Hepatitis E virus (HEV) protease: a chymotrypsin-like enzyme that processes both non-structural (pORF1) and capsid (pORF2) protein

Daizy Paliwal, Subrat Kumar Panda, Neeraj Kapur, Satya Pavan Kumar Varma and Hemlata Durgapal

Hepatitis E virus (HEV), a major cause of acute viral hepatitis across the world, is a non-enveloped, plus-strand RNA virus. Its genome codes three proteins, pORF1 (multifunctional polyprotein), pORF2 (capsid protein) and pORF3 (multi-regulatory protein). pORF1 encodes methyltransferase, putative papain-like cysteine protease, helicase and replicase enzymes. Of these, the protease domain has not been characterized. On the basis of sequence analysis, we cloned and expressed a protein covering aa 440–610 of pORF1, expression of which led to cell death in *Escherichia coli* BL-21 and Huh7 hepatoma cells. Finally, we expressed and purified this protein from *E. coli* C43 cells (resistant to toxic proteins). The refolded form of this protein showed protease activity in gelatin zymography. Digestion assays showed cleavage of both pORF1 and pORF2 as observed previously. MS revealed digestion of capsid protein at both the N and C termini. N-terminal sequencing of the ~35 kDa methyltransferase, ~35 kDa replicase and ~56 kDa pORF2 proteins released by protease digestion revealed that the cleavage sites were alanine15/isoleucine16, alanine1364/valine1365 in pORF1 and leucine197/valine198 in pORF2. Specificity of these cleavage sites was validated by site-directed mutagenesis. Further characterization of the HEV protease, carried out using twelve inhibitors, showed chymostatin and PMSF to be the most efficient inhibitors, indicating this protein as a chymotrypsin-like protease. The specificity was further confirmed by cleavage of the chymotrypsin-specific fluorogenic peptide N-succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin. Mutational analysis of the conserved serine/cysteine/histidine residues suggested that H443 and C472/C481/C483 are possibly the active site residues. To our knowledge, this is the first direct demonstration of HEV protease and its function.

**INTRODUCTION**

Hepatitis E virus (HEV), a feco-orally transmitted, non-enveloped virus, is the causative agent of hepatitis E and has been classified as a member of the genus *Hepevirus* in the family *Hepeviridae* (Viswanathan 1957; Purdy *et al.*, 1993). HEV has a 7.2 kb positive-strand, capped, polyadenylated RNA genome (Tam *et al.*, 1991; Reyes *et al.*, 1993; Kabrane-Lazizi *et al.*, 1999) that, like those of other viruses, needs to be translated and produce viral enzymes before the replication cycle can start. The HEV *ORF1* codes for the functional polyprotein (pORF1) (Koonin *et al.*, 1992) containing active methyltransferase (MT; Magden *et al.*, 2001), helicase (Karpe & Lole, 2010 a, b) and replicase (Agrawal *et al.*, 2001) domains, which have been partially characterized. A putative papain-like cysteine protease (PCP) and two domains (X and Y domains) with unknown functions have also been proposed (Koonin *et al.*, 1992), but these have not been fully characterized yet. However, it is suggested that the X domain is a macrodomain (PFAM family code: pf01661), involved in nucleic acid binding. The HEV *ORF2* codes for a capsid protein (pORF2), which undergoes endoplasmic recycling and proteolytic processing (Torresi *et al.*, 1999; Zafrullah *et al.*, 1999; Surjit *et al.*, 2007). The HEV *ORF3* produces a small 121 aa phosphoprotein (pORF3) with putative multi-regulatory functions (Zafrullah *et al.*, 1997; Korkaya *et al.*, 2001; Graff *et al.*, 2006).

A functional HEV protease, essential for understanding the biology of this virus, has been neither expressed nor characterized for its nature and activity. This protease may well serve as an important antiviral drug target.
We describe here, for the first time to our knowledge, cloning, expression and purification of the active recombinant HEV protease, its target cleavage sites and possible active site amino acids. On the basis of the site of digestion, inhibitor sensitivity, cleavage of a chymotrypsin-specific peptide, and serine/cysteine/histidine mutagenesis, we propose HEV protease to be a chymotrypsin-like protease.

**RESULTS**

**In silico analysis of HEV ORF1 protein (pORF1)**

Homology analysis of 36 HEV strains available in the NCBI GenBank database showed conservation of the N terminus of the protease domain (aa 440–520) across the strains (Fig. S1, available in the online Supplementary Material). On the basis of these observations, we cloned 1345–1857 nt to look for the functional protease domain in this region of **ORF1**.

**Expression of HEV protease in both *Escherichia coli* and Huh7 cells**

HEV protease overexpression killed *E. coli* BL-21 cells (Stratagene) after IPTG induction (1 mM, 0.8 mM, 0.5 mM and 0.2 mM) at 37 °C. Reduction of the induction temperature to 25 °C did not prevent cell death. However, HEV protease could be expressed in *E. coli* C43 (Lucigen; Dumon-Seignovert *et al.*, 2004) (known to express toxic recombinant proteins) without causing cell death. The best expression levels were obtained at 25 °C for 12 h with 0.5 mM IPTG. The purified protease appeared as a single band in SDS-PAGE (Fig. S2a) at the expected molecular mass, which was further confirmed and validated by Western blotting using anti-histidine antibody (Sigma-Aldrich) (Fig. S2b).

To confirm that the HEV protease led to Huh7 cell death, we used a pcDNA3-protease–linker–EGFP (PLE) construct. The Huh7 cells transfected with pcDNA3–PLE turned dysmorphic as compared with control cells expressing pcDNA3–EGFP (Fig. S3a), indicating stress and signs of cell death. PLE-expressing cells also stained positive for propidium iodide, indicating loss of membrane integrity (a marker for cell death) due to protease expression (Fig. S3c). The cells transfected with pcDNA3–PLE showed no DNA fragmentation (Fig. S4), indicating cell death by lysis and not apoptosis. The killing of both bacterial and mammalian cells presented initial and indirect proof that the HEV protease was functional.

Immunoprecipitation of 35S-radiolabelled PLE (expressed in a rabbit reticulocyte transcription and translation system) with anti-EGFP polyclonal antibody showed two proteins of ~51 kDa and ~27 kDa, corresponding to full-length PLE fusion protein and EGFP, respectively (Fig. S3b). In a parallel control reaction, expression of ~27 kDa EGFP was observed. This indicated the possible cleavage of the glycine/alanine-rich linker by the protease of the PLE fusion protein.

**Functional characterization of HEV protease**

Refolded HEV protease cleaved gelatin (a standard non-specific protein substrate used to assay proteases) in a zymography assay (Fig. 1a). No contaminating protease activity was observed in the controls (Fig 1b, c).

Immunoprecipitation of protease-digested 67 kDa MT-PCP with polyclonal rabbit anti-MT antibody yielded a ~35 kDa protein (Fig. 2a, lanes 2 and 3). No self-cleavage was observed in MT-PCP protein incubated for 2 h in the absence of protease (Fig. 2a, lane 1). Two replicate fragments of ~55 and ~35 kDa (Fig. 2b, lanes 3 and 4) were immunoprecipitated with anti-replicase antibody from protease-digested 84 kDa helicase-replicase protein at 1 h and 2 h time intervals. In parallel control reactions (Fig. 2b, lane 1 and 2), helicase-replicase did not self-cleave in the absence of protease, indicating specificity of the assay. HEV protease also cleaved *E. coli*-expressed and purified full-length capsid protein (76 kDa) into a stable ~56 kDa fragment (Fig. 2c). The digestion pattern of capsid protein was analysed for 30 min. We also observed some intermediates between the full-length 76 kDa capsid protein (72 kDa pORF2 + 4 kDa N-terminal histidine tag from the pRSET vector) and the stable ~56 kDa fragment.

![Fig. 1.](image-url)
at the end of the 10 min digestion reaction (Fig. 2c, lane 2). However, with increased digestion time (20 and 30 min.), ~56 kDa appeared to be the most stable digestion product (Fig. 2c, lane 3 and 4). The specificity of the ~56 kDa digestion product was confirmed by Western blot using anti-ORF2 antibody (Fig. 2d).

Mass spectrometric analysis (using MASCOT software) of peptide fingerprints obtained for both full-length 76 kDa and ~56 kDa capsid proteins showed that the peptides towards the N and C termini were missing in the ~56 kDa fragment (Fig. 3a). This indicated that HEV protease cleaved capsid protein at both the N and C termini.

**Identification of HEV protease cleavage sites**

N-terminal sequencing of the ~35 kDa MT, ~35 kDa replicase and ~56 kDa capsid protein fragments (released after digestion of MT-PCP, helicase-replicase and capsid protein) revealed alanine15/isoleucine16, alanine1364/valine1365 in pORF1 (GenBank accession number: FJ457024) and leucine197/valine198 in capsid protein (FJ457024) as HEV protease cleavage sites.

**Validation of observed cleavage sites**

The alanine15/isoleucine16, alanine1364/valine1365 and leucine197/valine198 sites were mutated to methionine/methionine in the pRSET B-Methyltransferase, pRSET B-Helicase-Replicase and pRSET A-ORF2 constructs, respectively. On digestion assay, these purified mutant proteins were found to be resistant to HEV protease cleavage, validating the specificity of the cleavage sites (Fig. 4). We observed that the ~35 kDa MT, ~35 kDa replicase and ~56 kDa pORF2 fragments were not obtained in the digestion of mutant proteins (Fig. 4a, lanes 5–7; Fig. 4b, lanes 1–4; and Fig. 4c, lanes 5–7). However, all the other fragments were identical in the wild-type and mutant protein digestion patterns.

Mapping of the leucine197/valine198 cleavage site in the capsid protein structure predicted using GOR 4 software (Garnier-Osguthorpe-Robson; Garnier et al., 1996) showed that this site is present in the unstructured region of pORF2 (Fig. 3b). The X-ray crystal structure of HEV capsid protein, available from the Protein Data Bank (PDB; http://www.ebi.ac.uk/pdbe-srv/view/entry/2ztn/secondary.html), also validated the lack of secondary structure in the protein region encompassing amino acids leucine197/valine198.

**Sensitivity of HEV protease towards known inhibitors**

Activity of HEV protease was tested in the presence of twelve different protease inhibitors in independent reactions...
Fig. 3. (a) Mass spectrometric analysis of full-length and ~56 kDa capsid proteins. (i) and (ii) represent the analysis of MS results and output files obtained from MASCOT software for the full-length 76 kDa and processed ~56 kDa pORF2 fragment, respectively. Matched peptides are shown in red. Peptides towards both the N and C termini were not present in the ~56 kDa pORF2 fragment indicating that the HEV Protease cleaves pORF2 at both the N and C termini. Subsequent N-terminal analysis...
of the ~56 kDa fragment showed that the cleavage occurred between leucine197/valine198 [circled in (ii)]. (b) Mapping of the HEV protease cleavage site on pORF2. The leucine197/valine198 site was found to be present in the unstructured region of pORF2 in the gOR 4 software predicted structure shown here.

(Fig. 5). We observed that HEV protease was most efficiently inhibited by PMSF (99.6 %) and chymostatin (99.4 %). E-64d, a known cell-permeable cysteine protease inhibitor (derivative of E-64) did not demonstrate any effective inhibition of HEV protease. The cell death in PLE-expressing Huh7 cells did not differ in the presence of E-64d (100 μM, dissolved in DMSO) for 72 h (Fig. S5a, b). As a control, an equal volume of DMSO was added for the same length of time to parallel pcDNA3-PLE and pcDNA3-EGFP transfected Huh7 cells (Fig. S5c, d). However, no difference was observed in the presence of E-64d or DMSO in either PLE- or EGFP-expressing Huh7 cells.

Characterization of HEV protease using a fluorogenic chymotrypsin-specific peptide

The efficient cleavage of a chymotrypsin-specific, internally quenched, fluorogenic peptide [N-succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin (AMC)] by recombinant HEV protease validated leucine/valine as the preferential cleavage site and its classification as a chymotrypsin-like enzyme (Table 1).

The activity of PMSF-inhibited HEV protease was not reversible in the presence of DTT (1 mM, 5 mM or 10 mM), indicating irreversible inhibition of protease by PMSF. The percentage of HEV protease inhibition achieved at all the tested DTT concentrations was similar and did not differ (Table S1).

Mutational analysis of HEV protease for the identification of active site residues

Mutation of the conserved serine (S454, S468, S505, S512, S515 and S518), cysteine (C457, C459, C471, C472, C481 and C483) and histidine (H443 and H590) residues to alanine demonstrated the significance of the H443, C472, C481 and C483 residues in HEV protease activity (Fig. 6). Based on the digestion assays performed with purified mutant proteases, H443 and one of the cysteines (C472, C481 or C483) appear to be the possible active site residues.

DISCUSSION

The positive-strand RNA viruses need to produce a negative-sense intermediate in order to replicate (Lai, 1998). To synthesize this replication intermediate, they need a functional replicase and other viral enzymes. Most of these viruses use either cap-mediated translation or an internal ribosome entry site (IRES) to translate their early proteins (Buck, 1996).

HEV, being a capped, plus-stranded RNA virus, uses cap-dependent translation to synthesize ORF1, which is the first ORF from the 5’ end. The HEV MT (Magden et al., 2001), helicase (Karpe & Lole 2010 a, b) and replicase (Agrawal et al., 2001) proteins have been cloned, expressed and partially characterized. However, attempts to produce an
active protease and characterize it have failed consistently. It is in the backdrop that pORF1, when expressed as such in liver cells, produces a polyprotein of ~186 kDa, which does not undergo any self-processing (Ansari et al., 2000; Suppiah et al., 2011; Perttilä et al., 2013). Few investigators have shown processing of pORF1 in heterologous expression systems such as insect cells (Sehgal et al., 2006) and vaccinia virus systems (Ropp et al., 2000). However, the processed fragments produced in these systems are different in size from those obtained from HEV-replicon-transfected HepG2 hepatoma cells (Panda et al., 2000). In addition, the activity of these products has never been assayed. Therefore, the ORF1 polyprotein is possibly processed only in the context of the virus life cycle. As observed previously, the yield of processed MT, helicase and replicase domains was low and not enough to perform N-terminal sequencing to define the protease cleavage sites of pORF1.

We initiated our investigations by expressing the predicted protease domain (aa 440–610) in hepatoma cells. We observed that transfection of the pcDNA3-Protease into Huh7 cells resulted in cell death. The pcDNA3-PLE construct was generated to visualize and verify the association of cell death with HEV protease expression. All PLE-transfected cells were dysmorphic and stained positively with propidium iodide (Fig. S3a, c) in addition to EGFP fluorescence. Since the genomic DNA extracted from PLE-expressing dead Huh7 cells was not fragmented (Fig. S4), we concluded that HEV protease led to cell lysis and not apoptosis. Similarly, attempts to express the putative protease in E. coli BL-21 cells led to bacteriolysis.

These results provided the initial proof of toxicity and activity associated with expression of this protein. Similar toxicity and death of E. coli have been shown previously to

### Table 1. Digestion assay with fluorogenic chymotrypsin-specific peptide

<table>
<thead>
<tr>
<th>Inhibitor*</th>
<th>Specificity of inhibitor</th>
<th>FSU reading at 460 nm</th>
<th>Normalized FSU reading</th>
<th>Inhibition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncut peptide</td>
<td>–</td>
<td>1332.44</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Without inhibitor (negative control)</td>
<td>–</td>
<td>985 564.31</td>
<td>983 851.42</td>
<td>–</td>
</tr>
<tr>
<td>AEBSF</td>
<td>Irreversible inhibitor of serine proteases</td>
<td>514 889.44</td>
<td>513 176.55</td>
<td>47.8</td>
</tr>
<tr>
<td>ALLN</td>
<td>Peptide aldehyde inhibitor of neutral cysteine proteases</td>
<td>747 817.62</td>
<td>746 104.73</td>
<td>24.3</td>
</tr>
<tr>
<td>Aprotinin</td>
<td>Inhibits serine proteases</td>
<td>514 858.52</td>
<td>513 145.63</td>
<td>47.8</td>
</tr>
<tr>
<td>Bestatin</td>
<td>Inhibits amino peptidases and exopeptidases</td>
<td>750 485.65</td>
<td>848 772.73</td>
<td>13.7</td>
</tr>
<tr>
<td>Antipain</td>
<td>Inhibits papain, trypsin and plasmin</td>
<td>703 419.31</td>
<td>701 706.44</td>
<td>28.6</td>
</tr>
<tr>
<td>Chymostatin</td>
<td>Inhibitor of chymotrypsin</td>
<td>7 432.66</td>
<td>5 719.77</td>
<td>99.4</td>
</tr>
<tr>
<td>E-64</td>
<td>Inhibitor of cysteine proteases</td>
<td>750 526.56</td>
<td>748 813.67</td>
<td>23.8</td>
</tr>
<tr>
<td>EDTA</td>
<td>Inhibits metalloproteases</td>
<td>878 742.75</td>
<td>877 029.86</td>
<td>10.8</td>
</tr>
<tr>
<td>Leupeptin</td>
<td>Inhibits serine and cysteine proteases</td>
<td>896 988.09</td>
<td>895 275.22</td>
<td>9.0</td>
</tr>
<tr>
<td>Pepstatin</td>
<td>Inhibits aspartic acid proteases</td>
<td>893 942.37</td>
<td>892 229.48</td>
<td>9.3</td>
</tr>
<tr>
<td>Phosphoramidon</td>
<td>Inhibitor of metalloproteases</td>
<td>872 789.06</td>
<td>871 076.17</td>
<td>11.4</td>
</tr>
<tr>
<td>PMSF</td>
<td>Irreversible inhibitor of serine proteases and reversible inhibitor of cysteine proteases</td>
<td>5 217.03</td>
<td>3 504.14</td>
<td>99.6</td>
</tr>
</tbody>
</table>

*Full names of inhibitors are given in Methods.
be associated with viral protease expression (Baum et al., 1990). The mutants of HIV protease were isolated on the basis of toxicity of the enzyme in E. coli (Baum et al., 1990). HIV protease inhibitors have also been screened by their ability to confer protection against activity-mediated cytotoxicity in E. coli (Cheng et al., 2006).

The E. coli C43-expressed, purified protein (Fig. S2a) was dialysed and refolded. Digestion of gelatin in zymography (Fig. 1a) gave the first direct evidence of protease activity in the expressed putative protease domain (440 to 610 aa) of pORF1. This permitted us to study pORF1 processing with the active recombinant HEV protease. However, pORF1, being a large (~186 kDa) protein, does not give high yields when expressed, as required for protease digestion assays and subsequent N-terminal sequencing. To circumvent this technical hurdle, we cloned pORF1 in two fragments so as to cover two domains and the inter-domain region, which contains the predicted sites for proteolysis. The ~35 kDa MT protein and ~35 kDa replicase fragments were obtained from the protease digestion of the 67 kDa MT-PCP protein (Fig. 2a) and 84 kDa helicase-replicase proteins, respectively (Fig. 2b). The above-obtained fragments were similar in molecular mass to the ones observed in our earlier immunoprecipitation studies with replicon-transfected HepG2 cells (Panda et al., 2000). In addition, we also observed a ~55 kDa immediately processed form of the replicase protein. The immunoprecipitation of similar-sized domains from both our protease digestion assays and earlier experiments with HEV-replicon-transfected hepatoma cells suggests the involvement of HEV protease in pORF1 processing observed in the context of the virus life cycle. This is in contrast with another study where a protein covering the MT-PCP domain (nt 1–1710) failed to cleave the helicase-replicase protein (Karpe & Lole, 2011). Previously, we reported that the molecular size of HEV capsid protein on the secreted virions is lower (~56 kDa) than the full-length capsid protein (Kapur et al., 2012). The endoplasmic recycling and cytoplasm retrotranslocation of HEV capsid protein have been described (Torresi et al., 1999; Zafrullah et al., 1999; Surjit et al., 2007). With this background, we explored the possibility of involvement of HEV protease in capsid protein maturation. The recombinant HEV protease digested the 76 kDa capsid protein into a ~56 kDa fragment (Fig. 2c). Mass spectrometric analysis (Fig. 3a) of the eluted, trypsin-digested 76 kDa and ~56 kDa capsid protein fragments revealed that HEV protease cleaved at both the N and C termini.

Amino (N) terminal analysis of the ~35 kDa MT, ~35 kDa replicase and ~56 kDa capsid proteins showed that the recombinant protease cleaved at alanine15/isoleucine16, alanine1364/valine1365 in pORF1 and leucine197/valine198 in capsid protein. Based on these results, the consensus sequence we obtained for HEV protease cleavage sites is [A,L]/[I,V]. Viral proteases, unlike digestive protease enzymes, cleave inter-domain regions in polyproteins through the process of limited proteolysis (Babe´ & Craik, 1997). Therefore, although the sites may be present multiple times in the genome, the limited proteolysis by viral proteases preferentially cleaves accessible sites present in the folded polypeptide.

Site-directed mutagenesis of the observed cleavage site to methionine/methionine in pRSET B-MT, pRSET B-Helicase-Replicase and pRSET A-ORF2 constructs confirmed their specificity. The above mutations of the protease cleavage site abolished proteolysis (Fig. 4). The
protease-digested forms of the MT and replicase proteins begin with isoleucine16 and valine1365 of pORF1, respectively.

Mapping of the leucine/valine site on the known pORF2 X-ray crystallographic structure (Yamashita et al., 2009) available in the PDB (entry Q1AHU7) showed that this site is present in an unordered region devoid of secondary structure (Fig. 3b). The cleavage of the flexible linker (unstructured glycine/alanine-rich region) observed during pcDNA3-PLE expression in rabbit reticulocyte lysate (Fig. S3b) also indicated the preference of HEV protease to cleave in unstructured protein segments, like polypeptide inter-domain regions (Hubbard, 1998).

In the absence of active site identification, protease inhibitors give valuable hints regarding the category of functional active sites. We tested twelve inhibitors for HEV protease inhibition and observed that HEV protease was most effectively inhibited by chymostatin (99.4 % inhibition) and PMSF (99.6 % inhibition) (Fig. 5, Table 1). In the case of HEV protease, the PMSF interaction was not reversible in the presence of 10 mM DTT (Table S1).

Lastly, efficient cleavage of chymotrypsin-specific, internally quenched, fluorogenic peptide substrate confirmed our findings of the similarity of HEV protease to chymotrypsin. On the basis of (1) the inhibition of HEV protease by chymostatin and PMSF, (2) specific recognition and cleavage of a leucine-valine bond by HEV protease (known cleavage site for chymotrypsin; Zimmerman & Ashe, 1977), and (3) cleavage of a chymotrypsin-specific fluorogenic substrate, we propose HEV protease to be a chymotrypsin-like protease. Even among chymotrypsin-like enzymes, HEV protease cleavage sites are very similar to elastase cleavage sites.

HEV proteases carrying mutations of conserved serine, cysteine and histidine residues were purified from E. coli C43 cells, refolded and used for performing digestion assays. We observed the loss of activity in a protease carrying H443A and C472A/C481A/C483A mutations (GenBank accession no. FJ457024). Therefore, we suggest H443 and C472/C481/C483 as amino acids involved in the formation of active sites (Fig. 6).

Recently, the activity of HEV protease has been reported in the HEV-SAR55 replicon in human S10-3 cells (Parvez, 2013). In this study, HEV protease activity and recognition of the proposed substrate (conserved G815-G816-G817 in the X domain) have been studied with respect to the production of GFP in a SAR55 replicon containing GFP. The putative HEV protease domain (1345–1857 nt of HEV; GenBank accession no. FJ457024) was PCR-amplified (primer sequences in Table 1) using a combination of Taq (Promega)/Pfu (Stratagene) (9:1) DNA polymerases from a full-length HEV cDNA clone. The amplified product was cloned into BamHI- and XhoI-digested pRSET-A vector (Invitrogen) and pcDNA3 vector (Invitrogen). The pRSET-A-Protease construct was transformed into E. coli BL-21 DE3 (Stratagene) and E. coli C43 (Lucigen) cells.

A single colony from pRSET-A-Protease-transformed E. coli C43 cells was inoculated overnight at 37 °C in LB media. The next day, primary culture was inoculated into fresh LB media at 1:99 ratio and grown for 2–3 h until the optical density reached 0.6, at which point 0.5 mM IPTG (Sigma-Aldrich) was added for overnight induction at 25 °C. Cells were pelleted and lysate was prepared as described previously (Agrawal et al., 2001). The lysate was resolved on a 12 % SDS-PAGE gel and stained with Coomassie brilliant blue (Sigma-Aldrich) (Fig. S2a). The purification of induced protein was carried out from 1000 ml culture using affinity purification on Ni-NTA resin (Qiagen) as per the manufacturer’s instructions. The purity of the purified protein was checked by SDS-PAGE (Fig. S2a) and Western blot using anti-His antibody (Sigma-Aldrich) (Fig. S2b).

To visualize the cells expressing protease, the protease domain was subcloned in pcDNA3 vector (Invitrogen) as a fusion construct with EGFP, separated by a flexible linker (pcDNA3-PLE). The EGF fragment was amplified from pEGFP1 (Clontech). The sequence of the linker [(Gly)6-Ala-(Gly)6-Ala-(Gly)6] is given in Table 2. pcDNA3-PLE and pcDNA3-EGFP (2.0 μg) were transfected into 1 × 10⁶ Huh7 cells in independent 60 mm culture plates (Fig. S3a) using Lipofectamine 2000 reagent (Invitrogen) as per the manufacturer’s instructions and the fluorescence was observed using a Nikon ECLIPSE TE2000 U fluorescence microscope at 24 and 48 h post-transfection.

**METHODS**

Cloning and expression of putative HEV protease from a full-length HEV clone. The putative HEV protease domain (1345–1857 nt of HEV; GenBank accession no. FJ457024) was PCR-amplified (primer sequences in Table 1) using a combination of Taq (Promega)/Pfu (Stratagene) (9:1) DNA polymerases from a full-length HEV cDNA clone. The amplified product was cloned into BamHI- and XhoI-digested pRSET-A vector (Invitrogen) and pcDNA3 vector (Invitrogen). The pRSET-A-Protease construct was transformed into E. coli BL-21 DE3 (Stratagene) and E. coli C43 (Lucigen) cells.

A single colony from pRSET-A-Protease-transformed E. coli C43 cells was inoculated overnight at 37 °C in LB media. The next day, primary culture was inoculated into fresh LB media at 1:99 ratio and grown for 2–3 h until the optical density reached 0.6, at which point 0.5 mM IPTG (Sigma-Aldrich) was added for overnight induction at 25 °C. Cells were pelleted and lysate was prepared as described previously (Agrawal et al., 2001). The lysate was resolved on a 12 % SDS-PAGE gel and stained with Coomassie brilliant blue (Sigma-Aldrich) (Fig. S2a). The purification of induced protein was carried out from 1000 ml culture using affinity purification on Ni-NTA resin (Qiagen) as per the manufacturer’s instructions. The purity of the purified protein was checked by SDS-PAGE (Fig. S2a) and Western blot using anti-His antibody (Sigma-Aldrich) (Fig. S2b).

To visualize the cells expressing protease, the protease domain was subcloned in pcDNA3 vector (Invitrogen) as a fusion construct with EGFP, separated by a flexible linker (pcDNA3-PLE). The EGF fragment was amplified from pEGFP1 (Clontech). The sequence of the linker [(Gly)6-Ala-(Gly)6-Ala-(Gly)6] is given in Table 2. pcDNA3-PLE and pcDNA3-EGFP (2.0 μg) were transfected into 1 × 10⁶ Huh7 cells in independent 60 mm culture plates (Fig. S3a) using Lipofectamine 2000 reagent (Invitrogen) as per the manufacturer’s instructions and the fluorescence was observed using a Nikon ECLIPSE TE2000 U fluorescence microscope at 24 and 48 h post-transfection.

Among the different categories of protease, classified on the basis of active site nucleophile, serine and cysteine proteases form two distinct families. However, based upon experiments and secondary structure predictions, the relatedness between chymotrypsin, chymotrypsin-like serine proteases and cysteine proteases of some positive-strand RNA viruses (such as hepatitis A virus and poliovirus) became evident. Later, the crystal structure of cysteine proteases of plus-stranded RNA viruses [hepatitis A virus (Allaire et al., 1994), poliovirus (Mosimann et al., 1997) and rhinovirus (Petersen et al., 1999)] confirmed their classification as chymotrypsin-like cysteine proteases.

Such observations led to reidentification of principal catalytic residues of some viral proteases (Gorbalenya et al., 1989) and their classification to distinct protein superfamilies, which share common structural folds. Therefore, the classification of HEV protease depends not only on the involvement of histidine and cysteine but also on structure and sequence homology. Even though we have classified it as a ‘chymotrypsin-like protease’, it may alternatively be called a ‘class-specific cysteine protease’ until the structure and exact active site are determined.

The determination of the crystal structure and development of specific protease inhibitors may help in managing HEV infections, particularly those following organ transplantation and immuno-suppression.
HEV Protease

Propidium iodide staining. The pcDNA3-PLE-transfected, pcDNA3-EGFP-transfected and control Huh7 cells were stained with propidium iodide as previously described (Rieger et al., 2011) and observed with a Nikon ECLIPSE TE2000 U fluorescence microscope at 488 nm for EGFP and 535 nm for PI (Fig. S3c).

DNA fragmentation assay. The pcDNA3-PLE-transfected and normal Huh7 control cells were harvested at 24 and 48 h post-transfection and genomic DNA was isolated as described by Zhang et al. (1998). The samples were analysed further on 2% agarose gels (Fig. S4).

Gelatin zymography assay of purified recombinant putative HEV protease. Refolding of purified HEV protease was performed at 4°C by dialysis over reducing urea (Sigma-Aldrich) concentrations from 8 M to 0.5 M in a step-wise gradient (with buffer change after every 12 h) in 20 mM Tris/HCl (pH 7.4) (Sigma-Aldrich), 0.5 M NaCl (Sigma-Aldrich), 0.4 M L-arginine (Sigma-Aldrich), 0.5 mM DTT (Sigma-Aldrich) and 10% glycerol (Sigma-Aldrich).

The purified refolded HEV protease was quantified with Bradford reagent (Bio-Rad) using BSA as a standard. The activity of refolded

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protease</td>
<td>Forward: 5'-CGGGATCCATGAGCGCGGCTTTTCACTTTGACCCACGTG-3'</td>
</tr>
<tr>
<td>MT</td>
<td>Reverse: 5'-GGCTCGAGATATGTGACACTGAGTAAAGGTCAGCGGACCAGAT-3'</td>
</tr>
<tr>
<td>MT-PCP</td>
<td>Forward: 5'-CCTCGAGATATGTGACACTGAGTAAAGGTCAGCGGACCAGAT-3'</td>
</tr>
<tr>
<td>Helicase-replicase</td>
<td>Reverse: 5'-GGCTCGAGATATGTGACACTGAGTAAAGGTCAGCGGACCAGAT-3'</td>
</tr>
<tr>
<td>EGFP</td>
<td>Forward: 5'-GGCTCGAGATATGTGACACTGAGTAAAGGTCAGCGGACCAGAT-3'</td>
</tr>
<tr>
<td>L/V to M/M mutant (ORF2)</td>
<td>Reverse: 5'-GGCTCGAGATATGTGACACTGAGTAAAGGTCAGCGGACCAGAT-3'</td>
</tr>
<tr>
<td>A/I to M/M mutant (MT)</td>
<td>Reverse: 5'-GGCTCGAGATATGTGACACTGAGTAAAGGTCAGCGGACCAGAT-3'</td>
</tr>
<tr>
<td>A/V to M/M mutant (helicase-replicase)</td>
<td>Reverse: 5'-GGCTCGAGATATGTGACACTGAGTAAAGGTCAGCGGACCAGAT-3'</td>
</tr>
<tr>
<td>Linker</td>
<td>Forward: 5'-AACGACGATATGTGACACTGAGTAAAGGTCAGCGGACCAGAT-3'</td>
</tr>
</tbody>
</table>

Table 2. Primer sequences used in cloning
protease was checked using gelatin zymography (Fig. 1a). Briefly, 1% agar and 0.5% gelatin (Sigma-Aldrich) were prepared in 0.1 M Tris/HCl (pH 7.2) and boiled in a microwave oven. The mix was then poured over a glass slide and allowed to cool. The refolded protease was added to holes punched in the gel and allowed to diffuse overnight in a humidified chamber at room temperature. The gel was then fixed with 60% methanol, 7% acetic acid in water followed by Amido black (Sigma-Aldrich) staining and destaining. Dialysis buffer was used as a negative control in gelatin zymography (Fig. 1c). Similar purification and refolding were performed from the lysate of untransformed E. coli C43 cells and this was also used as a control in zymography to rule out the presence of any contaminating host protease (Fig. 1b).

**Construction of ORF1 fragments.** The ORF1 regions covering MT to PCP (27–1857 nt) and helicase to replicase (2833–5109 nt) were PCR-amplified using specific primers (Table 2) and HEV full-length cDNA as a template (GenBank accession no. FJ457024). The regions 27–1857 nt and 2833–5109 nt were then cloned into the pcDNA3 vector (Invitrogen) using EcoRI/Xhol restriction sites. The pcDNA3-MT-PCP and pcDNA3-Helicase-Replicase proteins were expressed and radiolabelled with [35S]methionine-cysteine (BRT) using an in vitro rabbit reticulocyte Transcription and Translation kit (Promega) as per the manufacturer’s instructions. The MT region from 27–1293 nt was also PCR-amplified from HEV full-length cDNA and cloned into the pRSET B expression vector using EcoRI/Xhol restriction sites.

**Protease digestion assays.** The expressed MT-PCP, helicase-replicase and capsid protein (pORF2) were used as substrate proteases in digestion assays (Fig. 2a–d). The pORF2 capsid protein was used and purified using the pRSET A-ORF2 construct as described previously (Kapur et al., 2012).

In digestion assays, 10 ng recombinant HEV protease was incubated with 500 ng pORF2 in 0.1 M Tris/HCl buffer (pH 7.2) in a 50 μL reaction for varying time intervals at 37 °C. Similar, the 100 μL transcription and translation reaction was divided into three equal aliquots: one served as a control for monitoring the auto-protolytic activity of protein substrates (Fig. 2a, lane 1; Fig. 2b, lanes 1 and 2); to the other aliquots, 10 ng functional HEV protease was added in 0.1 M Tris/HCl (pH 7.2) and incubated at 37 °C for 1 h (Fig. 2a, lane 2; Fig. 2b, lane 3) and 2 h (Fig. 2a, lane 3; Fig. 2b, lane 4). After the proteolytic digestion, rabbit anti-MT and anti-replicase antibodies were used to immunoprecipitate the specific proteins from the digestion reaction using Protein-G Sepharose beads (GE healthcare) as described previously (Kapur et al., 2012). These immunoprecipitated proteins were resolved by 12% SDS-PAGE. The gel was vacuum dried and fluorography was performed (Ansari et al., 2000). The rabbit anti-MT and anti-replicase antibodies were generated in the laboratory using purified recombinant MT and replicase proteins as described previously (Ansari et al., 2000).

**MS analysis of digested capsid protein.** The desired protein bands (76 and ~56 kDa capsid proteins) were excised from the gel and samples for MS analysis were prepared as described by Mukherjee & Roy (2013). MS analysis was performed in a QSTAR XL MS/MS system (Applied Biosystems) and the output results were analysed using MASCOT software (Fig. 3a).

**Amino (N)-terminal sequencing.** Unlabelled MT-PCP and helicase-replicase proteins were digested, immunopurified and PAGE-resolved as mentioned above. The proteins were then blotted onto PVDF membrane (Millipore) in CAPS (Sigma-Aldrich) buffer and samples for N-terminal sequencing were prepared as described by Bass et al. (2000) and loaded into the reaction cartridge in a Procise Protein Sequencer (Applied Biosystems). The sequence obtained in the output file was blasted to the PDB to find matched hits.

**Site-directed mutagenesis of cleavage sites.** The cleavage sites of MT (alanine15/isoleucine16), replicase (alanine1364/valine1365) and pORF2 (leucine197/valine198) were mutated to methionine/methionine in the pRSET B- MT (27–1293), pRSET B-Helicase-Replicase (2833–5109) and pRSET A-ORF2 (5147–7129) constructs using the Gene Art Site-directed mutagenesis system (Invitrogen) as per the manufacturer’s instructions. The primer sequences are given in Table 2. The resultant mutant constructs were confirmed by sequencing using the dyeoxy chain-termination method in an ABI 310 sequencer (Applied Biosystems). The 51 kDa MT protein (47 kDa MT N-terminal His tag) was expressed from the pRSET B-MT construct. The mutant proteins were used as a protease substrate in digestion assays (described above) to validate the specificity of cleavage (Fig. 4). The digestion reaction with wild-type proteins served as a positive control.

**Analysis of recombinant HEV protease inhibition.** To study the inhibition profile, 10 ng HEV protease was pre-incubated with 1× protease inhibitors AEBSF [4-(2-aminoethyl)-benzenesulfonyl-fluoride hydrochloride], ALLN (N-acetyl-Leu-Leu-Nle-CHO), antipain, aprotinin, bestatin, chymostatin, sodium EDTA, E-64 [N-[l-1 trans-carboxyamine-2-carboxy]-l-leucyl]-agmatine], leupeptin, pepstatin, phosphoramid and PMSF in independent 30 μL reactions containing 0.1 M Tris/HCl buffer (pH 7.2) (Protease inhibitor set; G Biosciences) for 1 h at 37 °C, following which 500 ng pORF2 was added to each reaction as a specific substrate (Fig. 5, lanes 3–14). A protease-negative control with 500 ng pORF2 was kept in parallel under the same conditions to monitor pORF2 self-degradation (Fig. 5, lane 1). These reactions were incubated for 5 h at 37 °C for over-digestion. All the reactions and controls were resolved by 12% SDS-PAGE (Fig. 5). The inhibition of activity of HEV protease was visualized in the presence of different inhibitors.

**Confirmation of specificity of HEV protease by quenched fluorogenic peptide digestion.** A synthetic, internally quenched, fluorogenic peptide, N-succinyl-Leu-Leu-Val-Tyr-7-AMC (Sigma-Aldrich), known to be specific for chymotrypsin-like proteases, was used as a substrate for HEV protease. The peptide was dissolved in a GLOMAX MULTI JR detection system (Promega) at 460 nm with excitation at 380 nm. Cleavage in the peptide sequence results in a fluorescence signal. The rabbit anti-MT and helicase-replicase antibodies were used to immunoprecipitate the specific proteins from the digestion reaction using Protein-G Sepharose beads (GE healthcare) as described previously (Kapur et al., 2012). These immunoprecipitated proteins were resolved by 12% SDS-PAGE. The gel was vacuum dried and fluorography was performed (Ansari et al., 2000). The rabbit anti-MT and anti-replicase antibodies were generated in the laboratory using purified recombinant MT and replicase proteins as described previously (Ansari et al., 2000).

**Identification of possible active site amino acids of protease by mutagenesis.** The conserved serine (S454, S468, S505, S512, S515 and S518), cysteine (C457, C459, C471, C472, C481 and C483) and histidine (H443 and H590) residues were mutated to alanine in independent pRSET- Protease constructs using the Gene Art site-directed mutagenesis system (Invitrogen) as per the manufacturer’s instructions (primer sequences in Table 2). The mutations in the constructs were confirmed by sequencing using the dyeoxy chain-termination method in an ABI 310 sequencer (Applied Biosystems). These mutant proteins were expressed and purified from E. coli C43 cells, and refolded as described for the wild-type protease. A digestion assay was performed to screen these protease mutants using pORF2 as a substrate (as described above) (Fig. 6).
ACKNOWLEDGEMENTS

This study was funded by a grant in aid project from the Department of Biotechnology, Ministry of science and Technology, Government of India (BT/PR13220/Med/29/156/2009) to S.K.P. In addition, S.K.P. was supported by the Department of Science and Technology, Ministry of Science and Technology, Government of India, in the form of a J. C. Bose fellowship. D. P. was the recipient of a fellowship from the Council of Scientific and Industrial Research (CSIR), Govt of India. We sincerely thank Professor Dinakar M. Salunke and Mr Atin (Regional centre of Biotechnology, India) for conducting aminoterminal protein sequencing.

REFERENCES


