer than HEV⁵¹⁰⁻⁶⁹¹ and was analyzed by NMR. Comparison of HEV⁵¹⁰⁻⁷²⁰ with HEV⁵¹⁰⁻⁶⁹¹ showed that the additional residues had amide chemical shifts in the region of 7.5 to 8.5 ppm, which is indicative of disordered amino acids (Fig. 6A) and coincided with the performed secondary-structure prediction (Fig. 1B). Analysis by DSF showed that in the presence of zinc, HEV⁵¹⁰⁻⁷²⁰ had a greater increase in thermal stability (5.5°C) than did HEV⁵¹⁰⁻⁶⁹¹ (Fig. 6C). 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 to 696 (HEV⁵¹⁰⁻⁶⁹⁶) was identified to be the optimal construct and exhibited a 6°C increase in thermal stability in the presence of zinc (Fig. 6C) but did not contain an additional increase in disorder in the NMR spectrum (Fig. 6B). 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] heteronuclear single quantum coherence (HSQC) spectra of HEV⁵¹⁰⁻⁶⁹¹ and HEV⁵¹⁰⁻⁶⁹⁶, we found that 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 coordination residues, the residues at the C terminus and a stretch of amino acids on β -strand 4, between residues 572 and 574 (Fig. 6D). Although we were unable to crystallize HEV⁵¹⁰⁻⁶⁹⁶, [¹H-¹⁵N] HSQC-NMR experiments provided valuable data to gain insight in the structure of the longer construct. In general, chemical shift perturbations (CSPs) of NMR signals occur due to a change in the local magnetic environment of the affected resonances. The changes we observed for residues 572 to 574 are therefore indicative of the additional five residues (692 to 696) interacting with the three amino acids. Although irrefutable

FIG 6 Optimization of the C terminus of HEV⁵¹⁰⁻⁶⁹¹. (A and B) Superposition of 2D [¹⁵N, ¹H]-HSQC spectra acquired with HEV⁵¹⁰⁻⁶⁹¹ (black) and HEV⁵¹⁰⁻⁷²⁰ (green) (A) and HEV⁵¹⁰⁻⁶⁹⁶ (magenta) (B). (C) DSF analysis of HEV⁵¹⁰⁻⁶⁹¹ (black), HEV⁵¹⁰⁻⁷²⁰ (green), and HEV⁵¹⁰⁻⁶⁹⁶ (magenta) in the presence of increasing concentrations of ZnCl₂. (D) Structure of HEV⁵¹⁰⁻⁶⁹¹ showing all residues (magenta) that experience a significant chemical shift perturbation compared to HEV⁵¹⁰⁻⁶⁹⁶.

structural information cannot be obtained from this type of NMR experiment, we assume from analysis of the solved crystal structure that the five additional residues could interact with residues 572 to 574 by forming a β -strand that runs 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 coordination site. To confirm if His 671, Glu 673, and His 686 were involved with zinc coordination, each of the three residues was mutated individually and together to an alanine residue. Mutation of each residue contributed approximately a 2°C reduction in thermal stability compared to that of HEV⁵¹⁰⁻⁶⁹⁶, corresponding to 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 (Fig. 7), indicating that all three residues are involved with zinc coordination. If the identified domain were to correspond to a metalloprotease, an acidic residue would need to be in close proximity to the coordinated zinc, to act as a catalytic residue. We identified Glu 583 as the residue that potentially plays this role, and mutagenesis of Glu 583 to alanine did not have any significant effect on the induced increase in thermal stability upon addition of zinc (Fig. 5A). This was to be expected since this residue is not directly involved with zinc coordination and thus does not directly stabilize the proteins' tertiary structure.

Work characterizing the predicted HEV protease has been performed with constructs that more closely matched the boundaries identified by Koonin et al. (18, 25, 31)

FIG 7 DSF analysis of HEV⁵¹⁰⁻⁶⁹⁶ (black) and ΔE583A (red), ΔH671A (blue), ΔE673A (orange), ΔH686A (green), and ΔH671A/ΔE673A/ΔH686A (magenta) HEV⁵¹⁰⁻⁶⁹⁶ in the presence of increasing concentrations of ZnCl₂.

(Fig. 1A). To investigate the effect of additional residues at the N terminus of HEV⁵¹⁰⁻⁶⁹¹, constructs were produced that extended the N-terminal boundaries to residues 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 showed that *E. coli* BL21 cells expressing residues 440 to 610 experienced cell death upon induction at different temperatures and various concentrations of isopropyl-β-D-1-thiogalactopyranoside (IPTG) (25). Consistent with these observations, when the 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. In the presence of 1× trace metals (Table 2), no expression was observed with the HEV⁴⁰³⁻⁶⁹¹ construct; however, HEV⁴⁴⁰⁻⁶⁹¹, which coincided with the previously identified putative protease domain (18), had protein expression comparable to that of HEV⁵¹⁰⁻⁶⁹¹. Following gel filtration of HEV⁴⁴⁰⁻⁶⁹¹, it was apparent that the purified protein was eluted from the gel filtration column at a much lower retention volume (Fig. 8A) than expected (at a retention volume corresponding to a protein with an *M_w* of 270 kDa) and was brown. Mass spectrometry was performed with the protein species under denaturing conditions, and a single peak at 27,136.33 Da was observed (Fig. 8B), 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 determine if this was the case, the protein was thermally denatured and spun down and the supernatant was analyzed with 1,2-bis-(2-amino-5-fluorophenoxy)ethane-*N,N,N',N'*-tetraacetic acid (5F-BAPTA). 5F-BAPTA is a divalent cation chelator with two fluorine atoms positioned at symmetric positions, which yields a

TABLE 2 Composition of trace metal solution^a

Vol	Component	1× concn
50 ml	0.1 M FeCl ₃ in 0.12 M HCl	50 μM Fe
2 ml	1 M CaCl ₂	20 μM Ca
1 ml	1 M MnCl ₂ ·4H ₂ O	10 μM Mn
1 ml	1 M ZnSO ₄ ·7H ₂ O	10 μM Zn
1 ml	0.2 M CoCl ₂ ·6H ₂ O	2 μM Co
2 ml	0.1 M CuCl ₂ ·2H ₂ O	2 μM Cu
1 ml	0.2 M NiCl ₂ ·6H ₂ O	2 μM Ni
2 ml	0.1 M Na ₂ MoO ₄ ·5H ₂ O	2 μM Mo
2 ml	0.1 M Na ₂ SeO ₃ ·5H ₂ O	2 μM Se
2 ml	0.1 M H ₃ BO ₃	2 μM B

^aFor a 1,000× trace metal mixture, the components shown are added to 36 ml of water. All stock solutions of the individual metals should be autoclaved, except for the 0.1 M FeCl₃ in 1/100 volume concentrated HCl. Trace metal solution should be stored at room temperature.

FIG 8 Purification and analysis of HEV⁴⁴⁰⁻⁶⁹¹. (A) Elution profile measured at 280 nm of HEV⁴⁴⁰⁻⁶⁹¹ from a HiLoad 16/60 Superdex 75 gel filtration column. (B) Mass spectrometry analysis of the purified HEV⁴⁴⁰⁻⁶⁹¹. (C) ¹⁹F 1D analysis of metals copurified with HEV⁴⁴⁰⁻⁶⁹¹, in complex with 5F-BAPTA.

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. Although a strong brown color was observed in the protein pellet following denaturation and centrifugation, the metal that remained in solution induced a 27.8-ppm chemical shift perturbation to the 5F-BAPTA signal (Fig. 8C). According to a previous benchmark study (32), a chemical shift of 28.1 ppm is indicative of the formation of a 5F-BAPTA–Fe²⁺ complex. As none of the other known 5F-BAPTA–metal complexes show ¹⁹F chemical shifts at around 28 ppm, we concluded that the observed brown hue comes from HEV-bound, divalent iron.

DISCUSSION

Detailed characterization of the nonstructural 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 to 691) sandwiched between two regions of disorder (residues 492 to 509 and residues 692 to 779). Overexpression of this region resulted in a soluble, stable protein, suitable for crystallization. The resulting protein structure of HEV⁵¹⁰⁻⁶⁹¹ displayed a high structural homology to fatty acid binding domains; comparison 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 (Fig. 4C). 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 (Fig. 7). Each single mutant had a similar thermal melting point (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 (33). 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 acidic residues in close structural proximity to the bound zinc identified the highly conserved Glu 583 residue, which is positioned between β -strands 5 and 6 and could potentially serve as the catalytic residue (Fig. 5A). 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, reorientating 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 the binding of a specific fatty acid to the protease could be used by the virus as a regulatory mechanism.

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 revealing several distinct cleavage products (22–25, 29, 34, 35). 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 on the well-studied pSK-HEV2 construct (GenBank accession number AF444002.1) (16, 36). 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 (36), were synthesized with a NanoLuc-PEST reporter positioned in frame with ORF2 and sequence verified. Studies with the pSK-HEV2 replicon typically require the use of the Huh7-S10-3 liver cell line (37), to which we were unable to secure access, 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 to 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. Addition of 70 amino acids to the N terminus (HEV⁴⁴⁰⁻⁶⁹¹) of 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 due to the presence of bound iron. The 70 additional N-terminal residues added contain 6 cysteine residues, some of which

TABLE 3 Primers used to clone all HEV constructs

Forward		Reverse	
Residue no.	Primer	Residue no.	Primer
403	ACCTGTATTCCAAGGATCCGCGCAGAAGTTTATTACTCG	635	CAATGATAATGATGCTCGAGTTAACCCGGTGCGAAGACTGCGC
440	ACCTGTATTCCAAGGATCCGCGCGCTTTCACCTGGATCC	691	CAATGATAATGATGCTCGAGTTAGGCAGATTCCCACACGTGGC
510	ACCTGTATTCCAAGGATCCGCTGAGTCCGCTATCAGCGA	696	CAATGATAATGATGCTCGAGTTAGCCACAGAACGGGTTGGCAG

had previously been implicated as potential catalytic residues for the HEV protease (25). Contrary to this earlier report, we would like to propose that these cysteine residues are involved in the formation of an iron-sulfur (Fe-S) cluster. Analysis of the HEV ORF1 sequence for possible Fe-S motifs only identified approximate motifs in the HEV sequence. However, this does not mean that the CXCX₁₁CCX₈CXC sequence or any subset of this found between residues 457 and 483 does not correspond to a potential Fe-S motif. Although many Fe-S motifs have already been discovered and documented, new motifs are being discovered on a daily basis (38). This possible presence of an Fe-S cluster is also consistent with observations of Fe-S clusters in other hepatitis viruses (39).

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 nonstructural protein between the predicted Y domain and hypervariable 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 are ambiguous about the presence of a protease in ORF1, the results in this article confirm the presence of a nonstructural protein in the HEV ORF1 between residues 510 and 691. Having characterized the domain both structurally and biophysically, it is clear that the protein has structural homology to a fatty acid binding domain, contains a noncharacteristic catalytic zinc-binding motif, and could potentially act as a zinc metalloprotease. While we have shown *in vitro* that the identified zinc coordinating residues are essential for metal coordination, we have been unable to confirm the importance of the zinc-binding motif in viral replication. Although we were unable to perform these studies, it is our hope that another research group will follow up on characterizing the identified domain *in vivo*. It is our belief that the data presented here provide a significant advance in the understanding and characterization of the HEV nonstructural proteins, and we hope that they will form the building blocks for a more detailed understanding of HEV biology.

MATERIALS AND METHODS

Cloning of HEV constructs. The complete HEV ORF1 (Uniprot code P33424) was synthesized as an *E. coli* codon-optimized gene from Synthetic Genomics Inc. (Fig. 2). Constructs were amplified from the complete optimized ORF1 sequence with primers containing 20 bp of homology (Table 3) to the pET24a plasmid, which had been modified to contain a His₆-MBP-tobacco etch virus (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 QIAquick gel extraction kit (Qiagen). The purified PCR product was assembled into the vector using Gibson Assembly (NEB Gibson Assembly 2× master mix) and transformed into *E. coli* 10G chemically competent cells (Lucigen, Middleton, WI). Colonies resistant to kanamycin 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 HEV⁵¹⁰⁻⁶⁹⁶ was performed using the Q5 site-directed mutagenesis kit (New England BioLabs [NEB]).

Growth and expression of HEV MBP fusion proteins. *E. coli* BL21(DE3) chemically competent cells (NEB product number C2530H) were transformed with 100 ng of pET24a His₆-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). A total of 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 liter of terrific broth (TB) supplemented with 50 mM 3-(N-morpholino)propanesulfonic acid (MOPS; pH 7.5), 1× trace metals (Table 2), 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 in which 1 g/liter of [¹⁵N]ammonium chloride and 3 g/liter of

protonated ^{13}C -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 OD_{600} 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 h 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 was performed using the same protocol. *E. coli* cells containing the overexpressed HEV proteins were resuspended 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 EDTA and passed five times through a cell homogenizer (Microfluidics model M-110P) at 18,000 lb/in². Cell lysate was centrifuged at $35,000 \times g$ for 60 min 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/min. The Ni resin was washed with 100 ml of binding buffer to remove all nonspecifically bound protein before the desired MBP fusion proteins were eluted from the column using 100 ml of 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 U of His-tagged TEV protease per 1 mg of MBP fusion protein and dialyzed against 2 liters of 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 immobilized-metal affinity chromatography (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/min. The Ni resin was washed with binding buffer until all cleaved protein was 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 preequilibrated 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/min. Samples were concentrated to 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) was incubated against 60 μl of precipitant. Screening was run at 4 and 20°C, and trays were monitored for crystallization events using robotics. Suitable crystals grew at 4°C using 18% (wt/vol) PEG 3350, 180 mM LiNO_3 , and 10 mM NiCl_2 . Crystals were cryopreserved using precipitant supplemented with 20% (vol/vol) glycerol for data collection. All X-ray data were collected on a Rigaku FR-E+ SuperBright generator using a Pilatus R 300K charge-coupled device (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 thimerosal for 3 h. 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 (40). This initial model comprised approximately 2/3 of the protein sequence, and the remainder of the model was manually built into 2Fo-Fc and Fo-Fc maps using COOT (41). All coordinates were refined to convergence using BUSTER and the PHENIX suite of programs (42–44). The statistics of data used for refinement are reported in Table 1.

NMR experiments. All NMR experiments were conducted with 500 μl of sample containing 5% D_2O 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 [^{15}N , ^1H]-HSQC experiments were acquired with uniform sampling collecting totals of 1,024 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% nonuniform sampling (45), collecting 2,048 points in the direct dimension and a total of 3,300 points across all indirect dimensions. Nonuniformly sampled data sets were collected using Poisson gap sampling schemes and processed in Topspin 3.5 using the hms1ST algorithm (46, 47). All data for the assignment of backbone atoms were analyzed in CARA (48), and residues which experienced a change in chemical shift between the different constructs were mapped onto the protein structure in PyMOL (49).

Differential scanning fluorimetry. A total of 19 μl of DSF master mix (5 μM apo HEV construct, $10 \times$ Sypro orange fluorescent dye in crystallography buffer) was mixed with 1 μl of additive. The 20- μl reaction mixture was incubated at 25°C for 5 min before the temperature was gradually increased to 95°C at a rate of 1.8°C/min, with fluorescence readings being taken every 0.35°C. Additives were screened using an 8-point 2-fold dilution series of metals, with the top concentration analyzed being 1.25 mM. All measurements were performed in replicates of four, with an average standard deviation of 0.27°C observed between the replicates.

ACKNOWLEDGMENTS

We thank Andreas Lingel and Donald Ganem for their advice and mentorship during the project. We also thank Barbara Leon and Tobias Flecken for advice about construct design and protein expression and purification, as well as their useful comments and

discussion. Finally, we thank Suzanne Emerson and Dianjun Cao for helpful discussions about the HEV replicon.

All authors were employed by the Novartis Institutes for BioMedical Research while this work was conducted.

REFERENCES

- Nan Y, Zhang YJ. 2016. Molecular biology and infection of hepatitis E virus. *Front Microbiol* 7:1–21.
- Abbravanel F, Lhomme S, Dubois M, Peron J-M, Alric L, Kamar N, Izopet J. 2013. Hepatitis E virus. *Med Mal Infect* 43:263–270. <https://doi.org/10.1016/j.medmal.2013.03.005>.
- Navaneethan U, Al Mohajer M, Shata MT. 2008. Hepatitis E and pregnancy—understanding the pathogenesis. *Liver Int* 28:1190–1199. <https://doi.org/10.1111/j.1478-3231.2008.01840.x>.
- Cao D, Meng XJ. 2012. Molecular biology and replication of hepatitis E virus. *Emerg Microbes Infect* 1:1–10. <https://doi.org/10.1038/emi.2012.7>.
- Mushahwar IK. 2008. Hepatitis E virus: molecular virology, clinical features, diagnosis, transmission, epidemiology, and prevention. *J Med Virol* 80:646–658. <https://doi.org/10.1002/jmv.21116>.
- 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. <https://doi.org/10.1155/2007/289059>.
- 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. <https://doi.org/10.1002/jmv.21033>.
- Kamar N, Rostaing L, Abbravanel F, Garrouste C, Esposito L, Cardeau-Desangles I, Mansuy JM, Selves J, Peron JM, Ota 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. <https://doi.org/10.1086/650488>.
- Kamar N, Rostaing L, Abbravanel F, Garrouste C, Lhomme S, Esposito L, Basse G, Cointault O, Ribes D, Nogier MB, Alric L, Peron JM, Izopet J. 2010. Ribavirin therapy inhibits viral replication on patients with chronic hepatitis E virus infection. *Gastroenterology* 139:1612–1618. <https://doi.org/10.1053/j.gastro.2010.08.002>.
- Smith DB, Simmonds P, Jameel S, Emerson SU, Harrison TJ, Meng XJ, Okamoto H, Van der Poel WHM, Purdy MA. 2014. Consensus proposals for classification of the family Hepeviridae. *J Gen Virol* 95:2223–2232. <https://doi.org/10.1099/vir.0.068429-0>.
- Ahmad I, Holla RP, Jameel S. 2011. Molecular virology of hepatitis E virus. *Virus Res* 161:47–58. <https://doi.org/10.1016/j.virusres.2011.02.011>.
- Tam AW, Smith MM, Guerra ME, Huang CC, Bradley DW, Fry KE, Reyes GR. 1991. Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* 185:120–131. [https://doi.org/10.1016/0042-6822\(91\)90760-9](https://doi.org/10.1016/0042-6822(91)90760-9).
- Tsarev SA, Emerson SU, Reyes GR, Tsareva TS, Legters LJ, Malik IA, Iqbal M, Purcell RH, Legters LJ, Li IAM, Li MI, Purcell RH. 1992. Characterization of a prototype strain of hepatitis E virus. *Proc Natl Acad Sci U S A* 89:559–563. <https://doi.org/10.1073/pnas.89.2.559>.
- Ding Q, Heller B, Capuccino JMV, Song B, Nimgaonkar I, Hrebikova G, Contreras JE, Ploss A. 2017. Hepatitis E virus ORF3 is a functional ion channel required for release of infectious particles. *Proc Natl Acad Sci U S A* 114:1147–1152. <https://doi.org/10.1073/pnas.1614955114>.
- Cao D, Huang YW, Meng XJ. 2010. The nucleotides on the stem-loop RNA structure in the junction region of the hepatitis E virus genome are critical for virus replication. *J Virol* 84:13040–13044. <https://doi.org/10.1128/JVI.01475-10>.
- Emerson SU, Zhang M, Meng XJ, Nguyen H, St Claire M, Govindarajan S, Huang YK, Purcell RH. 2001. Recombinant hepatitis E virus genomes infectious for primates: Importance of capping and discovery of a cis-reactive element. *Proc Natl Acad Sci U S A* 98:15270–15275. <https://doi.org/10.1073/pnas.25155098>.
- Parvez MK. 2017. The hepatitis E virus nonstructural polyprotein. *Future Microbiol* 12:915–924. <https://doi.org/10.2217/fmb-2017-0016>.
- Koonin EV, Gorbalenya AE, Purdy MA, Rozanov MN, Reyes GR, Bradley DW. 1992. Computer-assisted assignment of functional domains in the nonstructural polyprotein of hepatitis E virus: delineation of an additional group of positive-strand RNA plant and animal viruses. *Proc Natl Acad Sci U S A* 89:8259–8263. <https://doi.org/10.1073/pnas.89.17.8259>.
- Magden J, Takeda N, Li T, Auvinen P, Ahola T, Miyamura T, Merits A, Kaariainen L. 2001. Virus-specific mRNA capping enzyme encoded by hepatitis E virus. *J Virol* 75:6249–6255. <https://doi.org/10.1128/JVI.75.14.6249-6255.2001>.
- Agrawal S, Gupta D, Panda SK. 2001. The 3' end of hepatitis E virus (HEV) genome binds specifically to the viral RNA-dependent RNA polymerase (RdRp). *Virology* 282:87–101. <https://doi.org/10.1006/viro.2000.0819>.
- Karpe YA, Lole KS. 2010. NTPase and 5' to 3' RNA duplex-unwinding activities of the hepatitis E virus helicase domain. *J Virol* 84:3595–3602. <https://doi.org/10.1128/JVI.02130-09>.
- Ansari IH, Nanda SK, Durgapal H, Agrawal S, Mohanty SK, Gupta D, Jameel S, Panda SK. 2000. Cloning, sequencing, and expression of the hepatitis E virus (HEV) nonstructural open reading frame 1 (ORF1). *J Med Virol* 60:275–283. [https://doi.org/10.1002/\(SICI\)1096-9071\(200003\)60:3<275::AID-JMV5>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1096-9071(200003)60:3<275::AID-JMV5>3.0.CO;2-9).
- Ropp SL, Tam AW, Beames B, Purdy M, Frey TK. 2000. Expression of the hepatitis E virus ORF1. *Arch Virol* 145:1321–1337. <https://doi.org/10.1007/s007050070093>.
- Sehgal D, Thomas S, Chakraborty M, Jameel S. 2006. Expression and processing of the hepatitis E virus ORF1 nonstructural polyprotein. *Virol J* 3:1–9.
- Paliwal D, Panda SK, Kapur N, Varma SPK, Durgapal H. 2014. Hepatitis E virus (HEV) protease: a chymotrypsin-like enzyme that processes both non-structural (pORF1) and capsid (pORF2) protein. *J Gen Virol* 95:1689–1700. <https://doi.org/10.1099/vir.0.066142-0>.
- Parvez MK, Khan AA. 2014. Molecular modeling and analysis of hepatitis E virus (HEV) papain-like cysteine protease. *Virus Res* 179:220–224. <https://doi.org/10.1016/j.virusres.2013.11.016>.
- Proudfoot A, Hyrini A, Holdorf M, Frank AO, Bussiere D. 2019. First crystal structure of a nonstructural hepatitis E virus protein identifies a putative novel zinc-binding protein. *bioRxiv* <https://doi.org/10.1101/538884>.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg M. 2015. The Phyre2 web portal for protein modelling, prediction, and analysis. *Nat Protoc* 10:845–858. <https://doi.org/10.1038/nprot.2015.053>.
- Parvez MK. 2013. Molecular characterization of hepatitis E virus ORF1 gene supports a papain-like cysteine protease (PCP)-domain activity. *Virus Res* 178:553–556. <https://doi.org/10.1016/j.virusres.2013.07.020>.
- Holm L, Laakso LM. 2016. Dali server update. *Nucleic Acids Res* 44:W351–W355. <https://doi.org/10.1093/nar/gkw357>.
- Parvez MK. 2015. The hepatitis E virus ORF1 “X-domain” residues form a putative macrodomain protein/Appr-1”-pase catalytic-site, critical for viral RNA replication. *Gene* 566:47–53. <https://doi.org/10.1016/j.gene.2015.04.026>.
- Benters J, Flögel U, Schäfer T, Leibfritz D, Hechtenberg S, Beyersmann D. 1997. Study of the interactions of cadmium and zinc ions with cellular calcium homeostasis using 19F-NMR spectroscopy. *Biochem J* 322:793–799. <https://doi.org/10.1042/bj3220793>.
- McCall KA, Huang C, Fierke CA. 2000. Function and mechanism of zinc metalloenzymes. *J Nutr* 130:1437S–1446S. <https://doi.org/10.1093/jn/130.5.1437S>.
- Suppliah S, Zhou Y, Frey TK. 2011. Lack of processing of the expressed ORF1 gene product of hepatitis E virus. *Virol J* 8:245. <https://doi.org/10.1186/1743-422X-8-245>.
- Perttola J, Spuul P, Ahola T. 2013. Early secretory pathway localization and lack of processing for hepatitis E virus replication protein pORF1. *J Gen Virol* 94:807–816. <https://doi.org/10.1099/vir.0.049577-0>.
- Emerson SU, Nguyen H, Graff J, Stephany DA, Brockington A, Purcell RH. 2004. In vitro replication of hepatitis E virus (HEV) genomes and of an HEV replicon expressing green fluorescent protein. *J Virol* 78:4838–4846. <https://doi.org/10.1128/JVI.78.9.4838-4846.2004>.
- Graff J, Torian U, Nguyen H, Emerson SU. 2006. A bicistronic subgenomic mRNA encodes both the ORF2 and ORF3 proteins of hepatitis E virus. *J Virol* 80:5919–5926. <https://doi.org/10.1128/JVI.00046-06>.
- Fuss JO, Tsai CL, Ishida JP, Tainer JA. 2015. Emerging critical roles of Fe-S

- clusters in DNA replication and repair. *Biochim Biophys Acta* 1853: 1253–1271. <https://doi.org/10.1016/j.bbamcr.2015.01.018>.
39. Brea RJ, Cole CM, Lyda BR, Ye L, Prosser RS, Sunahara RK, Devaraj NK. 2017. Hepatitis B virus protein has unusual iron-sulfur cluster. *C&EN* 95(15):11.
 40. Terwilliger T. 2004. SOLVE and RESOLVE: automated structure solution, density modification, and model building. *J Synchrotron Radiat* 11: 49–52. <https://doi.org/10.1107/S0909049503023938>.
 41. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501. <https://doi.org/10.1107/S0907444910007493>.
 42. Bricogne G, Blanc E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart OS, Vonrhein C, Womack TO. 2016. BUSTER. 2.11.6. Global Phasing Ltd, Cambridge, United Kingdom.
 43. Adams PD, Afonine PV, 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 D Biol Crystallogr* 66:213–221. <https://doi.org/10.1107/S0907444909052925>.
 44. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD. 2008. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64:61–69. <https://doi.org/10.1107/S090744490705024X>.
 45. Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY. 2005. High-resolution four-dimensional ^1H - ^{13}C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. *J Am Chem Soc* 127:2767–2775. <https://doi.org/10.1021/ja044032o>.
 46. 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. <https://doi.org/10.1007/s10858-012-9611-z>.
 47. 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. <https://doi.org/10.1021/ja908004w>.
 48. Keller R. 2004. The computer aided resonance assignment tutorial. Cantina Verlag, Goldau, Switzerland.
 49. Schrödinger, LLC. 2015. The PyMOL molecular graphics system, version 1.8. Schrödinger, LLC, New York, NY.
 50. Nicesola D, Perduca M, Capaldi S, Carrizo ME, Righetti PG, Monaco HL. 2004. Crystal structure of chicken liver basic fatty acid-binding protein complexed with cholic acid. *Biochemistry* 43:14072–14079. <https://doi.org/10.1021/bi0489661>.