Structural features stabilized by divalent cation coordination within hepatitis E virus ORF1 are critical for viral replication

Robert LeDesma[1], Abhishek Biswas[1], Stephanie Maya[1], Stefania Gili[2], John Higgins[2], Alexander Ploss[1,*]

[1]Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, Princeton, NJ 08544, USA.

[2]Department of Geosciences, Princeton University, Princeton, NJ, 08544, USA.

*Correspondence should be addressed to A.P.: aploss@princeton.edu; +1 (609) 258-7128

Keywords: viral hepatitis; Hepatitis E; Hepatitis E virus, ORF1, viral replication; structural virology

Abstract:

Hepatitis E virus (HEV) is an RNA virus responsible for over 20 million infections annually. HEV's open reading frame (ORF)1 polyprotein is essential for genome replication, though it is unknown how the different subdomains function within a structural context. Our data show that ORF1 operates as a multifunctional protein, which is not subject to proteolytic processing. Supporting this model, scanning mutagenesis performed on the putative papain-like cysteine protease (pPCP) domain revealed six cysteines essential for viral replication. Our data are consistent with their role in divalent metal ion coordination, which governs local and interdomain interactions that are critical for the overall structure of ORF1; further, the "pPCP" domain can only rescue viral genome replication *in trans* when expressed in the context of the full-length ORF1 protein but not as an individual subdomain. Taken together our work provides a comprehensive model of the structure and function of HEV ORF1.

Significance Statement: The development of non-teratogenic and potent antiviral therapies against HEV have been hindered by an incomplete understanding of the viral replication cycle. Our work provides a mechanistic insight into the complex replicative cycle of this understudied human pathogen and identifies a novel domain-domain interaction that is vital for replicative fitness.

Introduction:

Hepatitis E has a global disease burden of over 20 million cases per annum, leading to approximately 70,000 fatalities (Rein et al., 2012). This burden is especially pronounced in immunocompromised individuals and pregnant women, the latter of whom experience a close to 30% mortality rate in the third trimester (Khuroo & Kamili, 2003) and/or approximately 3,000 stillbirths (Rein et al., 2012). HEV infection can be prevented with a prophylactic vaccine which is currently only licensed in China. Presently, treatment is limited to ribavirin (RBV) and pegylated type I interferon (IFN) (Kamar et al., 2014). However, these therapies are plagued with considerable side effects, as RBV is teratogenic and thus cannot be administered during pregnancy, and IFN therapy can lead to transplant rejection in organ transplant recipients (Haagsma et al., 2010). Furthermore, HEV strains with fitness-enhancing mutations have been identified in patients showing clinical resistance to ribavirin treatment (Todt et al., 2018). One clinical case study has suggested that sofosbuvir, a drug approved for hepatitis C virus (HCV) treatment, may have a beneficial additive effect when used in combination with ribavirin, however other studies have observed no therapeutic benefit, and the use of this drug for HEV treatment remains controversial (van der Valk et al., 2017; van Wezel et al., 2019). Other clinical drugs are currently not approved for the treatment of hepatitis E, obviating the need for direct acting antivirals and a better understanding of the viral replication cycle.

HEV is a (+) ssRNA virus in the *Orthohepevirus* genus in the *Hepeviridae* family of viruses (Smith et al., 2014). The genome of HEV has a 5'-methylated cap and a 3'-poly (A) tail and is comprised of three partially overlapping open reading frames (ORFs). ORF1 encodes the viral replicase (Koonin et al., 1992), ORF2 encodes the capsid protein (Reyes et al., 1993; Tam et al., 1991), and ORF3 encodes a viroporin necessary for viral egress (Ding et al., 2017). While the primary functions of ORF2 and ORF3 have been characterized, much of ORF1 remains to be understood. ORF1 has been organized into seven domains based on prior bioinformatic analysis (**Figure 1A)** (Koonin et al., 1992), only four of which have been functionally characterized in some detail (namely the methyltransferase (Magden et al., 2001), macrodomain (Parvez, 2015), helicase (Devhare et al., 2014; Karpe & Lole, 2010), and RNA-dependent RNA polymerase (RdRp) (Koonin, 1991; van der Heijden & Bol, 2002)). Of the

remaining domains of HEV ORF1, the putative papain-like cysteine protease (pPCP) has been the subject of some debate; several groups assessing the functionality of this region have disagreed on the presence of protease activity (Ansari et al., 2000; Kanade et al., 2018; Karpe & Lole, 2011; Paliwal et al., 2014; Parvez, 2013; Parvez & Khan, 2014; Perttilä et al., 2013; Ropp et al., 2000; Sehgal et al., 2006; Suppiah et al., 2011) and if so, whether this region has viral protein or host cellular protein targets. Evidence supporting and refuting these activities have been building for almost three decades, however the possibility remains that this region may exert orthogonal activities.

Full characterization of ORF1's functions as well as HEV's full replication cycle have been hampered by a dearth of structural information of the ORF1 protein. Though two structures of small regions with ORF1 have been recently obtained - one being the amphipathic "thumb" of the RNA dependent RNA polymerase (RdRp) (Oechslin et al., 2022), the other being an intra-ORF1 region spanning portions of the putative PCP and hyper-variable region (HVR) (Proudfoot et al., 2019) -, neither are functional outside of the context of the ORF1 protein, limiting our understanding of how uncharacterized regions of ORF1 fold and operate. Without structural information readily available, *in silico* analyses have been previously attempted to glean information of ORF1's functions, with limited success. For instance, bioinformatic analysis of the HEV pPCP of genotype 1 SAR55 strain predicted three disulfide bridges and a putative zinc-binding motif (Parvez & Khan, 2014; Saraswat et al., 2019), though previous computational approaches have been lacking in power and iterative ability. With the advent of AlphaFold (Jumper et al., 2021; Tunyasuvunakool et al., 2021), research into protein structure has entered a new era where testable hypotheses can be generated in an iterative, high-throughput manner.

To understand how the putative PCP - and ORF1 more broadly - operate, we opted to combine biochemical, genetic, mass spectrometric, and computational approaches. We identified amino acid motifs within the pPCP that are vital for viral replication via an unbiased triplet alanine scanning mutagenesis, as well as characterized the necessity or dispensability of eight conserved cysteines within the putative PCP domain. Of these eight, six were identified as indispensable, forming a hexa-cysteine motif commonly seen in host metal-binding proteins. We established a transcomplementation system to demonstrate that the putative PCP is only functional within the context of the full length ORF1 protein. We have been able to validate A.I. driven protein structure prediction programs with testable genetic and biochemical data; using the AlphaFold algorithm(Jumper et al., 2021), we determined the replicative capacity lost via site-directed mutagenesis to not be due to a deficiency of proteolytic activity, but rather a loss in structural integrity due to an inability to bind divalent metal ions. Moreover, we identified a novel interdomain divalent metal ion binding interaction between the pPCP and the upstream uncharacterized Y-domain of HEV ORF1. Further, utilizing a tolerable epitope locus within ORF1's hyper-variable region (HVR) (Szkolnicka et al., 2019), we were able to purify ORF1 protein for downstream inductively-coupled plasma mass spectrometry (ICP-MS) analysis, and discovered that point mutants in either the pPCP or Y-domain differed in divalent ion binding activity. Taken together, our work demonstrates that HEV ORF1 likely functions as one large multidomain protein that does not undergo processing, and that the putative catalytic residues predicted by prior bioinformatic analyses are actually structural in nature via their ability to bind divalent metal ions.

Results:

Mutations within the HEV putative protease domain render the virus replication incompetent. The functional domains and genome organization of HEV **(Figure 1A)** were first

suggested based on bioinformatic alignments of the genotype 1 Burma reference strain with other well-characterized viruses in 1992 (Koonin et al., 1992), and the putative protease domain was proposed based on limited sequence identity with the distantly related rubella virus (RUBV). These analyses suggested the existence of a putative papain-like protease within HEV ORF1, and that the proposed catalytic dyad were residues C483 (which is highly conserved across all known HEV genotypes) and H590 -which is variable across all 8 known HEV genotypes (**Supplementary Figure 1).** Since, tools have been developed to systematically interrogate the HEV genome, such as the development of infectious clones of cell-culture adapted strains (Johne et al., 2014; Shukla et al., 2012; Shukla et al., 2011), and reporter replicons utilizing green fluorescent protein (GFP) (Emerson et al., 2004) or Gaussia luciferase (Gluc) (Graff et al., 2005) **(Figure 1A)**.

To determine the importance of the residues that have been proposed as the putative catalytic dyad, we mutated C483 and/or Y590 to chemically similar amino acids, alanine, or in the case of Y590, residues found in other genotypes of HEV. Huh7.5 human hepatoma cells were transfected with *in vitro* transcribed RNA from a recombinant version of an HEV genome derived from the KernowC1/p6 strain (Shukla et al., 2011) in which ORF2 and ORF3 are replaced by a secreted version of Gaussia luciferase (Gluc) (Shukla et al., 2011), termed Kc1/p6 Gluc **(Figure 1A)**. Gluc activity as a measure for the efficiency of RNA dependent RNA polymerase-mediated viral replication was quantified in the culture supernatants over 4 days post RNA transfection (4 d.p.t). Transfection of the wild type (wt) Kc1/p6 Gluc into naïve Huh7.5 human hepatoma cells led to a ca. 34,000-fold increase in luminescence over mock cells. Transfection of a polymerase deficient genome harboring a mutation in the highly conserved GDD motif of the RNA-dependent RNA polymerase (deemed pol (-)) expectedly did not augment Gluc activity (**Figure 1 B-C**). Notably, genomes harboring mutations in the C483 and/or Y590 positions were incapable of establishing stable replication (**Figure 1 B-C**).

We next sought to understand if this lack of viral replication brought on by mutating the highly conserved C483 was unique to the Kc1/p6 cell culture adapted strain of HEV, or if it translated to other known human-tropic HEV strains. Thus, we mutated the C483 residues in the Gluc reporter genome configurations of HEV strains SAR55 (genotype 1), SHEV3 (genotype 3), and TW6196E (genotype 4) (Ding, Nimgaonkar, et al., 2018), and transfected *in vitro* transcribed RNA into HepG2C3A human hepatoma cells similar to **(Figure 1 B-C)**. In line with our observations using Kc1/p6 replicon, HEV RNA replication was severely impaired in SAR55, SHEV3 and TW6196E genomes harboring the C483A mutation (**Figure 1 D-G)**. Notably, Gaussia luciferase levels were equivalently as low as the pol (-) versions of the reporter replicons.

To determine if this deficiency was due to a disruption in RNA folding, we mutagenized C483 into each available codon for cysteine and alanine **(Figure 1H).** These analyses demonstrated that encoding the alternate cysteine had a mild negative affect on viral replication efficiency, while any alanine codon usage brought replication levels down to those of the Pol (-) mutant **(Figure 1H),** suggesting that the deficiency primarily lies in protein folding or function. Collectively, these data demonstrate the necessity of these residues for viral fitness, despite Y590 being heterogeneous across HEV viral genotypes.

HEV ORF1 putative PCP cannot function outside of the context of the full-length protein, and C483A replication deficiency is rescuable *in trans*. To probe further the mechanism underlying the functional impairments of the C483 mutants, we devised an experimental system in which HEV RNA replication is uncoupled from protein translation. Following our previously

established successful transcomplementation approach for studying HEV ORF3's viroporin function (Ding et al., 2017) and cis-regulatory elements responsible for regulating transcription of the subgenomic RNA (Ding, Nimgaonkar, et al., 2018), we lentivirally expressed a wild-type (wt), pol (-), or C483A version of ORF1, or the pPCP alone in HepG2C3A human hepatoma cells (**Figure 2A**). These cells were subsequently transfected with *in vitro* transcribed RNA from Kernow C1/p6 Gluc wt, pol (-), or C483A genome (**Figure 2A**). Gluc activity as a measure of the efficiency for HEV replication was quantified in the culture supernatants over 4 days post RNA transfection. Notably, when the mock signal fold change over WT luciferase signal was examined, cells expressing a mutant form of ORF1 demonstrate a deleterious effect on WT replicon replication, likely due to competitive inhibitory binding of the mutant protein with the replicon RNA **(Figure 2B)**. Impairments in viral genome replication due to the pol (-) or C483A mutations could be rescued *in trans* by expression of WT ORF1 near to levels of those following transfection of WT replicon RNA (**Figure 2 D-E)**. Of note, expression of the putative PCP failed to restore replication of Kc1/p6 Gluc C483A suggesting that the functions of this region of ORF1 are not adequately maintained outside of the context of the ORF1 polyprotein. This transcomplementation platform provides further means to uncouple the putative functions of the pPCP, e.g. polyprotein processing or modulation of the host cellular environment, from viral genome replication.

Point mutations of highly conserved cysteines and alanine scanning mutagenesis within the putative PCP identifies residues and regions indispensable for viral replication. We examined the putative PCP sequences across all 8 known HEV genotypes **(Supplementary Figure 1)**, and noticed an octa-cysteine motif highly conserved across all HEV genotypes **(Figure 3A)**. Previous work has shown the core hexa-cysteine motif encompassing cysteines 457-483 in SAR55 genotype 1 HEV were necessary for viral replication (Parvez, 2013). To determine the necessity of each cysteine to the viral replication cycle in the Kc1/p6 strain of HEV, we mutated each in turn to alanine within the Kc1/p6 Gluc reporter replicon and quantified the luciferase signal 4 d.p.t. **(Figure 3B)**. We noticed that of the 8 conserved cysteines, only the core 6 that form a $CxC[x_{11}]CC[x_8]CxC$ hexa-cysteine motif are vital for viral replication, with the first cysteine at position C434 being completely dispensable for replication, and the final cysteine at position C563 having a slight detriment to replication when mutated to alanine **(Figure 3B)**.

To determine more broadly which other regions of the HEV pPCP are indispensable for viral replication, we sought to conduct an unbiased genetic mutagenesis screen of the entire putative PCP region of 160 amino acids. Site-directed triplicate alanine scanning mutagenesis was conducted across the entirety of the HEV pPCP, identifying several triplicates that offer pro-viral activity, as well as identifying the majority of triplicates vital for viral replicative fitness **(Figure 3C).** Notably, triplicates containing any of the conserved cysteines, as well as the variable amino acid at position 590 were indispensable for viral replication, despite point mutations at positions C434 and C563 being tolerated **(Figure 3B)**. Additionally, the amino acid stretches 55-63 and 73-81 directly downstream of the hexa-cysteine motif tolerate mutagenesis quite well, suggesting a possible structural/linker function of these amino acids; these triplicates are not highly conserved across the eight HEV genotypes **(Supplementary Figure 1).**

Hexa-cysteine motif ($CxC[x]_{11}CC[x]_8CxC$) within HEV Kernow putative PCP vital for viral replication shares homology with host divalent metal ion binding proteins. Interrogating the results of point mutations to the 8 highly conserved cysteines within the putative PCP led us to further bioinformatic analysis to determine the function of the vital hexa-cysteine motif. Utilizing motif searches with ScanProSite (de Castro et al., 2006) and TrEMBLE(Bairoch &

Apweiler, 2000; Boeckmann et al., 2003; O'Donovan et al., 2002), we began by searching for a relaxed expression of the HEV hexa-cysteine motif ($CxC[x]_{3-20}CC[x]_{3-20}CxC$, where x can be any amino acid). From this analysis over 33,000 proteins emerged. To further refine our approach, we identified proteins that matched the HEV CxC motif exactly ($CxC[x]_{11}CC[x]_{8}CxC$, hereafter referred to as the HEV motif); however, all of the protein hits that emerged are as of now uncharacterized, offering little insight as to the function of this motif **(Supplementary Figure 2A)**. Relaxing the criteria to (+/-) 1 for each of the stretches of [x] ($CxC[x]_{(10-12)}CC[x]_{(7-9)}CxC$) brought forth 26 proteins with known functions **(Supplementary Figure 2A)**, enriched for proteins with divalent metal ion binding activity **(Supplementary Figure 2B).** We hypothesized that this region within the pPCP is necessary for metal ion coordination.

Structural prediction models of HEV pPCP demonstrates low-confidence scores, suggesting lack of highly ordered secondary structure. The dearth of structural information of the HEV ORF1 protein has hampered the complete understanding of the viral replication cycle. To glean more information about the domain organization and protein folding of HEV ORF1, we turned to AlphaFold(Jumper et al., 2021) to predict the structure of ORF1. The complete sequence of HEV ORF1 was fed into the AlphaFold algorithm, and the best ranked model (**Figure 4A**) was chosen for further analysis. To gain confidence in the best ranked model, we analyzed the AlphaFold prediction in several ways. First, we analyzed the confidence levels produced by AlphaFold (pLDDT score) for each residue across the ORF1 structure prediction (**Figure 4 B-C**). This analysis revealed varying levels of confidence throughout the entirety of ORF1, and importantly, low confidence throughout much of the pPCP. We further looked at how AlphaFold predicts the pPCP outside of the context of ORF1 and found that the pLDDT averages are very similar (**Figure 4C**), with the pPCP alone averaging a pLDDT score of 65.92, and the pPCP within the context of ORF1 scoring a slightly better average of 66.05. Second, we tested how closely the AlphaFold prediction aligned with two separate solved structures of fragments of HEV ORF1, as well as the known solved structures of the macro domains in other distantly related viruses. To this end, we used a combinatorial approach of sequence based alignments with structure based alignments to gain the most accurate distance matrices of relevant atomic coordinates within each alignment of regions of ORF1 and the corresponding known structures (approach reviewed in(Carpentier & Chomilier, 2019)). To accomplish this, we took the ORF1 structure prediction, and using the tool TM-Align(Zhang & Skolnick, 2005) we aligned: a region of HEV ORF1 that spans parts of the putative PCP and HVR (AAs 510-691) of the SAR55 strain of genotype 1 HEV (PDB: 6NU9)(Proudfoot et al., 2019), the amphipathic “thumb” of HEV genotype 3 strain 83-2-27 RNA dependent RNA Polymerase (amino acids 1628-1647; ORF1 kc1/p6 residues 1684-1709)(Oechslin et al., 2022), and the macro domains of Sindbis virus (SINV amino acids 1342 – 1509 of PDB: 4GUA)(Shin et al., 2012), and Chikungunya virus (CHIKV) (PDB: 3GPG)(Malet et al., 2009), (**Supplementary Figure 3A-C,** respectively).

Alignment of 6NU9 with the corresponding region of Kc1/p6 ORF1 shows high local alignment identity, with the average distance of this superimposition aligning at 0.67 Å **(Supplementary Figure 3A)**. Alignment of the amphipathic RdRp “thumb” domain similarly shows incredibly high local alignment identity with a 0.48 Å average differential (**Supplementary Figure 3B**). Understandably, alignment of the macrodomains of SINV and CHIKV also show high local alignment, though less robust than that of other HEV strains to Kc1/p6 ORF1 of HEV. Notably, the HEV macro domain and the macro domains of SINV and CHIKV share little amino acid similarity and identity, with the Kc1/p6 HEV macro domain sharing only 36.45% sequence identity and 54.21% sequence similarity with the SINV macro domain, and sharing only 35.58% sequence identity and 56.73% sequence similarity with the CHIVK macrodomain (**Supplementary Figure 4**). Despite this sequence disparity, these viral macrodomains share

high structural identity, with the ORF1 and SINV superimposition aligning at 1.79 Å on average, and similarly, aligning with the CHIKV superimposition at 1.89 Å on average (**Supplementary Figure 3C**).

Importantly, these aforementioned highly ordered secondary structures within HEV ORF1 all exhibit much higher AlphaFold pLDDT scores on average than the whole of the pPCP (**Figure 4C**). This prompts several considerations: 1.) AlphaFold is not able to accurately predict the amino acid backbone or side chains of the pPCP with a high amount of confidence. 2.) If this region were a protease domain, it would likely have a highly ordered, and readily predicted secondary structure. To test this latter point, we analyzed AlphaFold's ability to accurately predict the structure of a known protease, which it does very well. When comparing the known structure of the hepatitis A virus (HAV) 3C protease(Bergmann et al., 1999) with the AlphaFold prediction, they align with an average superimposition distance of 1.03 Å (**Supplementary Figure 3D**). Interestingly, AlphaFold was able to fold the HAV protease with very high confidence, with an average pLDDT score of 97.14 across the entirety of prediction, giving high confidence of interpretation to both the amino acid backbones and side chain chains (**Supplementary Figure 3D**). Due to the low pLDDT score of the pPCP, we visualized the predicted folding of this region by color-coding the outcomes of the alanine-scanning mutagenesis to better understand whether the secondary structures predicted by AlphaFold stand. We found that the majority of the predicted alpha-helices and beta-sheets do not tolerate mutations well, while the majority of the tolerated mutants lie within predicted regions of disorder (**Figure 4D**), (**Supplementary Movie 1**). Taken together, these results suggest the structure predictions of HEV ORF1 by AlphaFold demonstrate high confidence within known regions that possess a high level of secondary structure. AlphaFold is able to predict the structure of a known viral protease with high confidence, suggesting further that the HEV pPCP domain does not possess the necessary secondary structure of a protease. This prediction of the HEV ORF1 pPCP is bolstered by the enrichment of replication tolerant triplets from the alanine scanning mutagenesis to preferentially localize to predicted areas of disorder **(Figure 4D**), (**Supplementary Movie 1**), suggesting as a whole that the pPCP does not fold as a protease with a catalytic pocket.

Structural prediction models suggest that mutating cysteines within putative PCP disrupts divalent ion coordination pockets and novel domain-domain interaction with upstream Y-domain. Upon demonstrating that mutations within the putative PCP domain prevent HEV from replicating, we sought to further elucidate potential mechanisms by which this deficiency occurs. Utilizing the structural predictions of HEV ORF1 generated with AlphaFold, we began by analyzing the predicted folding structure of WT ORF1, and two of the point mutants within the putative PCP: C483A and C563A (chosen based on the heterogeneity of their phenotypes). C483A is fully replication deficient, while C563A is blunted in replication at two orders of magnitude lower than WT four days post transfection (**Figure 3B**).

We next fed the sequences of the C483A and C563A mutant sequences of ORF1 into the AlphaFold algorithm. Upon inspection of the best ranked models for each of these versions of ORF1 (WT, C483A, C563A), we noticed a novel pseudo-zinc-finger formed by the amino acids underlying the HEV hexa-cysteine motif, despite the low pLDDT scores within this region (**Figure 5A**, left) (**Supplementary Movies 2-4, respectively**). Mutation of C483A, but not C563A is predicted to disrupt the molecular architecture of this pseudo-zinc-finger, however mutation of either causes a relaxation of several predicted bond lengths to beyond biological relevance (Zheng et al., 2008) (**Supplementary Table 2, Supplementary Movies 2-3**). Cysteine side chain residues at positions 457 and 459 at the end of the predicted beta sheet leading into the pseudo-zinc-finger were shown to be projecting into inter-domain space. These

residues are predicted to form a potential tetrahedral divalent ion binding pocket (Laitaoja et al., 2013) with D248 and H249 in the upstream Y-domain (**Figure 5A**, middle), which had pLDDT scores of 82.99 and 87.67, respectively, giving high confidence in the predicted location of the amino acid backbones and potentially their side-chains (**Figure 4C**). Cysteine, histidine, and aspartic acid residues are well known to bind divalent metal ions and form tetrahedral geometry (Laitaoja et al., 2013; Zheng et al., 2008). To further test whether this predicted interaction between C457, C459, D248, and H249 was vital to viral replication, we mutated either D248 or H249 to alanine in the Kc1/p6 Gluc replicon and quantified Gluc expression 4 days post RNA transfection. D248 was dispensable for viral replication while H249 is not (**Supplementary Figure 5**); D248 is variable across all 8 known HEV genotypes while H249 is highly conserved (**Figure 5B**). This observation led us to inquire as to what the disruptions these point mutations could have on ORF1 globally. We noticed that by comparing the structure of ORF1 WT (**Supplementary Movie 2**) next to the mutants while highlighting domains of ORF1 with well-defined functions such as the methyltransferase, helicase, and RdRp, structural differences emerge. Mutating C483A, C563A, or H249A are predicted to cause regions of the methyltransferase, helicase, and RdRp to reconfigure (**Figure 5A**, right) (**Supplementary Movies 3, 4, and 6, respectively**). Further, mutating C563A, D248A, or H249A causes a predicted membrane association domain that is exposed in the WT ORF1 protein to become buried, possibly preventing the association with intracellular membranes and preventing/hindering the formation of a replication compartment(Metzger et al., 2022; Szkolnicka et al., 2019) (**Supplementary Movies 4, 5, and 6, respectively**). Taken together, these results suggest that by interfering with divalent metal ion binding domains within the putative PCP, structural domains vital to viral replication form aberrantly, and prevent HEV ORF1 from efficiently replicating.

HEV hexa-cysteine motif coordinates biologically relevant divalent metal ions. Outside of our own analysis that suggests the putative PCP of HEV ORF1 has metal ion binding activity, other groups have identified regions within ORF1 predicted to harbor Ca^{2+} and Zn^{2+} ion binding sites (Parvez & Khan, 2014; Proudfoot et al., 2019). However, the low abundance of ORF1 protein within infected cells and the lack of well-characterized ORF1 specific antibodies has hampered attempts at purifying a replication competent ORF1 protein. This barrier has only recently been overcome when tolerant epitope tag insertion sites were discovered within the HVR region of ORF1 and characterized(Metzger et al., 2022; Szkolnicka et al., 2019).

Utilizing an HA-tag insertion site flanked on either side by the linker sequence (AAAPG-HA-AAPG, hereafter referred to as HA-tagged) within the HVR of ORF1(Szkolnicka et al., 2019), we generated an overexpression system of WT or mutant HA-tagged ORF1 via lentiviral transduction of Huh7 human hepatoma cells (**Figure 6A**). To determine if the mutations at C483A, C563A, D248A or H249A affected ORF1's ability to bind divalent ions, we immune purified each protein via the HA-tag mediated immunopurification (**Figure 6B**), and subjected each purified protein eluant to inductively-coupled plasma mass spectrometry (ICP-MS) (**Figure 6C**). Element signatures indicate that while each mutant is heterogenous with each other in binding activity for most divalent ions, they all bind less Zn^{2+} than WT. These results suggest that divalent ion binding capacity likely affects proper structural confirmation of ORF1 leading to differential replication capacity based on mutation position.

Loss of divalent ion binding activity leads to differences in subcellular localization of HEV ORF1. To determine whether the differential in divalent ion binding potential due to mutations within the putative PCP or upstream Y-domain is responsible for a loss of subcellular localization, we turned to confocal microscopy. Cells bicistronically expressing zsGreen as a

marker of transduction as well as WT ORF1 without an epitope tag, WT ORF1-HA-tagged, C483A-HA-tagged, C563A-HA-tagged, D248A-HA-tagged, or H249A-HA-tagged were imaged for zsGreen, and using antibodies against the HA-tag and nuclei were imaged to visualize ORF1 subcellular localization patterning. Cells expressing ORF1 WT-HA showed significant expression of ORF1 in both the cytoplasm and nucleus as previously reported(Metzger et al., 2022), and showed many puncta aggregates throughout the cytoplasm (**Figure 7**). In contrast, the replication deficient ORF1 C483A-HA expressing cells lost the ability for ORF1 to localize to the nucleus and was found dispersed throughout the cytoplasm (**Figure 7**). Cells expressing the C563A-HA mutant of ORF1 showed similar localization patterns to ORF1 WT-HA, with puncta forming in the cytoplasm, and maintaining the ability to localize to the nucleus (**Figure 7**).

Cells expressing mutations in the novel upstream interacting Y-domain were also varied in their localization when compared to ORF1 WT-HA. Replication competent ORF1 D248A-HA, like C563A-HA, shared localization patterns with ORF1 WT-HA, forming cytoplasmic puncta and localizing within the nucleus (**Figure 7**). In contrast, replication deficient ORF1 H249A-HA showed very disperse cytoplasmic localization, no aggregate formation, and a lesser ability to localize to the nucleus (**Figure 7**). Taken together with the replication data of these epitope-tagged ORF1 mutants (**Supplementary Figure 5**), localization of ORF1 seems to correlate with replicative capacity, with nuclear localization being lost or diminished in mutants that cannot replicate (C483A-HA, H249A-HA), and mutants able to establish replication (C563A-HA, D248A-HA) share localization patterns with ORF1 WT-HA. These results shed light on a potential novel mechanism by which mutants in the putative PCP and upstream Y-domain interfere with HEV's replication cycle.

Discussion:

With the advent of reporter replicons for the HEV replicase in the early 2000s(Emerson et al., 2004), research into the functional domains of ORF1 become possible. Perturbations to the viral replicase could be assessed qualitatively and be quantified for the first time, allowing researchers to begin dissecting regions of ORF1 necessary to viral replication. In our study, we first sought to determine whether our results with the Kc1/p6 genotype 3 HEV were in agreement with previous results that utilized a GFP reporter replicon of HEV genotype 1 SAR55 strain in S10-3 cells(Parvez, 2013). We were able to demonstrate that mutations in the core six of the eight highly conserved cysteines within the pPCP, as well as the heterogenous residue at position 590, renders Kc1/p6 HEV replication incompetent. We were also able to show that the putative catalytic cysteine in the HEV pPCP renders HEV in three additional genotypes replication incompetent. Taking this analysis further, we demonstrate that the dysfunction in our reporter replicon is likely at the protein level, due to the replicon being able to tolerate an alternate codon for cysteine, and none of the codons for alanine, suggesting a conserved need of this amino acid residue for HEV.

Of all the domains within ORF1, the functions of the putative PCP remain the most debated. Though evidence for and against proteolytic cleavage continues to mount on both sides, it is important to take the scientific results, as well as the functionality of the ORF1 protein, in context. Our data in this study has shown that the pPCP of ORF1 cannot function outside of the context of the full length protein, which is rather uncommon for many RNA viruses such as hepatitis A virus(Lemon et al., 1991), HCV(Yang et al., 2000), and flaviviruses such as Zika virus(Ding, Gaska, et al., 2018). While most characterized (+) ssRNA viruses rely on proteases to liberate individual gene products from their encoded polyprotein, HEV may be an exception. While it remains conceivable that host proteases may post-translationally process ORF1, there

is rather limited evidence that subunits of ORF1 itself harbors proteolytic activity. Furthermore, if processing were to occur, it is likely that only a small fraction of ORF1 might be cleaved, as suggested previously(Metzger et al., 2022); however, the smaller species of ORF1 in the previously cited study were unable to be characterized by mass spectrometric analysis, leaving the processing of ORF1 still subject to debate. The inability for the putative HEV PCP to act outside of the context of the full length ORF1 protein suggests that it has some orthogonal activity, and that HEV ORF1 likely functions as one large multi-domain protein.

To take an unbiased approach to analyzing the putative PCP domain, we attempted to identify motifs within the region that were either vital or dispensable to viral replication. To this end, we conducted an alanine scanning mutagenesis screen in triplets across the entire viral region, and found that while the majority of the region is needed for replication, there were 11/54 triplets that were able to replicate at near WT levels. Several of these triplets fell very near to the HEV ($CxC[x_{11}]CC[X_8]CxC$) motif within the putative PCP; we aimed to identify the potential function of this region, as well as obtain as much structural information. To this end, we were able to identify proteins with known functions that shared close homology to the HEV motif, and found that these proteins were enriched for metal ion binding or for disulfide bond formation, suggesting a structural activity rather than a catalytic one. We then capitalized on the power of AlphaFold to gain structural insights into the nature of this vital region. When looking at the structure predictions of the HEV motif, we noticed a striking pseudo-zinc finger formation, as well as a tetrahedral binding pocket canonically associated with Zn^{2+} binding with two residues in the upstream Y-domain. When mutations to conserved cysteines or the aspartic acid or histidine in the upstream Y-domain were modeled, several changes noticeable: first, for the two mutants that are rendered replication incompetent (C483A, H249A), the alpha-helix of the pseudo zinc-finger is disrupted. Further, several bond lengths between potential coordinating residues are relaxed to beyond biological relevance. Further, a putative membrane contact site predicted in(Parvez, 2017) becomes buried in these mutants, hinting at a potential mechanism behind the loss of replicative ability. Many RNA viruses are known to adopt a similar strategy of metal ion coordination to carry out necessary functions(Chasapis, 2018). For instance, HCV utilizes a metalloprotein in its replicase, the nonstructural protein NS5A(Tellinghuisen et al., 2004; Tellinghuisen et al., 2005); HCV utilizes four cysteines within a $C[x]_{17}CxC[x]_{20}C$ motif, conserved among *Hepacivirus* and *Pestivirus* genera, for Zn^{2+} coordination and proper function of the HCV replicase machinery.

To determine whether mutants in these predicted metal ion binding motifs actually led to a decrease in ion binding activity, we needed to be able to purify ORF1 protein for subsequent analyses. One of the many difficulties in interrogating HEV ORF1 is its low expression level in infected cells, as well as the lack of a well-characterized commercial antibody(Lenggenhager et al., 2017). However, recent identification of sites within ORF1 that tolerate epitope tags without sacrificing viral replicative capacity have opened up new avenues for researchers to investigate ORF1, and the putative PCP, more robustly and critically. Utilizing one such insertion site, we were able to purify WT and mutant ORF1 proteins and subject them to ICP-MS. We found that across all mutants, none were able to bind zinc ion species as well as the WT ORF1, lending validity to our structural hypotheses generated with AlphaFold. We then determined whether the predicted burying of the putative membrane contact site in the replication deficient mutants affected ORF1 localization within cells expressing the epitope tagged ORF1. Utilizing confocal microscopy, we were able to demonstrate that the C483A mutation loses ORF1 nuclear localization, while the H249A mutation decreases nuclear localization of ORF1, and prevents puncta formation throughout the cytoplasm, which lies in stark contrast to WT ORF1. Taken together, with the advent of new tools such as reporter replicons for HEV, AlphaFold, and tolerable epitope insertion sites discovered within ORF1, our bioinformatic and genetic analyses

have been able to go far beyond previous attempts at predicting functional domains within ORF1: we have been able to demonstrate a powerful, iterative pipeline of testing A.I. driven predictions via hypothesis generation and testing with the tools at our disposal. We have been able to demonstrate a novel domain-domain interaction between the upstream Y-domain and the metal-coordinating structural domain of HEV, previously (and incorrectly called the PCP), and we suggest a change in the accepted nomenclature of this vital, and enigmatic viral region to reflect this function.

Materials and Methods:

Cell lines and cell culture.
HepG2C3A cells (ATCC, CRL-10741) Huh7, and Huh7.5 cells (kindly provided by Charles Rice, The Rockefeller University) were maintained in Dulbecco's modified Eagle medium (DMEM) (Gibco, NY, USA) supplemented with 10% (vol/vol) fetal bovine serum (FBS), 50 IU/ml penicillin and streptomycin, in a humidified 5% (vol/vol) CO_2 incubator at 37°C.

Multiple sequence alignment.
The following hepatitis E virus (HEV) strains were used for sequence alignment: GenBank identifiers [IDs] or accession numbers M73218 (genotype 1a, Burma strain), M74506 (2a, Mex), JQ679013.1 (3,Kernow-C1/p6), AB197673 (4a), AB573435 (5a), AB856243 (6, wbJNN_13), KJ496144 (7, 180C), and KX387866 (8, 48XJ). M73218 (genotype 1a, Burma strain) was used as the reference strain for numbering. Sequence alignments were conducted with the SnapGene software (from Insightful Science; available at snapgene.com) using the multiple sequence alignment tool. HEV genotype alignments were conducted using CLUSTALW alignment algorithms embedded within the software.
Alignment of the HEV macro domain with the macrodomain of SINV (PDB: 4GUA) and CHIKV (PDB: 3GPG) were conducted with the SnapGene software (from Insightful Science; available at snapgene.com) using multiple sequence alignment tool with the local alignment Smith-Waterman algorithm.

Hexa-cysteine motif bioinformatics.
The HEV hexa-cysteine motif $CxC[x]_{11}CC[X]_8CxC$ motif sequence identified was searched using ScanProsite(de Castro et al., 2006) on all UniProtKB/Swiss-Prot (release 2020_02 of 22-Apr-20: 562253 entries), UniProtKB/TrEMBL (release 2020_02 of 22-Apr-20: 0 entries) databases sequences using the regular expression [C-X-C-X(3,20)-C-C-X(3,20)-C-X-C]. The regular expression allows from 3 up to 20 residues in the 2 long stretches of amino acid residues where X can be any amino acid shown by previous bioinformatic analysis and sequence alignments across HEV genotypes **(Supplementary Figure 1A)**. The search produced 429 hits in SwissProt and 32,764 hits in TrEMBL. We further refined the regular expression [C-X-C-X(11)-C-C-X(8)-C-X-C to match the HEV hexa-cysteine motif exactly, and produced 25 protein hits with no known function. We refined this search yet again to include the regular expression [C-X-C-X(10,12)-C-C-X(7,9)-C-X-C, which allows (+/-) one residue in each long stretch of amino acids where X can be any amino acid, and found 26 proteins hits with known functions (**Supplementary Figure 2**) and were used to predict the function of the motif sequence.

ORF1 Protein Structure Predictions.
FASTA files of each species of ORF1 (WT or mutants) were submitted to the AlphaFold algorithm(Jumper et al., 2021) (DeepMind, United Kingdom, v. 2.0.0--model preset=monomer))

(or in the case of the HAV 3C protease, v. 2.1.1--model preset=monomer)) run on the Princeton Research Computing DELLA Cluster at Princeton University. 5 models of each protein prediction were produced, and the best ranked model for each was used for subsequent analysis.

Atomic Distance Calculations.

Atomic distance calculations between the HEV ORF1 structure predications produced via AlphaFold (WT or point mutants) and known structures within ORF1 or distantly related viral macrodomains were done via the structure based alignment tool TM-Align(Zhang & Skolnick, 2005). Briefly, the known crystal structure of a region of the ORF1 putative PCP/HVR (PDB: 6NU9(Proudfoot et al., 2019)) was directly fed into the TM-Align software. The amphipathic "thumb" of the RdRp solved via nuclear magnetic resonance (NMR)(Oechslin et al., 2022), encountered a multi-mapping problem due to its alpha-helical nature and short amino acid sequence when fed directly into TM-Align; to generate the correct distance plot, the AlphaFold ORF1 prediction was trimmed of amino acids not corresponding to the region of the amphipathic RdRp thumb domain (Kc1/p6 AAs 1690-1708) utilizing UCSF Chimera. The trimmed ORF1 structure was aligned with the amphipathic thumb domain PDB, kindly provided by Jérôme Gouttenoire, using TM-Align. The distance plot generated between the AlphaFold ORF1 prediction and the macro domain of SINV was done by trimming a single chain of the trimer to remove amino acids outside the macrodomain of the SINV prediction P23pro-zbd(Shin et al., 2012) (utilizing AAs 1342 – 1509 of PDB: 4GUA) and aligning it to the ORF1 prediction using TM-Align. The CHIKV macrodomain PDB file 3GPG(Malet et al., 2009) was edited to remove the three additional chains comprising the hetero-tetramer and aligned with the ORF1 prediction

Molecular graphics and analysis.

Molecular graphics and analyses performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 (Pettersen et al., 2004).

Note: Kernow c1/p6 strain contains an s17 insertion within the HVR, so amino acid positions shift downstream of this insertion at amino acids 751-806.

Plasmid construction.

To construct lentiviral constructs encoding ORF1 of Kernow C1/p6 (GenBank accession number JQ679013), the Kernow C1/p6 ORF1 cDNA was amplified by PCR from a plasmid encoding the full-length (FL) infectious HEV clone Kernow C1/p6 (kindly provided by Suzanne Emerson, NIH) and then cloned into pLVX-IRES-zsGreen1 vector using an In-Fusion HD cloning kit (Clontech, Mountain View, CA, USA). The GAD mutant of ORF1 inactivating the polymerase was generated by QuikChange (Stratagene) site-directed mutagenesis. The HEV Kernow-C1 p6-Gluc (Shukla et al., 2011) and pSAR55-GLuc were kindly provided by Suzanne Emerson and Patricia Farci. pGEM-9zf-pSHEV3 and pGEM-7Zf(-)-TW6196E encoding the infectious pSHEV3 (gt 3) and TW6196 (gt 4) clone, respectively were gifts from X.J. Meng. Site-directed mutagenesis of these plasmids for C483A or Pol(-) mutants were obtained with the QuikChange kit (Stratagene) using primers listed in **Supplementary Table 1**. HEV Kernow-C1 p6-Gluc was used to generate the triplicate mutants for the alanine scanning mutagenesis of the entire pPCP domain, as well as the C434A, C457A, C459A, C471A, C472A, C481A, C483A, C563A, C483S, C483C (TGT), Y590A, Y590F, C483A-Y590A, C483G-Y590L, C483A (GCT), C483A (GCA), C483A (GCG), D248A, and H249A point mutants by QuikChange XL site-directed mutagenesis kit (Stratagene, La Jolla, CA). All primers used for site-directed mutagenesis can be found in **(Supplementary Table 1).** AAAPG-HA tag-AAAPG insert was generated by amplifying out the HA tag sequence from pLVX ORF2-HA using primers listed in (**Supplementary table 1**)**.** pLVX Kernow C1/p6 ORF1 AAAPG-HA tag-AAAPG IRES zsGreen

was generated via XhoI and XbaI digestion of pLVX IRES zsGreen plasmid and In-Fusion HD cloning of ORF1 AAAPG-HA tag-AAAPG from p6/BSR-2A-ZsGreen AAAPG-HA tag-AAAPG HVR using primers listed in **Supplementary Table 1.** Overexpression constructs for ORF1 AAAPG-HA tag-AAAPG point mutants C483A, C563A, D248A, and H249A were generated from the parent pLVX Kernow C1/p6 ORF1AAAPG-Ha tag-AAAPG IRES zsGreen construct via QuikChange XL site-directed mutagenesis kit (Stratagene, La Jolla, CA). using primers listed in (**Supplementary Table 1**).

Generation of HEV reporter genomes.

The generation of p6/BSR-2A-ZsGreen was described previously (Nimgaonkar et al., 2021). The generation of pSK-SAR55-Gluc, pGEM-9Zf-pSHEV3-Gluc, and pGEM-7Zf(-)-TW6196E-Gluc reporter constructs were described previously (Ding, Nimgaonkar, et al., 2018). Generation of pSK-Kernow AAAPG-HA tag-AAAPG HVR GLuc was conducted via PCR linearization of pSK Kernow WT GLuc and In-Fusion HD cloning of AAAPG-HA tag-AAAPG cloned out of p6/BSR-2A-ZsGreen AAAPG-HA tag-AAAPG HVR using primers listed in **Supplementary table 1.**
All DNA fragments were cloned into the respective vectors using an In-Fusion HD cloning kit (Clontech, Mountain View, CA, USA). All constructs or primers used to construct the HEV reporter genomes in **Supplementary Table 1** have been validated through Sanger sequencing and are available upon request.

***In vitro* transcription assay and viral RNA transfection.**

HEV Kernow-C1 p6-Gluc, HEV Kernow-C1 p6 C483A-Gluc, HEV Kernow-C1 p6 GAD-Gluc and all Kernow point mutant and alanine scanning triplite construct plasmids were linearized by MluI. pSAR55-Gluc, pSAR55 C483A-Gluc, and pSAR55 GAD-Gluc were linearized by BglII, pGEM-9Zf-pSHEV3-Gluc, pGEM-9Zf-pSHEV3 C483A-Gluc, and pGEM-9Zf-pSHEV3 GAD-Gluc were linearized by XbaI, and pGEM-7Zf(-)-TW6196E-Gluc, pGEM-7Zf(-)-TW6196E C483A-Gluc, and pGEM-7Zf(-)-TW6196E GAD-Gluc were linearized by SpeI. All capped viral RNA was *in vitro* transcribed from the corresponding linearized plasmid using the HiScribe T7 antireverse cap analog (ARCA) mRNA kit (New England Biolabs, Ipswich, MA) according to the manufacturer's protocol. *In vitro* transcribed viral RNA was purified by LiCl precipitation following DNAse1 digestion. *In vitro* transcribed viral RNA was transfected into HUH7, HUH7.5, or HepG2C3A cells via the TransIT-mRNA transfection reagent (Mirus Bio LLC, Madison, WI) according to the manufacturer's instructions.

***Gaussia* luciferase assays.**

Gaussia luciferase activity was determined using Luc-Pair Renilla luciferase HS assay kit (GeneCopoeia, Rockville, MD). Specifically, 10 µl of harvested cell culture medium was added per well of a 96-well solid white, flat-bottom polystyrene microplate (Corning, NY, USA), followed by the addition of Renilla luciferase assay substrate according to manufacturer protocol and the detection of luminescence was performed using a Berthold luminometer (Bad Wildbach, Germany).

Immunopurification of HA-tagged proteins.

Immunopurification of HA tagged proteins for ICP-MS analysis was conducted using the Pierce™ MS-Compatible Magnetic IP Kit, protein A/G (Thermo Scientific, Waltham, Massachusetts, Catalog number 90409). 750 ug of crude protein lysates of stably expressing ORF1 cells (WT or mutants) were subjected to each round of IP for subsequent analyses (ICP-MS or western blot). IP was done with rabbit anti-HA antibody (Cell Signaling Technology, Danvers, Massachusetts, catalog number C29F4) at a ratio of 1:50 in a laminar flow tissue culture hood that was sterilized with 70% ethanol and washed with MilliQ water prior to IP. The IP was conducted according to the manufacturer's instructions, and each collected fraction

(crude cell lysate, unbound fraction, and IP eluate) were split into large (80% total volume) and small (20% total volume) fractions. The large fractions were used for downstream nitric acid/hydrogen peroxide digestion and ICP-MS analysis, whereas the small fraction was used for protein quantification and western blot analysis. This was necessary to reduce chances of contaminating divalent ions being introduced into the sample during non-ICP-MS characterization.

HA-Tagged Immunopurified Protein Digestion.

Eluted protein samples were freeze dried and placed at -80C until nitric acid digestion. Briefly, protein samples were resuspended in 1.5mL Nitric Acid 67-69% Optima™, for Ultra Trace Elemental Analysis (Fisher Chemical, Fair Lawn, NJ) in ICP-MS grade Teflon at 80C for 24 hours in a laminar flow hood. 500uL Hydrogen Peroxide (Optima™, Fisher Chemical, Fair Lawn, NJ) was added to each sample and incubated at 80°C for 24 hours in a laminar flow hood, when the samples were subsequently dried for 24 hours at 40°C. Fully digested samples were then resuspended in 1mL 2% (vol/vol) nitric acid dilutes with MilliQ water, diluted in 1x, 2x, and 10x dilution series, and interrogated via ICP-MS (iCap, Thermo Scientific, Waltham, MA).

Inductively coupled plasma mass spectrometry (ICP-MS).

Samples were quantified via a single quadrupole iCap ICP-MS (Thermo Scientific, Waltham, MA). Briefly, 1mL of 2% (vol/vol) nitric acid running buffer blank was measured, followed by 1mL per dilution of a dilution series (1x, 2x, 5x, 10x, 50x) of the certified elemental standard 1643F (National Institute of Standards and Technology, Gaithersburg, MD) to generate an elemental calibration curve. All elements were measured in standard (STD) mode with the exception of lithium and iron, which were measured in kinetic (KED) mode to remove unwanted polyatomic interferences with the argon plasma. Samples were bracketed by an additional blank and standard curve to monitor instrument drift and ensure consistency throughout each experimental run. Samples were run via a series of three dilutions per sample, per run. Any measurement for a sample within the dilution series that fell outside of the dilution series range was excluded from further analysis. Contamination of trace elements from reagents and MilliQ water was monitored by processing MilliQ ‘samples’ through the entire immunopurification and protein digestion protocol on a run-by-run basis. Any trace metal analyses that was not statistically different from the MilliQ control ‘sample’ was excluded from further consideration. Iron and Zinc ions were binned prior to statistical analysis for ease due to having with multiple isotopes.

Immunofluorescence and confocal microscopy.

Naïve HUH7 cells or HUH7 cells expressing ORF1 WT, WT-HA-tag, C483A-HA-tag, C563-HA-tag, D248A-HA-tag, or H249A-HA-tag were seeded onto separate glass coverslips (#1.5; 10 mm; Thomas Scientific, Swedesboro, NJ) in a 24-well plate at 100,000 cells per well. Two days post seeding, the cells were fixed with 4% PFA for 15 minutes and subsequently permeabilized in 0.25% Triton x-100 for 15 minutes. The rabbit anti-HA tag, C29F4 (Cell Signaling Technology, Danvers, MA) primary antibody was used at a ratio of 1:1000 (V/V), and the AlexaFluor647 (goat anti-rabbit IgG (H+L), ThermoFisher Scientific, Waltham, MA) secondary antibody was used at a final concentration of 1 ug/mL. All antibodies were diluted with PBS and incubated for 40 minutes at room temperature (RT). Hoechst 33342 (ThermoFisher Scientific, Waltham, MA) was incubated at a final concentration of 1 ug/mL for 10 minutes at RT. The coverslips were then mounted onto glass microscopic slides (VWR International, Radnor, PA) with 5 µL of ProLong gold antifade reagent (ThermoFischer Scientific, Waltham, MA, United States). The stained samples were imaged using the Nikon A1R-Si microscope (Nikon, Melville, NY) in the Princeton University Confocal Microscopy Facility. The images were taken at 40x magnification. Images were then analyzed using Fiji (ImageJ2) image analysis software.

Statistical analysis.
All statistical analyses were performed using GraphPad Prism software version 9.3.1. One-way analysis of variance (ANOVA) with Dunnett's multiple comparison analysis or Brown-Forsythe one-way analysis of variance (ANOVA) with Dunnett's T3 multiple comparison analysis tests were used to test for statistical significance of the differences between the different group parameters in experiments utilizing the *Gaussia* Luciferase reporter replicon. *P* values of less than 0.05 were considered statistically significant. All data sets were analyzed for and cleaned of outliers using the ROUT method.

Conflict of Interest Statement:
The authors declare no conflicts of interest pertaining to the results reported in this study.

Funding Sources:
Work in the lab is supported by grants from the National Institutes of Health (R01 AI138797, R01 AI107301, R01 AI146917, R01 AI153236 to A.P.), a Burroughs Wellcome Fund Award for Investigators in Pathogenesis (#101539 to A.P.) and funding from Princeton University. R.L and S.M. were supported by the National Institute of General Medicine Sciences of the National Institutes of Health under Award Number T32GM007388. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. (DGE-2039656) awarded to R.L. The Molecular Biology Flow Cytometry Resource Facility is partially supported by the Cancer Institute of New Jersey Cancer Center Support grant (P30CA072720).

Acknowledgments:
We kindly thank Susan Emerson and Patricia Farci (NIAID) for providing us with the pSK SAR55, pBSK(+) Kc1 ORF1 WT GLuc plasmid, X.J. Meng (Virginia-Maryland College of Veterinary Medicine) for providing us the plasmids pGEM-9zf-pSHEV3 and pGEM-7Zf(-)-TW6196E encoding the infectious pSHEV3 (gt 3) and TW6196 (gt 4) clone, respectively. HEV ORF1 RNA dependent RNA Polymerase structural data was kindly provided by Jérôme Gouttenoire. We would also like to thank Christina DeCoste and Katherine Rittenbach in the Molecular Biology flow cytometry core facility and Dr. Gary Laevsky and the Molecular Biology Confocal Microscopy Facility which is a Nikon Center of Excellence for their excellent technical support. We further thank Dr. Frederick Hughson for indispensable advice and expertise, as well as Jaden Shirkey and Kevin DAmico of the Hughson lab, and all members of the Ploss lab for critical discussions and comments throughout experimentation and preparation of the manuscript.

Author Contributions:
This project was conceived by R.L. and A.P. All genetic experiments were performed by R.L. Bioinformatic analyses were performed by R.L. and A.B. ICP-MS experiments were designed by R.L., S.G, A.P., and J.H., and were performed by. R.L. and S.G. ICP-MS experiments conducted in facilities provided by J.H. Confocal imaging was performed by R.L. and S.M. All data was analyzed by R.L. and A.P. The draft of the manuscript was prepared by R.L. and A.P. with edits from all authors.

Materials Availability Statement: All materials generated by the Ploss lab will be available upon request from the corresponding author.

688 **Data Availability:** All source data has been provided and uploaded to eLIFE at initial
689 submission.
690

691 **References:**

692 Ansari, I. H., Nanda, S. K., Durgapal, H., Agrawal, S., Mohanty, S. K., Gupta, D., Jameel, S., &
693 Panda, S. K. (2000). Cloning, sequencing, and expression of the hepatitis E virus (HEV)
694 nonstructural open reading frame 1 (ORF1). *J Med Virol*, *60*(3), 275-283.
695 https://www.ncbi.nlm.nih.gov/pubmed/10630959
696 Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its
697 supplement TrEMBL in 2000. *Nucleic Acids Res*, *28*(1), 45-48.
698 https://doi.org/10.1093/nar/28.1.45
699 Bergmann, E. M., Cherney, M. M., McKendrick, J., Frormann, S., Luo, C., Malcolm, B. A.,
700 Vederas, J. C., & James, M. N. (1999). Crystal structure of an inhibitor complex of the
701 3C proteinase from hepatitis A virus (HAV) and implications for the polyprotein
702 processing in HAV. *Virology*, *265*(1), 153-163. https://doi.org/10.1006/viro.1999.9968
703 Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M.
704 J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., & Schneider, M. (2003). The
705 SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic*
706 *Acids Res*, *31*(1), 365-370. https://doi.org/10.1093/nar/gkg095
707 Carpentier, M., & Chomilier, J. (2019). Protein multiple alignments: sequence-based versus
708 structure-based programs. *Bioinformatics*, *35*(20), 3970-3980.
709 https://doi.org/10.1093/bioinformatics/btz236
710 Chasapis, C. T. (2018). Interactions between metal binding viral proteins and human targets as
711 revealed by network-based bioinformatics. *J Inorg Biochem*, *186*, 157-161.
712 https://doi.org/10.1016/j.jinorgbio.2018.06.012
713 de Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E.,
714 Bairoch, A., & Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches
715 and ProRule-associated functional and structural residues in proteins. *Nucleic Acids*
716 *Res*, *34*(Web Server issue), W362-365. https://doi.org/10.1093/nar/gkl124
717 Devhare, P., Sharma, K., Mhaindarkar, V., Arankalle, V., & Lole, K. (2014). Analysis of helicase
718 domain mutations in the hepatitis E virus derived from patients with fulminant hepatic
719 failure: effects on enzymatic activities and virus replication. *Virus Res*, *184*, 103-110.
720 https://doi.org/10.1016/j.virusres.2014.02.018
721 Ding, Q., Gaska, J. M., Douam, F., Wei, L., Kim, D., Balev, M., Heller, B., & Ploss, A. (2018).
722 Species-specific disruption of STING-dependent antiviral cellular defenses by the Zika
723 virus NS2B3 protease. *Proc Natl Acad Sci U S A*, *115*(27), E6310-E6318.
724 https://doi.org/10.1073/pnas.1803406115
725 Ding, Q., Heller, B., Capuccino, J. M., Song, B., Nimgaonkar, I., Hrebikova, G., Contreras, J. E.,
726 & Ploss, A. (2017). Hepatitis E virus ORF3 is a functional ion channel required for
727 release of infectious particles. *Proc Natl Acad Sci U S A*, *114*(5), 1147-1152.
728 https://doi.org/10.1073/pnas.1614955114
729 Ding, Q., Nimgaonkar, I., Archer, N. F., Bram, Y., Heller, B., Schwartz, R. E., & Ploss, A. (2018).
730 Identification of the Intragenomic Promoter Controlling Hepatitis E Virus Subgenomic
731 RNA Transcription. *MBio*, *9*(3). https://doi.org/10.1128/mBio.00769-18
732 Emerson, S. U., Nguyen, H., Graff, J., Stephany, D. A., Brockington, A., & Purcell, R. H. (2004).
733 In vitro replication of hepatitis E virus (HEV) genomes and of an HEV replicon
734 expressing green fluorescent protein. *J Virol*, *78*(9), 4838-4846.
735 https://doi.org/10.1128/jvi.78.9.4838-4846.2004
736 Graff, J., Nguyen, H., Kasorndorkbua, C., Halbur, P. G., St Claire, M., Purcell, R. H., &
737 Emerson, S. U. (2005). In vitro and in vivo mutational analysis of the 3'-terminal regions
738 of hepatitis e virus genomes and replicons. *J Virol*, *79*(2), 1017-1026.
739 https://doi.org/10.1128/JVI.79.2.1017-1026.2005

Haagsma, E. B., Riezebos-Brilman, A., van den Berg, A. P., Porte, R. J., & Niesters, H. G. (2010). Treatment of chronic hepatitis E in liver transplant recipients with pegylated interferon alpha-2b. *Liver Transpl*, *16*(4), 474-477. https://doi.org/10.1002/lt.22014

Johne, R., Reetz, J., Ulrich, R. G., Machnowska, P., Sachsenröder, J., Nickel, P., & Hofmann, J. (2014). An ORF1-rearranged hepatitis E virus derived from a chronically infected patient efficiently replicates in cell culture. *J Viral Hepat*, *21*(6), 447-456. https://doi.org/10.1111/jvh.12157

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

Kamar, N., Mallet, V., & Izopet, J. (2014). Ribavirin for chronic hepatitis E virus infection. *N Engl J Med*, *370*(25), 2447-2448. https://doi.org/10.1056/NEJMc1405191

Kanade, G. D., Pingale, K. D., & Karpe, Y. A. (2018). Activities of Thrombin and Factor Xa Are Essential for Replication of Hepatitis E Virus and Are Possibly Implicated in ORF1 Polyprotein Processing. *J Virol*, *92*(6). https://doi.org/10.1128/JVI.01853-17

Karpe, Y. A., & Lole, K. S. (2010). NTPase and 5' to 3' RNA duplex-unwinding activities of the hepatitis E virus helicase domain. *J Virol*, *84*(7), 3595-3602. https://doi.org/10.1128/JVI.02130-09

Karpe, Y. A., & Lole, K. S. (2011). Deubiquitination activity associated with hepatitis E virus putative papain-like cysteine protease. *J Gen Virol*, *92*(Pt 9), 2088-2092. https://doi.org/10.1099/vir.0.033738-0

Khuroo, M. S., & Kamili, S. (2003). Aetiology, clinical course and outcome of sporadic acute viral hepatitis in pregnancy. *J Viral Hepat*, *10*(1), 61-69. https://www.ncbi.nlm.nih.gov/pubmed/12558914

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol*, *72 (Pt 9)*, 2197-2206. https://doi.org/10.1099/0022-1317-72-9-2197

Koonin, E. V., Gorbalenya, A. E., Purdy, M. A., Rozanov, M. N., Reyes, G. R., & Bradley, D. W. (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*(17), 8259-8263. https://doi.org/10.1073/pnas.89.17.8259

Laitaoja, M., Valjakka, J., & Janis, J. (2013). Zinc coordination spheres in protein structures. *Inorg Chem*, *52*(19), 10983-10991. https://doi.org/10.1021/ic401072d

Lemon, S. M., Amphlett, E., & Sangar, D. (1991). Protease digestion of hepatitis A virus: disparate effects on capsid proteins, antigenicity, and infectivity. *J Virol*, *65*(10), 5636-5640. https://doi.org/10.1128/JVI.65.10.5636-5640.1991

Lenggenhager, D., Gouttenoire, J., Malehmir, M., Bawohl, M., Honcharova-Biletska, H., Kreutzer, S., Semela, D., Neuweiler, J., Hurlimann, S., Aepli, P., Fraga, M., Sahli, R., Terracciano, L., Rubbia-Brandt, L., Mullhaupt, B., Sempoux, C., Moradpour, D., & Weber, A. (2017). Visualization of hepatitis E virus RNA and proteins in the human liver. *J Hepatol*, *67*(3), 471-479. https://doi.org/10.1016/j.jhep.2017.04.002

Magden, J., Takeda, N., Li, T., Auvinen, P., Ahola, T., Miyamura, T., Merits, A., & Kääriäinen, L. (2001). Virus-specific mRNA capping enzyme encoded by hepatitis E virus. *J Virol*, *75*(14), 6249-6255. https://doi.org/10.1128/JVI.75.14.6249-6255.2001

Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X., & Canard, B. (2009). The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved

791 adenosine binding pocket. *J Virol*, *83*(13), 6534-6545. https://doi.org/10.1128/JVI.00189-
792 09
793 Metzger, K., Bentaleb, C., Hervouet, K., Alexandre, V., Montpellier, C., Saliou, J. M., Ferrie, M.,
794 Camuzet, C., Rouille, Y., Lecoeur, C., Dubuisson, J., Cocquerel, L., & Aliouat-Denis, C.
795 M. (2022). Processing and Subcellular Localization of the Hepatitis E Virus Replicase:
796 Identification of Candidate Viral Factories. *Front Microbiol*, *13*, 828636.
797 https://doi.org/10.3389/fmicb.2022.828636
798 Nimgaonkar, I., Archer, N. F., Becher, I., Shahrad, M., LeDesma, R. A., Mateus, A., Caballero-
799 Gomez, J., Berneshawi, A. R., Ding, Q., Douam, F., Gaska, J. M., Savitski, M. M., Kim,
800 H., & Ploss, A. (2021). Isocotoin suppresses hepatitis E virus replication through
801 inhibition of heat shock protein 90. *Antiviral Res*, *185*, 104997.
802 https://doi.org/10.1016/j.antiviral.2020.104997
803 O'Donovan, C., Martin, M. J., Gattiker, A., Gasteiger, E., Bairoch, A., & Apweiler, R. (2002).
804 High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform*,
805 *3*(3), 275-284. https://doi.org/10.1093/bib/3.3.275
806 Oechslin, N., Da Silva, N., Szkolnicka, D., Cantrelle, F. X., Hanoulle, X., Moradpour, D., &
807 Gouttenoire, J. (2022). Hepatitis E virus RNA-dependent RNA polymerase is involved in
808 RNA replication and infectious particle production. *Hepatology*, *75*(1), 170-181.
809 https://doi.org/10.1002/hep.32100
810 Paliwal, D., Panda, S. K., Kapur, N., Varma, S. P., & Durgapal, H. (2014). Hepatitis E virus
811 (HEV) protease: a chymotrypsin-like enzyme that processes both non-structural
812 (pORF1) and capsid (pORF2) protein. *J Gen Virol*, *95*(Pt 8), 1689-1700.
813 https://doi.org/10.1099/vir.0.066142-0
814 Parvez, M. K. (2013). Molecular characterization of hepatitis E virus ORF1 gene supports a
815 papain-like cysteine protease (PCP)-domain activity. *Virus Res*, *178*(2), 553-556.
816 https://doi.org/10.1016/j.virusres.2013.07.020
817 Parvez, M. K. (2015). The hepatitis E virus ORF1 'X-domain' residues form a putative
818 macrodomain protein/Appr-1″-pase catalytic-site, critical for viral RNA replication. *Gene*,
819 *566*(1), 47-53. https://doi.org/10.1016/j.gene.2015.04.026
820 Parvez, M. K. (2017). Mutational analysis of hepatitis E virus ORF1 "Y-domain": Effects on RNA
821 replication and virion infectivity. *World J Gastroenterol*, *23*(4), 590-602.
822 https://doi.org/10.3748/wjg.v23.i4.590
823 Parvez, M. K., & Khan, A. A. (2014). Molecular modeling and analysis of hepatitis E virus (HEV)
824 papain-like cysteine protease. *Virus Res*, *179*, 220-224.
825 https://doi.org/10.1016/j.virusres.2013.11.016
826 Perttilä, J., Spuul, P., & Ahola, T. (2013). Early secretory pathway localization and lack of
827 processing for hepatitis E virus replication protein pORF1. *J Gen Virol*, *94*(Pt 4), 807-
828 816. https://doi.org/10.1099/vir.0.049577-0
829 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., &
830 Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and
831 analysis. *J Comput Chem*, *25*(13), 1605-1612. https://doi.org/10.1002/jcc.20084
832 Proudfoot, A., Hyrina, A., Holdorf, M., Frank, A. O., & Bussiere, D. (2019). First Crystal Structure
833 of a Nonstructural Hepatitis E Viral Protein Identifies a Putative Novel Zinc-Binding
834 Protein. *J Virol*, *93*(13). https://doi.org/10.1128/JVI.00170-19
835 Rein, D. B., Stevens, G. A., Theaker, J., Wittenborn, J. S., & Wiersma, S. T. (2012). The global
836 burden of hepatitis E virus genotypes 1 and 2 in 2005. *Hepatology*, *55*(4), 988-997.
837 https://doi.org/10.1002/hep.25505
838 Reyes, G. R., Huang, C. C., Tam, A. W., & Purdy, M. A. (1993). Molecular organization and
839 replication of hepatitis E virus (HEV). *Arch Virol Suppl*, *7*, 15-25.
840 https://doi.org/10.1007/978-3-7091-9300-6_2

Ropp, S. L., Tam, A. W., Beames, B., Purdy, M., & Frey, T. K. (2000). Expression of the hepatitis E virus ORF1. *Arch Virol*, *145*(7), 1321-1337. https://www.ncbi.nlm.nih.gov/pubmed/10963340

Saraswat, S., Chaudhary, M., & Sehgal, D. (2019). Hepatitis E Virus Cysteine Protease Has Papain Like Properties Validated by in silico Modeling and Cell-Free Inhibition Assays. *Front Cell Infect Microbiol*, *9*, 478. https://doi.org/10.3389/fcimb.2019.00478

Sehgal, D., Thomas, S., Chakraborty, M., & Jameel, S. (2006). Expression and processing of the Hepatitis E virus ORF1 nonstructural polyprotein. *Virol J*, *3*, 38. https://doi.org/10.1186/1743-422X-3-38

Shin, G., Yost, S. A., Miller, M. T., Elrod, E. J., Grakoui, A., & Marcotrigiano, J. (2012). Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *Proc Natl Acad Sci U S A*, *109*(41), 16534-16539. https://doi.org/10.1073/pnas.1210418109

Shukla, P., Nguyen, H. T., Faulk, K., Mather, K., Torian, U., Engle, R. E., & Emerson, S. U. (2012). Adaptation of a genotype 3 hepatitis E virus to efficient growth in cell culture depends on an inserted human gene segment acquired by recombination. *J Virol*, *86*(10), 5697-5707. https://doi.org/10.1128/JVI.00146-12

Shukla, P., Nguyen, H. T., Torian, U., Engle, R. E., Faulk, K., Dalton, H. R., Bendall, R. P., Keane, F. E., Purcell, R. H., & Emerson, S. U. (2011). Cross-species infections of cultured cells by hepatitis E virus and discovery of an infectious virus-host recombinant. *Proc Natl Acad Sci U S A*, *108*(6), 2438-2443. https://doi.org/10.1073/pnas.1018878108

Smith, D. B., Simmonds, P., Jameel, S., Emerson, S. U., Harrison, T. J., Meng, X. J., Okamoto, H., Van der Poel, W. H., Purdy, M. A., & Group, I. C. o. T. o. V. H. S. (2014). Consensus proposals for classification of the family Hepeviridae. *J Gen Virol*, *95*(Pt 10), 2223-2232. https://doi.org/10.1099/vir.0.068429-0

Suppiah, S., Zhou, Y., & Frey, T. K. (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

Szkolnicka, D., Pollan, A., Da Silva, N., Oechslin, N., Gouttenoire, J., & Moradpour, D. (2019). Recombinant Hepatitis E Viruses Harboring Tags in the ORF1 Protein. *J Virol*, *93*(19). https://doi.org/10.1128/JVI.00459-19

Tam, A. W., Smith, M. M., Guerra, M. E., Huang, C. C., Bradley, D. W., Fry, K. E., & Reyes, G. R. (1991). Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology*, *185*(1), 120-131. https://www.ncbi.nlm.nih.gov/pubmed/1926770

Tellinghuisen, T. L., Marcotrigiano, J., Gorbalenya, A. E., & Rice, C. M. (2004). The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J Biol Chem*, *279*(47), 48576-48587. https://doi.org/10.1074/jbc.M407787200

Tellinghuisen, T. L., Marcotrigiano, J., & Rice, C. M. (2005). Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature*, *435*(7040), 374-379. https://doi.org/10.1038/nature03580

Todt, D., Meister, T. L., & Steinmann, E. (2018). Hepatitis E virus treatment and ribavirin therapy: viral mechanisms of nonresponse. *Curr Opin Virol*, *32*, 80-87. https://doi.org/10.1016/j.coviro.2018.10.001

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., . . . Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. *Nature*, *596*(7873), 590-596. https://doi.org/10.1038/s41586-021-03828-1

van der Heijden, M. W., & Bol, J. F. (2002). Composition of alphavirus-like replication complexes: involvement of virus and host encoded proteins. *Arch Virol*, *147*(5), 875-898. https://doi.org/10.1007/s00705-001-0773-3

van der Valk, M., Zaaijer, H. L., Kater, A. P., & Schinkel, J. (2017). Sofosbuvir shows antiviral activity in a patient with chronic hepatitis E virus infection. *J Hepatol*, *66*(1), 242-243. https://doi.org/10.1016/j.jhep.2016.09.014

van Wezel, E. M., de Bruijne, J., Damman, K., Bijmolen, M., van den Berg, A. P., Verschuuren, E. A. M., Ruigrok, G. A., Riezebos-Brilman, A., & Knoester, M. (2019). Sofosbuvir Add-on to Ribavirin Treatment for Chronic Hepatitis E Virus Infection in Solid Organ Transplant Recipients Does Not Result in Sustained Virological Response. *Open Forum Infect Dis*, *6*(8). https://doi.org/10.1093/ofid/ofz346

Yang, S. H., Lee, C. G., Song, M. K., & Sung, Y. C. (2000). Internal cleavage of hepatitis C virus NS3 protein is dependent on the activity of NS34A protease. *Virology*, *268*(1), 132-140. https://doi.org/10.1006/viro.1999.0168

Zhang, Y., & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*, *33*(7), 2302-2309. https://doi.org/10.1093/nar/gki524

Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., & Minor, W. (2008). Data mining of metal ion environments present in protein structures. *J Inorg Biochem*, *102*(9), 1765-1776. https://doi.org/10.1016/j.jinorgbio.2008.05.006

Figure Legends:

Figure 1. Mutations within the HEV putative protease domain render the virus replication incompetent.
a, Genome organization of HEV and Kernow strain genotype 3 reporter replicon. ORF2 and ORF3 were replaced by *Gaussia* Luciferase reporter in frame with subgenomic promoter and translation start site. MeT – methyltransferase. Y – Y-domain. PCP – putative papain-like cysteine protease. HVR – hypervariable region. X – macro-domain. Hel – helicase. RdRp – RNA-dependent RNA polymerase. organization. The putative catalytic dyad of HEV within the reporter replicon pPCP is denoted (C483 and Y590). GDD - catalytic triad of RdRp. Gluc – Gaussia Luciferase. **b**, Replication kinetics of Kc1/p6 WT HEV Gluc RNA or HEV pol (-) Gluc, or HEV RNA bearing mutations in putative PCP transfected into HUH7.5 cells. Cell culture supernatants were collected from transfected cells at time points indicated (drop in signal at D2 indicates wash step to eliminate signal from input RNA. **c**, End point analysis (day 4) of data from panel b. **d-g**) End point analysis (4 days post transfection) comparison of C483A mutant replication kinetics to WT and pol (-) in **d**, Kernow strain (genotype 3) **e**, SHEV (genotype 3) strain **f**, SAR55 (genotype 1) and **g**, TW6196E (genotype 4) of HEV when transfected into HepG2C3A human hepatoma cells. **h**, Alternate cysteine and alanine codons at C483 show viral replication deficiency caused by amino acid substitution and not RNA base substitution. These data represent an end point analysis (4 days post transfection) when HEV RNA replicons are transfected into HUH7 human hepatoma cells. One-Way ANOVA with Dunnett's multiple comparison analysis were conducted to determine significance. * - $p<0.05$, ** - $p<0.01$, *** - $p<0.001$, **** - $p<0.0001$. Data shown in d-g are from experiments done in technical and biological triplicate. Data shown in b, c, and h are from experiments done in technical triplicate and biological duplicate. Raw Gluc data provided in file Figure 1 – source data 1.

Figure 2. HEV ORF1 putative PCP cannot function outside of the context of the full-length protein, and C483A replication deficiency can be rescued *in trans*.
a, Schematic of the transcomplementation assay. HepG2C3A human hepatoma cells were transduced with lentivirus expressing HEV ORF1 wt, Pol (-), C483A, or putative PCP only (ORF1 AAs 433 – 592 in Kc1/p6) and subsequently transfected with *in vitro* transcribed RNA of wt, pol (-), or C483A replicons. End point analysis was conducted 4 days post transfection with *Gaussia* luciferase quantification. Cells transduced with WT ORF1 can rescue luciferase expression in pol (-) and C483A mutants *in* trans, whereas all other conditions, including the pPCP out of context of ORF1, cannot. **b-e**, End point analysis of fold change luciferase signal over WT replicon signal of **b**, Mock **c**, WT **d**, Pol (-) or **e**, C483A. One-Way ANOVA with Dunnett's multiple comparison analysis were conducted to determine significance. * - $p<0.05$, ** - $p<0.01$, *** - $p<0.001$, **** - $p<0.0001$. Schematic in panel a generated with BioRender. Pol (-) – replication incompetent replicon due to mutation in RNA dependent RNA polymerase. PCP – putative papain-like cysteine protease. Gluc – *Gaussia* luciferase. Data shown in b-e are from experiments done in technical and biological triplicate. Raw Gluc data provided in file Figure 2 – Source Data.

Figure 3. Point mutations of highly conserved cysteines and alanine scanning mutagenesis within putative PCP identifies residues and regions indispensable for viral replication.
a, Partial sequence of Kc1/p6 HEV ORF1 including putative PCP that shows 8 cysteines that are highly conserved across all 8 known HEV genotypes. Orange – upstream 40 amino acids within Y domain prior to putative PCP. Blue – downstream 40 amino acids within HVR after putative PCP. **b**, HepG2C3A cells were transfected with *in vitro* transcribed RNA of WT, polymerase deficient, or point mutants in one of the conserved cysteines within the putative

PCP. Cell culture supernatants were collected for four days post transfection prior to *Gaussia* luciferase quantification. **c**, Unbiased triplet alanine scanning mutagenesis of entire putative PCP region. HepG2C3A cells were transfected with *in vitro* transcribed RNA of triplet alanine scanning mutant replicons to assess viral replication capacity. Data shown are fold change of wild type replicon. Brown-Forsythe One-Way ANOVA with Dunnett's T3 multiple comparison analysis were conducted to determine significance. * - p<0.05, ** - p<0.01, *** - p<0.001, **** - p<0.0001. HVR – hypervariable region. Raw Gluc data provided in file Figure 3 – Source Data.

Figure 4. Structural prediction models of HEV pPCP demonstrates low-confidence scores, suggesting lack of highly ordered secondary structure.
a, HEV ORF1-WT AlphaFold structure prediction. **b**, HEV ORF1-WT AlphaFold structure prediction pseudo-colored by pLDDT score gradation (darker blue – higher pLDDT Score, darker red – lower pLDDT score). **c**, pLDDT score of AlphaFold prediction of HEV ORF1 across HEV genome organization for all of ORF1 (black, average 76.96), pPCP when measured with entirety of ORF1 (green, average 66.05), and pPCP when predicted by AlphaFold alone (purple, average 65.92). **d**, HEV ORF1 pPCP AlphaFold prediction pseudo-colored by alanine scanning mutagenesis ORF1 replication tolerance. Color based on fold change of WT in (**Figure 3C**). Orange – below (-2) Sea green – between (-2) and (-1). Cornflower blue – between (-1) and 0. Dark blue – replicated above WT levels. Beginning and end residues of pPCP noted in bold.

Figure 5. Structural prediction models suggest conserved cysteines within CxC[x_{11}]CC[x_8]CxC motif form divalent ion coordination pockets and novel domain-domain interaction with upstream Y-domain.
a, Alphafold structural predictions of domains within HEV ORF1. Left: Pseudo zinc-finger (amino acids 451-493 within putative PCP). Magenta: conserved cysteines C457, C459, C471, C472, C481, C483A. Potential divalent ion coordination tetrahedron outlined in yellow hatched line. Middle: amino acids 242-259 of HEV ORF1 Y-domain, and amino acids 451-462 of HEV putative PCP. Conserved cysteines C457 and C459 are outlined in magenta. D248 and H249 of upstream Y-domain highlighted in blue. Novel interdomain divalent ion coordination domain outlined in yellow hatched line. Right: HEV ORF1 protein demonstrating folding of WT and point mutant proteins. Orange (AAs 1-1036): methyltransferase(Magden et al., 2001). Yellow (AAs 1018-1262): Helicase(Devhare et al., 2014; Karpe & Lole, 2010). Red (AAs 1257-1709): RNA dependent RNA polymerase(Koonin et al., 1992; Oechslin et al., 2022). Cyan: Putative membrane association domain(Parvez, 2017). **b**, Multiple sequence alignment of HEV genotypes 1-8 of partial Y-domain containing variable residue D248 and highly conserved residue H249. (*) identical residue. (:) similar residue. (.) dissimilar residue. Yellow hatched line – bonds between coordinating amino acids

Figure 6. Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) Shows Bivalent Cation Coordination by HEV ORF1.
a, HA-epitope tag flanked by linker sequence inserted into HVR of HEV ORF1 lentiviral construct. **b**, Western blot analysis of HUH7 human hepatoma cell produced epitope tagged ORF1 purification (crude cell lysate, unbound fraction, purified ORF1). **c**, Workflow of protein sample preparation, purification, and HNO3/H2O2 digestion for ICP-MS. IP- Immuno-purification WT HA-tagged ORF1 (left), C483A HA-tagged ORF1 (middle), C563A HA-tagged ORF1 (right) d, Fold-change over WT of biologically relevant divalent metal ions(Zheng et al., 2008) bound by purified mutant epitope tagged ORF1. Any trace metal not measured as a value above a MilliQ water only purification control was excluded from this analysis on a run by run basis due to not being above IP buffer elution conditions. Two-Way ANOVA with Dunnett's multiple comparison analysis were conducted to determine significance. * - p<0.05, ** - p<0.01,

*** - p<0.001, **** - p<0.0001. Schematics in a and c were created with Biorender.com. Unedited western blots included in file Figure 6 – Source data.

Figure 7. Immunofluorescence of epitope tagged ORF1 demonstrates loss of membrane association when divalent ion coordination residues are mutated.
Confocal microscopy images of HUH7 cells stably expressing ORF1 bicistronic for zsGreen. ORF1-HA tagged proteins were imaged using a rabbit-anti-HA antibody and AlexaFluor647 (goat anti-rabbit IgG (H+L) secondary antibody. Nuclei were visualized with Hoechst 33342 stain. All images taken at 40x magnification, and analyzed using Fiji (ImageJ2) image analysis software. Source files of all TIFF images included in Figure 7 – Source Data.

Figure 1 - Source Data 1 - Mutations within the HEV putative protease domain render the virus replication incompetent. Spreadsheet of raw GLuc data kinetics for **b-c**) point and double mutant bearing replicons, **d**) Kc1/p6 genotype 3 HEV replicons, **e**) SHEV3 genotype 3 replicons, **f**) SAR55 genotype 1 HEV replicons, **g**) Tw6196E genotype 4 replicons, **h**) alternate codons of Kc1/p6 genotype 3 HEV replicons.

Figure 2 – Source Data - HEV ORF1 putative PCP cannot function outside of the context of the full-length protein, and C483A replication deficiency can be rescued *in trans.* Spreadsheet of raw GLuc data kinetics for transduced HepG2C3A cells (mock, ORF1 WT, ORF1 Pol (-), ORF1 C483A, HEV pPCP domain) transfected with HEV Kc1/p6 replicons (WT, Pol (-), C483A). Each sheet corresponds to different transduction condition.

Figure 3 – Source Data - Point mutations of highly conserved cysteines and alanine scanning mutagenesis within putative PCP identifies residues and regions indispensable for viral replication. Spreadsheet of raw GLuc data kinetics for Kc1/p6 conserved cysteine point mutant replicons and alanine scanning mutagenized triplets.

Figure 6 – Source Data - Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) Shows Bivalent Cation Coordination by HEV ORF1. Unedited western blots of ORF1 HA-IP.

Figure 7 – Source Data - Immunofluorescence of epitope tagged ORF1 demonstrates loss of membrane association when divalent ion coordination residues are mutated. TIFF files of confocal microscopy used in subcellular localization analysis.

AlphaFold Predictions Figures 4-5 – Source Data. These files are the best ranked (ranked 0) predictions generated by AlphaFold of HEV ORF1, its associated point mutants, and the hepatitis A virus (HAV) 3C protease.

Figure 1:

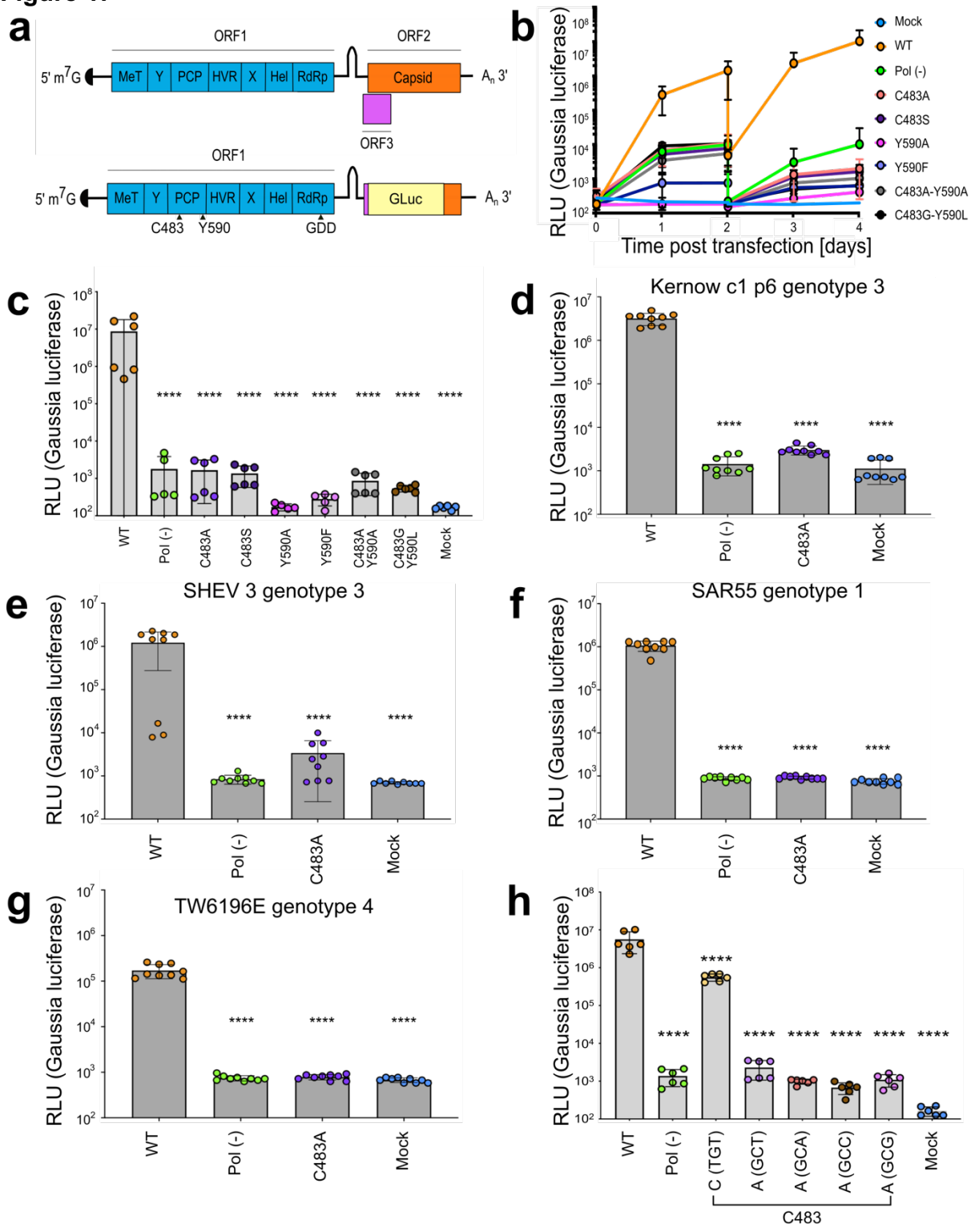


Figure 2:

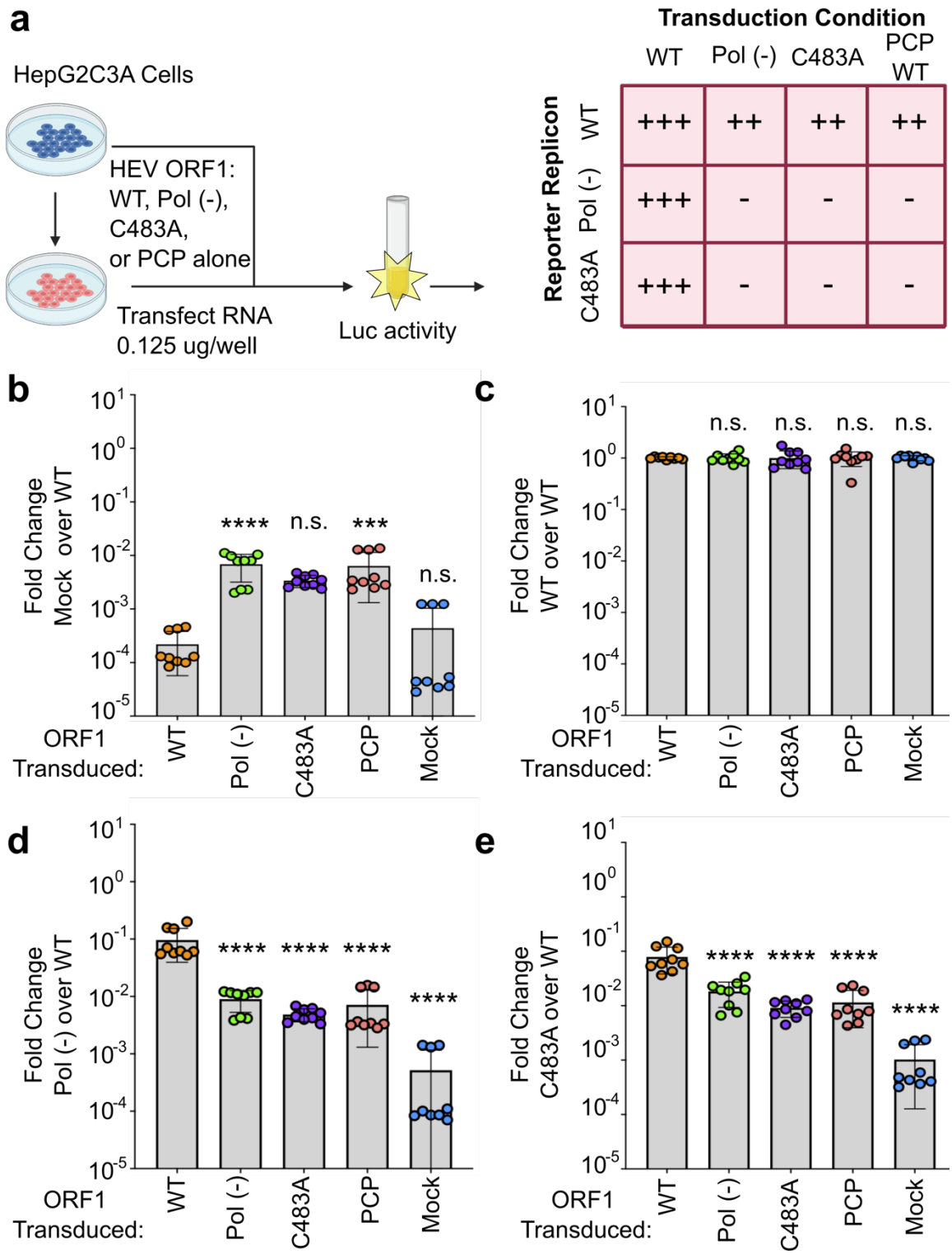


Figure 3:

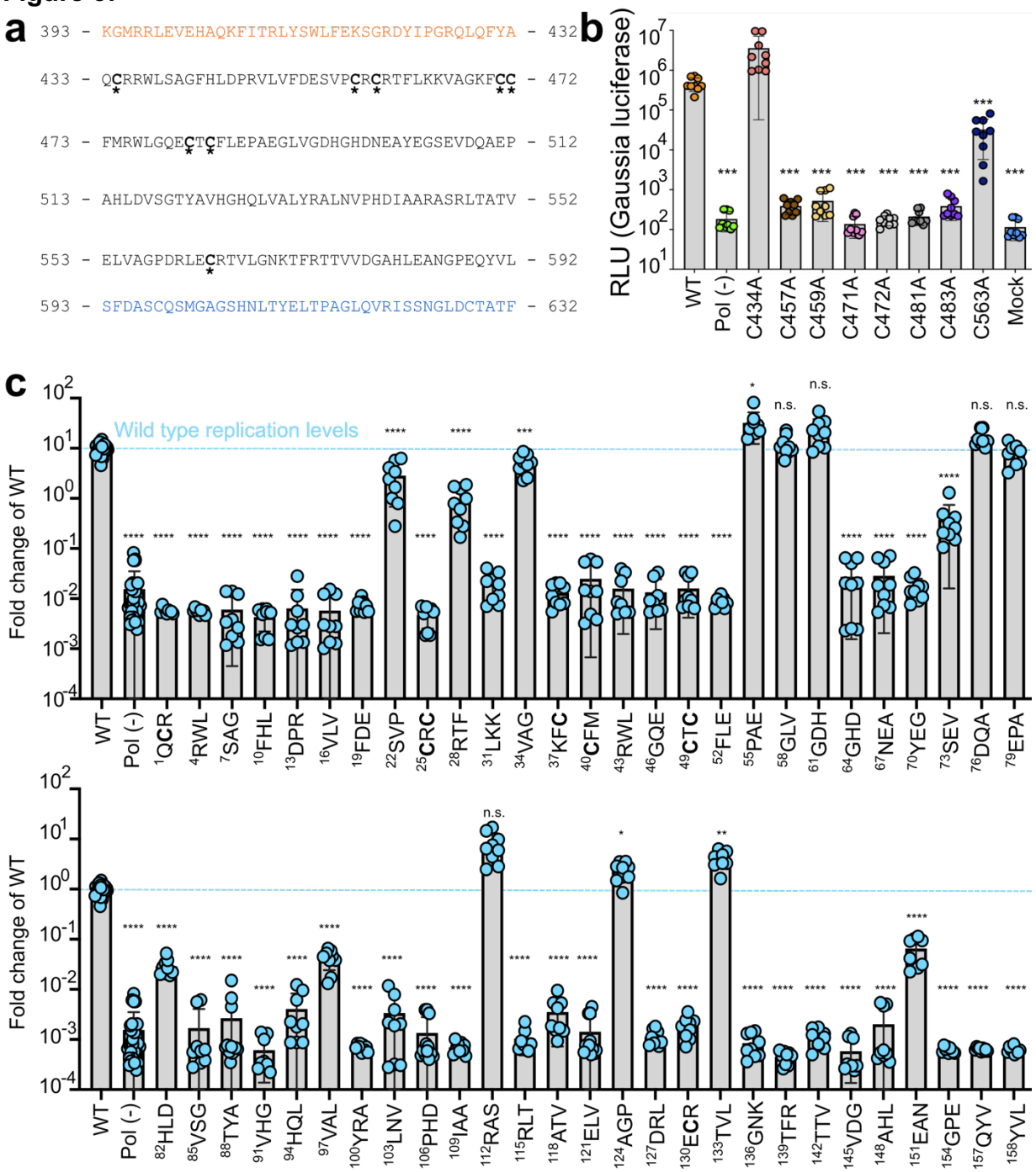


Figure 4:

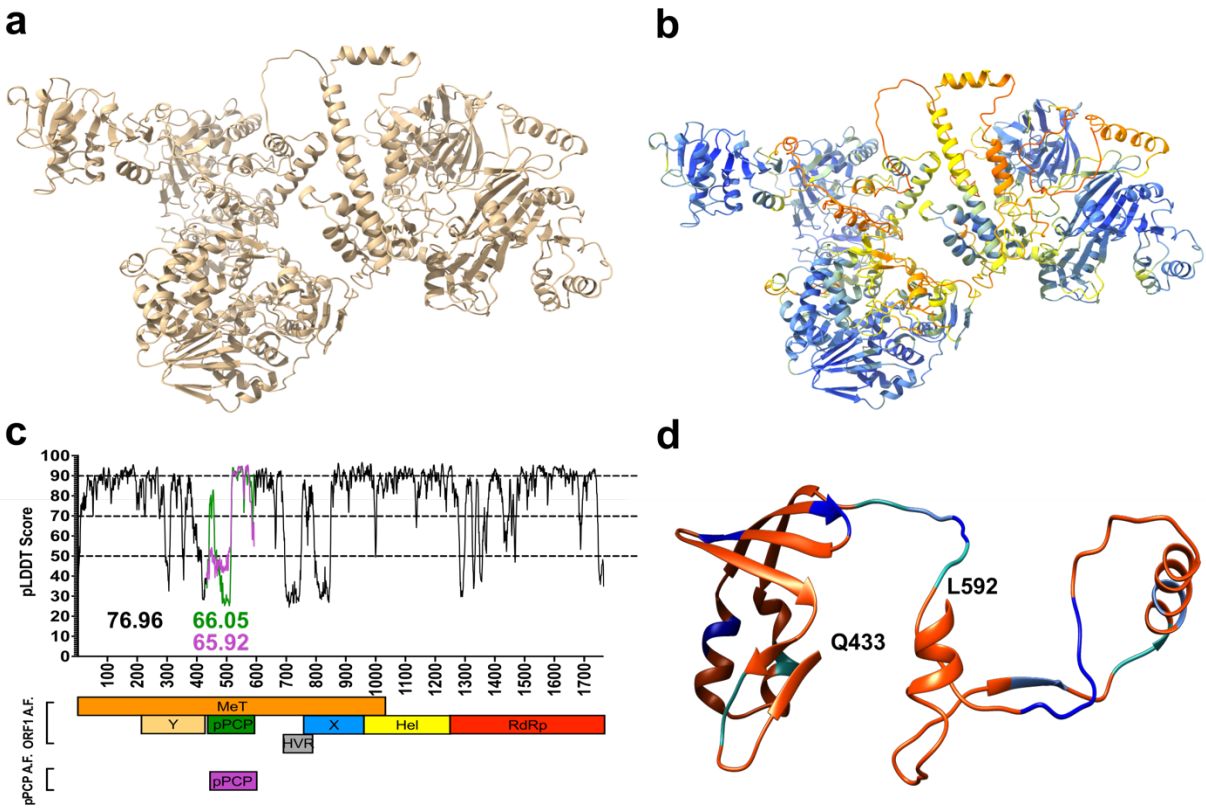
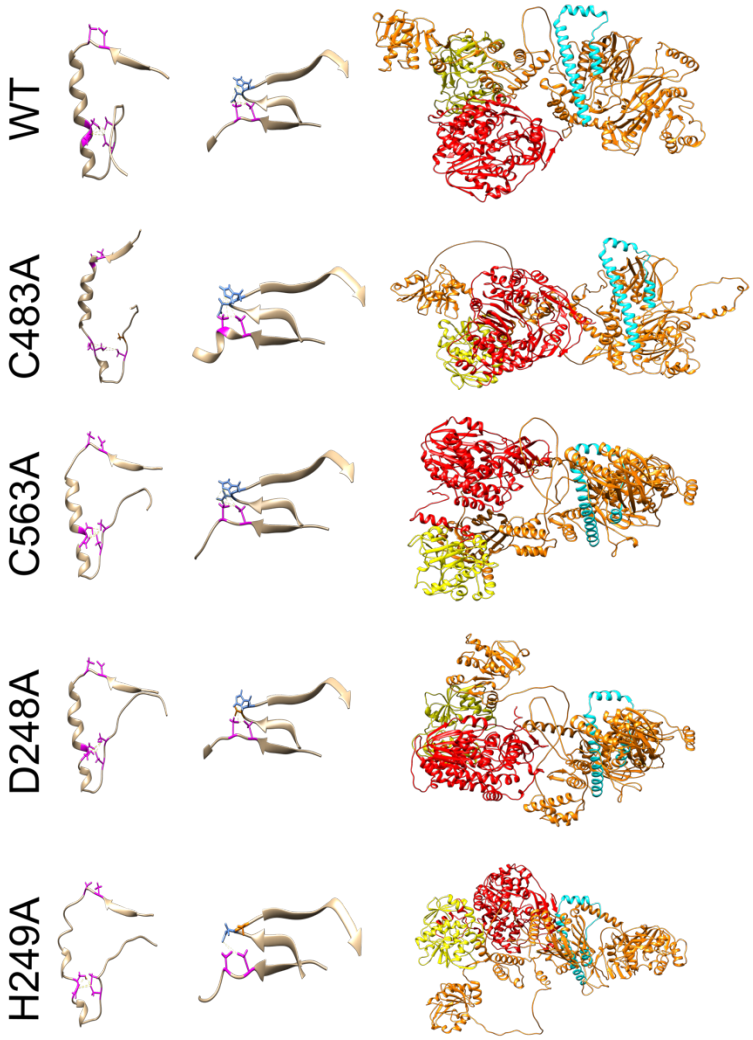


Figure 5:

a



b

```
                                    **:...******
HEV gt 1                  243 - TKVTGDHPLVIE - 254
HEV gt 1 (SAR55)          242 - TKVTGDHPLVIE - 253
HEV gt 2                  243 - TKVVGEHPLVIE - 254
HEV gt 3 (Kernow c1)      243 - TKIVGDHPLVIE - 254
HEV gt 3 (SHEV3)          243 - TKIVGDHPLVIE - 254
HEV gt 4                  243 - TKVTGDHPLVIE - 254
HEV gt 5                  243 - TKVTGSHPLVIE - 254
HEV gt 6                  243 - TKVTGNHPLVIE - 254
HEV gt 7                  243 - TKVTGNHPLVIE - 254
HEV gt 8                  243 - TKITTDHPLVIE - 254
```

Figure 6:

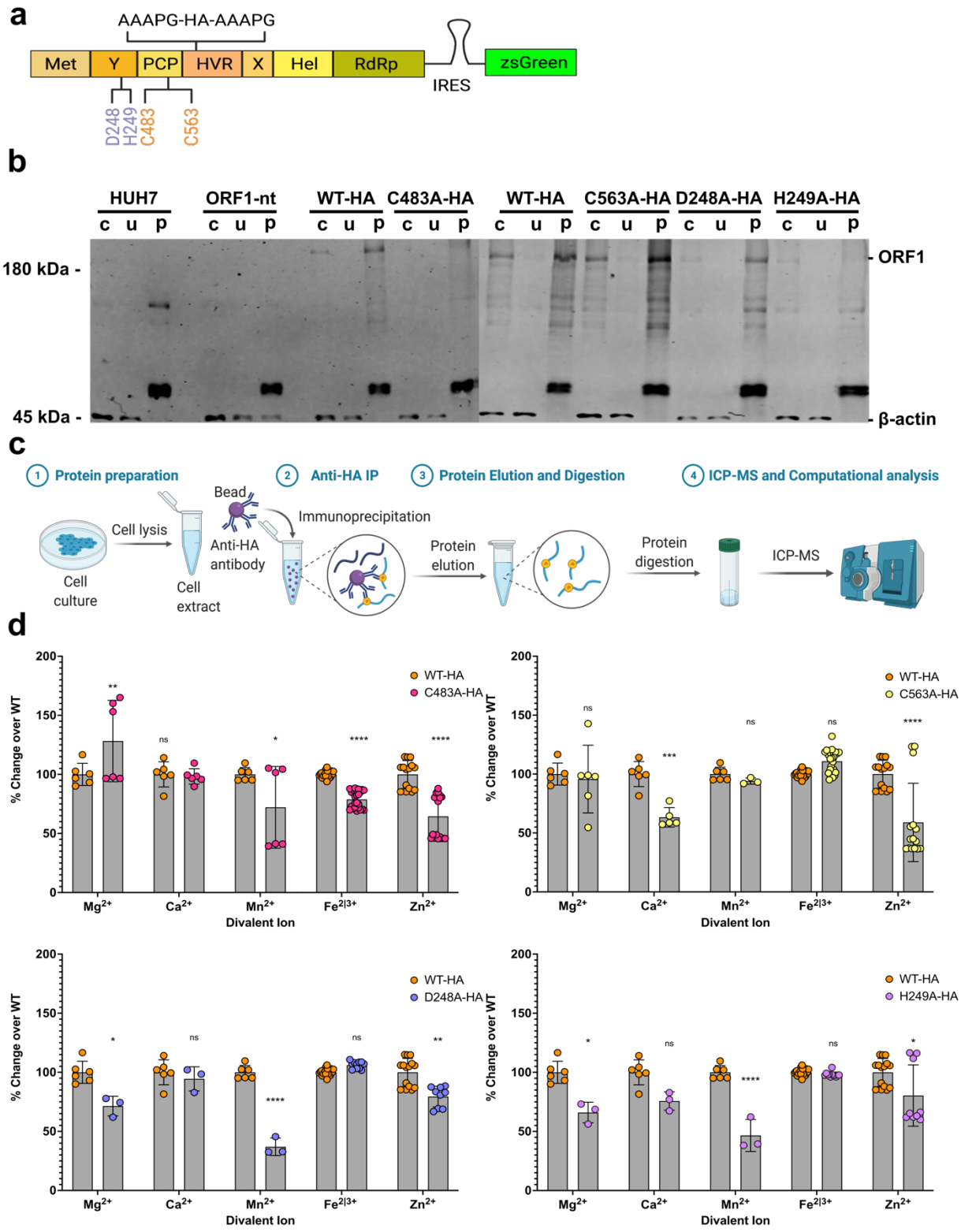
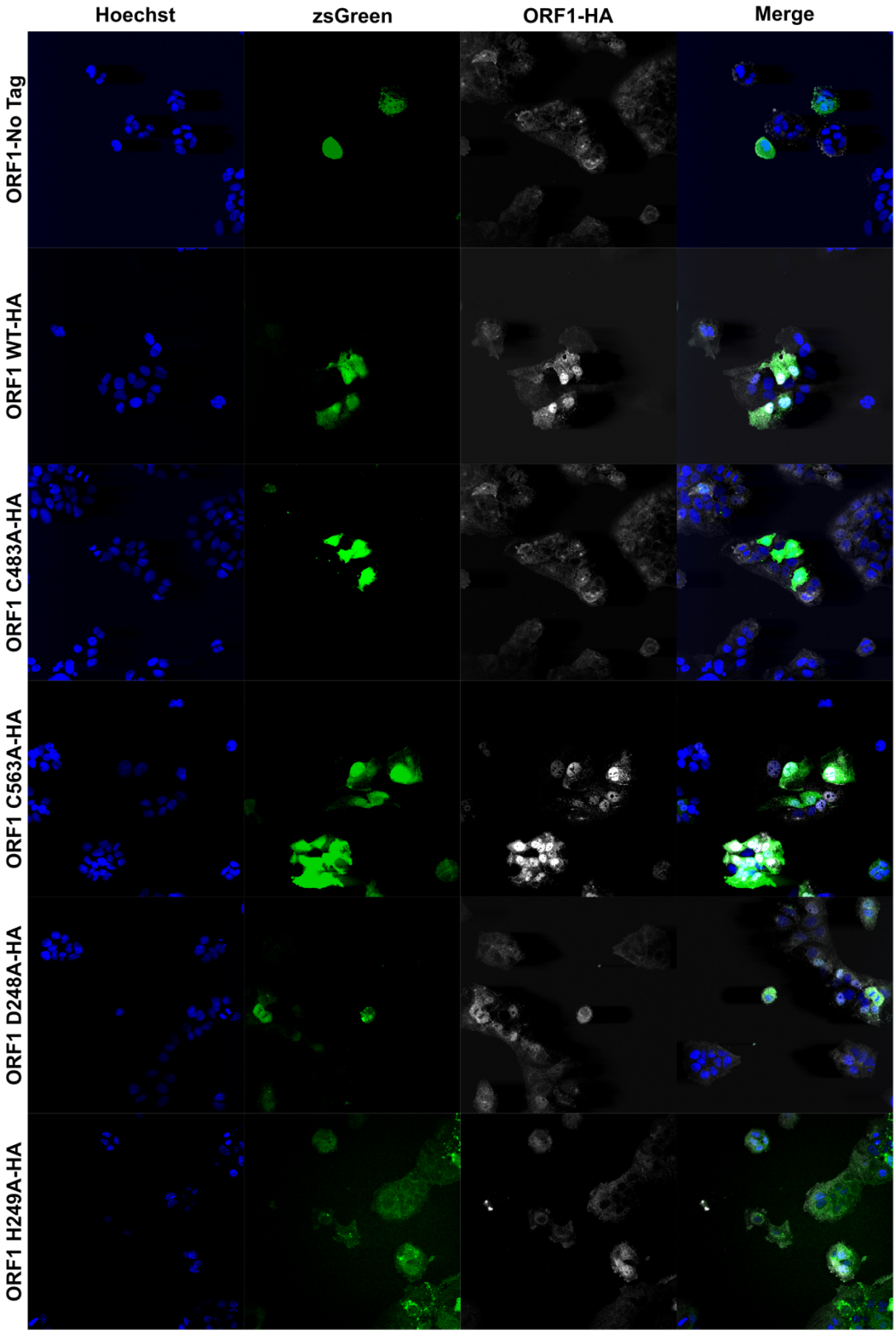


Figure 7:



SUPPLEMENTARY INFORMATION
FOR

Structural features stabilized by divalent cation coordination within hepatitis E virus ORF1 are critical for viral replication

Robert LeDesma[1], Abhishek Biswas[1], Stephanie Maya[1], Stefania Gili[2], John Higgins[2], Alexander Ploss[1],*

[1]Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, Princeton, NJ 08544, USA.

[2]Department of Geosciences, Princeton University, Princeton, NJ, 08544, USA.

*Correspondence should be addressed to A.P.: aploss@princeton.edu; +1 (609) 258-7128

Table of contents:

Supplementary Figures

Supplementary Figure 1. Alignment of one or more representative strain(s) from each known HEV genotype putative PCP domain with RUBV protease reveals highly conserved octa-cysteine motif in HEV.
Supplementary Figure 2. Bioinformatic analysis (PROSITE) predicts CxC[X11]CC[X8]CxC motif (CxC motif) to be necessary for divalent metal ion coordination.

Supplementary Figure 3. AlphaFold predicts structured domains of HEV ORF1 and viral protease of hepatitis A virus (HAV) with high confidence.

Supplementary Figure 4. Sequence alignments, percent similarity and identity of Kc1/p6 macrodomain with macrodomains of SINV and CHIKV.

Supplementary Figure 5. Replication end point analysis of WT-HA, D248A, and H249A Gluc constructs.

Supplementary Tables:
Supplementary Table 1. Primer sequences for all constructs in this study.

Supplementary Table 2. Predicted bond lengths (Å) of cysteines within HEV hexa-cysteine motif and D248/H249 of upstream Y-domain.

Supplementary Files:

Supplementary File 1: Raw trace element data for each sample series of ICP-MS

Supplementary Movies:
Supplementary Movie 1: AlphaFold Prediction of HEV ORF1-WT and alanine scan triplets highlighted demonstrates tolerable mutations largely localize to predicted disordered regions.
Supplementary Movie 2: AlphaFold Prediction of HEV ORF1-WT
Supplementary Movie 3: AlphaFold Prediction of HEV ORF1-C483A
Supplementary Movie 4: AlphaFold Prediction of HEV ORF1-C563A
Supplementary Movie 5: AlphaFold Prediction of HEV ORF1-D248A
Supplementary Movie 6: AlphaFold Prediction of HEV ORF1-H249A

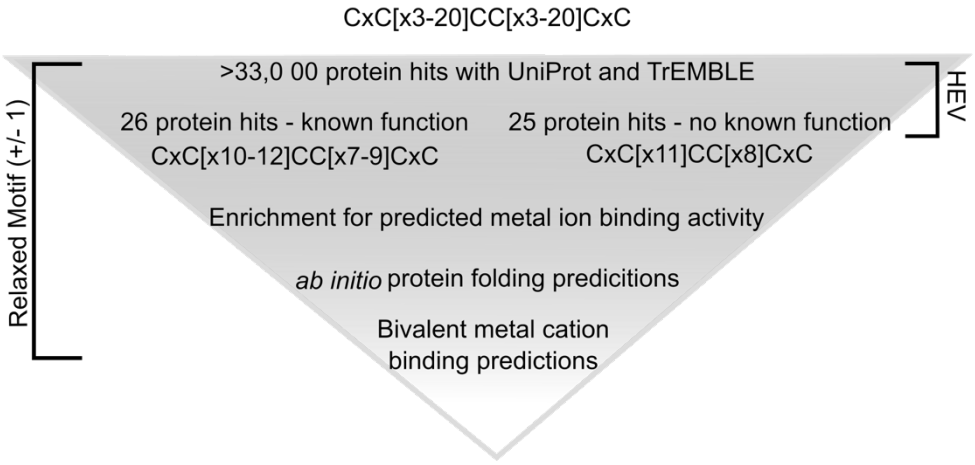
```
HEV gt 1                401 EHAQKFITRLYSWLFEKSGRDYIPGRQLEFYAQCRRWLSAGFHLDPRVLVFDESAPCHCRTAI-RKALSKFC  471
HEV gt 1 (SAR55)        400 EHAQKFITRLYSWLFEKSGRDYIPGRQLEFYAQCRRWLSAGFHLDPRVLVFDESAPCHCRTAI-RKAVSKFC  471
HEV gt 2                401 EHAQKFISRLYSWLFEKSGRDYIPGRQLQFYAQCRRWLSAGFHLDPRTLVFDESVPCSCRTTI-RRIAGKFC  471
HEV gt 3 (Kernow C1)    401 EHAQKFITRLYSWLFEKSGRDYIPGRQLQFYAQCRRWLSAGFHLDPRVLVFDESVPCRCRTFL-KKVAGKFC  471
HEV gt 3 (SHEV3)        401 EHAQKFITRLYSWLFEKSGRDYIPGRQLQFYAQCRRWLSAGFHLDPRVLVFDESVPCRCRTFL-KKVAGKFC  471
HEV gt 4                401 EHAQKFITRLYSWLFEKSGRDYIPGRQLQFYAQCRRWLSAGFHLDPRVLVFDEAAPCRCRSFL-RKAAHKFC  471
HEV gt 4 (TW4)          401 EHAQKFITRLYSWLFEKSGRDYIPGRQLQFYAQCRRWLSAGFHLDPRVLVFDESAPCRCRSFL-RKAAHKFC  471
HEV gt 5                401 EHAQKFITRLYSWLFEKSGRDYIPGRQLQFYAQCRRWLSAGFHLDPRVLVFDEAAPCRCRDLL-RRAAQKFC  471
HEV gt 6                401 EHAQRFVTRLYSWLFEKSGRDYIPGRQLQFYAQCRRWLSAGFHLDPRVLVFDEAAPCRCRDFL-RKAVKKFC  471
HEV gt 7                401 EHAQKFITRLYSWLFEKSGRDYIPGRQLEFYAQCRRWLSAGFHLDPRVLVFDEAASCRCRSSL-QKAVSRFC  471
HEV gt 8                401 EHAQKFITRLYSWLFEKSGRDYIPGRQVEFYAQCRRWLSAGFHLDPRVLVFDESAPCRCRSLL-RRAANSFC  471
RUBV                   1078 EHLATHFPLNHYSVLKPAEVRPPRGMCGSDMWRCRGW--HGMP-QVRCTPSNAHAA-LCRTGVPPRASTR-- 1143

HEV gt 1                472 CFMKWLGQECTCFLQPAEGAVGDQGHDNEAYEGSD-VDPA-ESAISDISGSYVVPGTALQ-PLYQALDLPAE  540
HEV gt 1 (SAR55)        471 CFMKWLGQECTCFLQPAEGVVGDQGHDNEAYEGSD-VDPA-ESAISDISGSYVVPGTALQ-PLYQALDLPAE  540
HEV gt 2                472 CFMKWLGQECSCFLQPAEGLAGDQGHDNEAYEGSD-VDTA-EPATLDITGSYIVDGRSLQ-TVYQALDLPAD  540
HEV gt 3 (Kernow C1)    472 CFMRWLGQECTCFLEPAEGLVGDHGHDNEAYEGSE-VDQA-EPAHLDVSGTYAVHGHQLV-ALYRALNVPHD  540
HEV gt 3 (SHEV3)        472 CFMRWLGQECTCFLEPAEGLVGDYGHDNEAYEGSE-VDPA-EPAHLDVSGTYAVHGRQLE-ALYRALNVPHD  540
HEV gt 4                472 CFMRWLGQDCTCFLQPIEGRVGEQGYDNEAFEGSD-VDPA-EEATVSISGSYIVTGRQLQ-PLYQALGIPSD  540
HEV gt 4 (TW4)          472 CFMRWLGQDCTCFLQPIEGRVGEQGYDNEAFEGSD-VDPA-EEATVSISGSYIVTGRQLQ-PLYQALGIPSD  540
HEV gt 5                472 CFMRWLGQECTCFLQPVEGRIGEQGYDNEAYEGSD-VDPA-EEVRVSVSGSYIVSGSQLQ-SYFKALGLPDD  540
HEV gt 6                472 CFMRWLGQECTCFLQPVEGHIGDQGHDNEAFEGSD-IDPA-EEVTFSISGTYIAAGNQLQ-PLYQALGIPAD  540
HEV gt 7                472 CFMRWLGQECNCFLQPVEGLIGDEGHDNEAYEGSE-VDPA-EPAAPDIDGIYTVSGAQLS-ALYRALGLPTD  540
HEV gt 8                472 CFMRWLGQECNCFLEPIEGHVGDEGHDNEAYEGSE-VDSA-DLAEPDPSGTFVVRGDQLL-PLFQALNLPSD  540
RUBV                   1144 G-GE-LDPN-TCWLRAA ANVAQAARACGAYTSAGCPKCAYGRALSEARTHEDFAALSQRWSASHADASPDGT 1211
                                 *

HEV gt 1                541 IVARAGRLTATV--KVSQV-DGR-IDC--ETLLGNKTFRTSFVD-GAVLETNG-PE---R--HNLSFDASQST 600
HEV gt 1 (SAR55)        540 IVARAGRLTATV--KVSQV-DGR-IDC--ETLLGNKTFRTSFVD-GAVLETNG-PE---R--HNLSFDASQST 599
HEV gt 2                541 LVARAARLSATV--TVTET-SGR-LDC--QTMIGNKTFLTTFVD-GARLEVNG-PE---Q--LNLSFDSQQCS 600
HEV gt 3 (Kernow C1)    541 IAARASRLTATV--ELVAG-PDR-LEC--RTVLGNKTFRTTVVD-GAHLEANG-PE---Q--YVLSFDASCQS 600
HEV gt 3 (SHEV3)        541 IAARASRLTATV--ELTAS-PDR-LEC--RTVLGNKTFRTTVVD-GAHLEANG-PE---Q--YVLSFDASRQS 600
HEV gt 4                541 LAARASRLAATV--EVSDA-EGR-LTC--KTILGNKTFSTVFTD-GAQLEANG-PE---Q--YVLSLDLTKQT 600
HEV gt 4 (TW4)          541 LAARASRLAATV--EVSDA-EGR-LTC--KTILGNKTFSTVFTD-GAQLEANG-PE---Q--YVLSFDLTKQT 600
HEV gt 5                541 LAARASRLTATV--EVDDT-DGH-FTC--RTVLGNKVFTTDFAD-GTVLESNG-PE---Q--YVLSFDSAKQC 600
HEV gt 6                541 LVARASRLAATV--EVEDT-NGR-LVC--RTTLGNKTFTTVFTD-GAELEANG-PE---T--YVLSFDSAKQC 600
HEV gt 7                541 LAARASRLVATV--EVTSH-PGR-IEC--KTTLGNKIFLTSFLD-GSRLEANG-PE---E--FILSFDADRQT 600
HEV gt 8                541 LVARASRLTATV--ELTVD-SNR-LDC--KTTLGNKVFRTTFID-GARLEANG-PE---Q--YVLSFDASKRS 600
RUBV                   1212 TGDPLDPLMETVGCACSRVWVGSEHEAPPDHLLVS-LHRAPNGPWGVVLEVRARPEGGNPTGHFVCAVGGGPR 1283
                                                                                      *
```

Supplementary Figure 1: Alignment of one or more representative strain(s) from each known HEV genotype putative PCP domain with RUBV protease reveals highly conserved octa-cysteine motif in HEV. Black – HEV Putative PCP. Brown - Upstream flanking 32 AAs – downstream flanking 8 AAs. (*) Catalytic dyad of RUBV protease at putative catalytic dyad positions within HEV putative PCP domain. Bold – 8 highly conserved cysteines across all known HEV genotypes. RUBV – Rubella virus. TW4 = HEV strain TW6196E. Alignments generated using ClustalW algorithm within the genome sequence viewing program SnapGene.

a



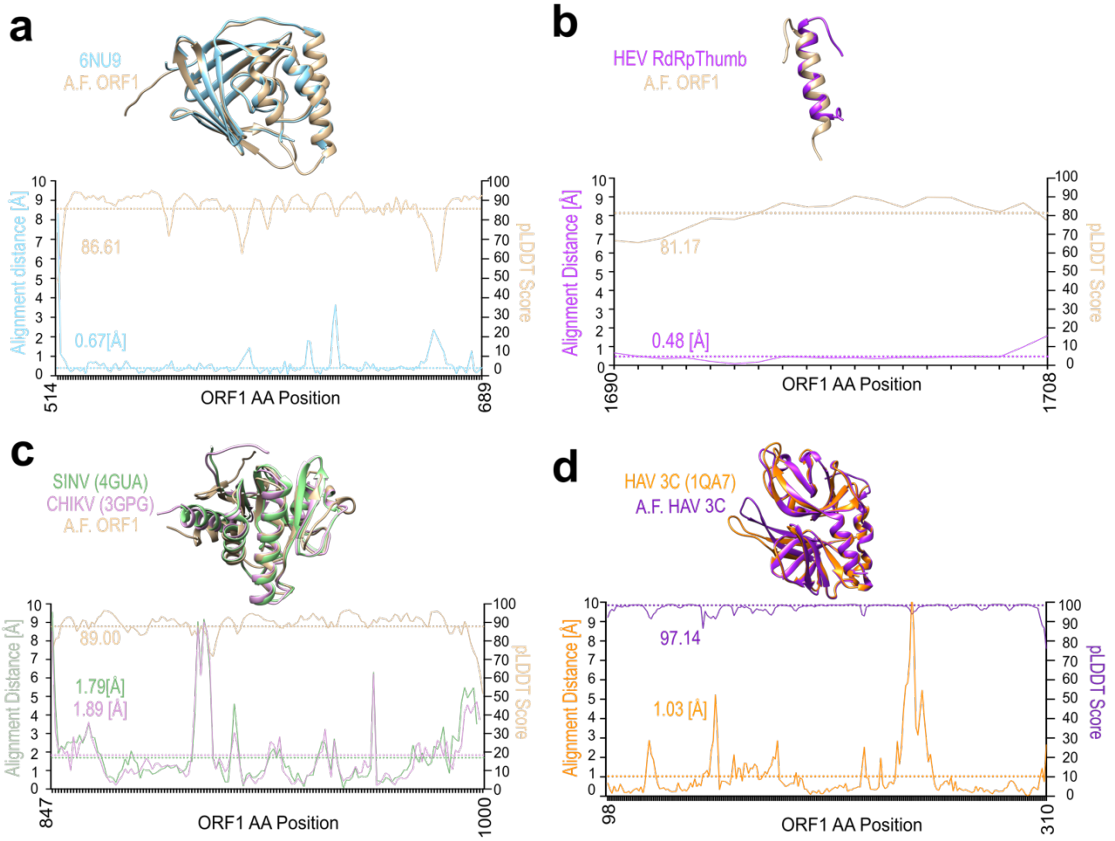
b

HEV Motif
CxC[x11]CC[x8]CxC
[(+/-)1]
CxC[10-12]CC[x7-9]CxC

Function	Gene name
Methyltransferase / transferase	ANM2
Colloidosmotic lysis	CR5AC, CRBA
Copper metallothioneins	MT2, MTCU
Calcium binding snake toxin	PA23
Zinc ion binding	TRIM42
FGF antagonist	SPRY2, SPRY3
Keratin Intermediate Filament	KRA53
GPCR	GG3

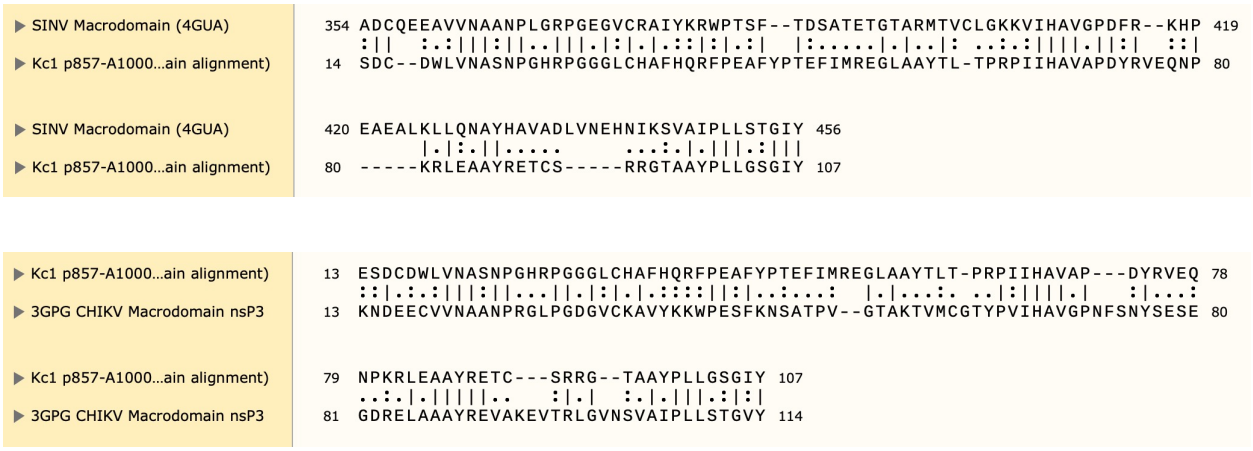
Function	Gene name
Cysteine-Rich Tail Protein	CRTP1
Colloidosmotic lysis	CR5AC, CRBA
Metallothioneins	MT MT1, MT2, MT3, MTA, MTB
Cell number regulator	CNR13
Zinc ion binding	TRIM42
Putative sodium transport	NPT2B
Keratin Intermediate Filament	KRA51 - KRA55, KRA58, SCGR8
GPCR	AGRF1

Supplementary Figure 2: Bioinformatic analysis (PROSITE) predicts CxC[X11]CC[X8]CxC motif (CxC motif) to be necessary for divalent metal ion coordination.
a, Bioinformatic pipeline depicting hexa-cysteine motif search across all deposited proteins in UniProt and TrEMBLE databases. Exact HEV motif search yielded 25 protein hits, all with unknown functions (right side). 11 amino acids upstream of core CC motif and 8 amino acid motif downstream of core CC were relaxed by (+/-) 1 to yield 26 protein hits, all with known functions and an enrichment for divalent ion binding activity. b, Utilizing motif searches with ScanProSite(de Castro et al., 2006) and TrEMBLE(Bairoch & Apweiler, 2000; Boeckmann et al., 2003; O'Donovan et al., 2002), we used a relaxed expression of the CxC motif where each stretch of [x] could be (+/-) 1: (CxC[x](10-12)CC[x](7-9)CxC). This analysis brought forth 26 proteins with known functions, enriched for proteins with divalent metal ion binding activity. Green: Proteins that matched the first stretch of [x 10-12]. Orange: Proteins that matched the second stretch of [x 7-9].



Supplementary Figure 3: AlphaFold predicts structured domains of HEV ORF1 and viral protease of hepatitis A virus (HAV) with high confidence.

a-c, Superimposition of AlphaFold (A.F.) prediction of corresponding Kernow C1/p6 ORF1 regions (Bronze) plus distance plot in (Å) and pLDDT score of A.F. prediction with: **a**, the crystal structure of HEV SAR-55 ORF1 amino acids 510-691 (Blue) (PDB: 6NU9)(Proudfoot et al., 2019). **b**, amphipathic "thumb" of HEV gt3 strain 83-2-27 RNA dependent RNA Polymerase (amino acids 1628-1647; ORF1 kc1/p6 residues 1684-1709) (Purple)(Oechslin et al., 2022). **c**, macro domains of Sindbis virus (Green) (SINV amino acids 1342 – 1509 of PDB: 4GUA)(Shin et al., 2012) and Chikungunya virus (Pink) (CHIKV) (PDB: 3GPG)(Malet et al., 2009). **d**, Superimposition of hepatitis a virus (HAV) 3C protease with corresponding A.F. prediction plus distance plot in (Å) and pLDDT score of A.F. prediction. Orange – HAV 3C protease (PDB: 1QA7)(Bergmann et al., 1999) and Purple – HAV 3C protease A.F. prediction. Hatched lines show averages of corresponding measurement.



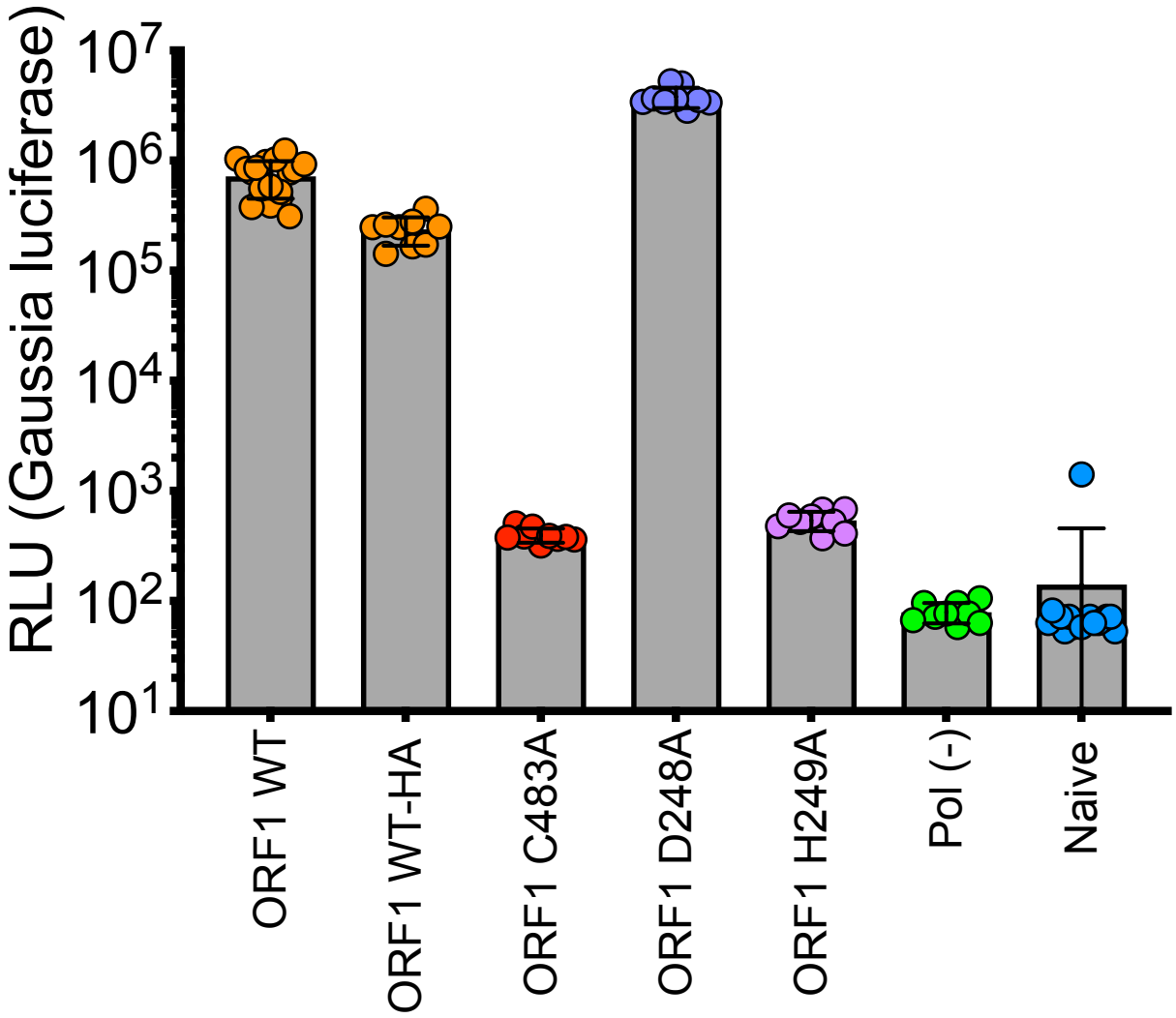
SINV vs. ORF1 macro domain

Matrix:	BLOSUM62
Gap open penalty:	10.0
Gap extend penalty:	1.0
Length:	104
Identity:	37 / 104 (35.58%)
Similarity:	59 / 104 (56.73%)
Gaps:	11 / 104 (10.58%)

CHIKV vs. ORF1 macro domain

Matrix:	BLOSUM62
Gap open penalty:	10.0
Gap extend penalty:	1.0
Length:	107
Identity:	39 / 107 (36.45%)
Similarity:	58 / 107 (54.21%)
Gaps:	17 / 107 (15.89%)

Supplementary Figure 4: Sequence alignments, percent similarity and identity of Kc1/p6 macrodomain with macrodomains of SINV and CHIKV. Alignments of ORF1 macrodomain (AAs 857-1000) with SINV macrodomain (AAs 1342 – 1509 of PDB: 4GUA)(Shin et al., 2012) (top, bottom left) and with CHIKV macrodomain (PDB: 3DPG)(Malet et al., 2009) (right, bottom right). Dash – gap in alignment. Colon - similar amino acid. Period – dissimilar amino acid. Line – identical amino acid.



Supplementary Figure 5: Replication end point analysis of WT-HA, D248A, and H249A Gluc constructs. HepG2C3A cells were transfected with *in vitro* transcribed RNA of WT, polymerase deficient, WT-HA-tagged, or point mutants in one of the upstream interacting Y-domain (D248A, H249A). Cell culture supernatants were collected for four days post transfection prior to *Gaussia* luciferase quantification.

Supplementary Table 1: Primer sequences for all constructs in this study.

Construct	Primer Name	Sequence
HEV Kernow-C1 p6 C483A-Gluc	PU-O-5702	CAGGAGTGCACCGCTTTCCTGGAGCCAGCCGAGGGTTTA
	PU-O-5703	TGGCTCCAGGAAAGCGGTGCACTCCTGCCCTAACCACCG
HEV Kernow-C1 p6 GAD-Gluc	PU-O-2711	CTGCCTTTAAGGGTGCTGATTCGGTGGTCCT
	PU-O-2712	AGGACCACCGAATCAGCACCCTTAAAGGCAG
pSAR55 C483A-Gluc	PU-O-5698	CAGGAGTGCACCGCTTTTCTACAACCTGCAGAAGGCGTCG
	PU-O-5699	AGGTTGTAGAAAAGCGGTGCACTCCTGGCCCAGCCACTTC
pSAR55 GAD-Gluc	PU-O-4635	CTGCCTTTAAAGGTGcTGATTCGATAGTGCT
	PU-O-4636	AGCACTATCGAATCAGCACCTTTAAAGGCAG
pGEM-9Zf-pSHEV3 C483A-Gluc	PU-O-5696	CAGGAGTGTACCGCTTTCTTGGAGCCAGCCGAAGGCTTG
	PU-O-5697	TGGCTCCAAGAAAGCGGTACACTCCTGCCCTAACCACCG
pGEM-9Zf-pSHEV3 GAD-Gluc	PU-O-4376	CCGCCTTCAAGGGTGcTGATTCGGTGGTCCT
	PU-O-4377	AGGACCACCGAATCAGCACCCTTGAAGGCGG
pGEM-7Zf(-)-TW6196E C483A-Gluc	PU-O-5700	CAAGACTGCACCGCTTTTCTTCAGCCTATTGAGGGGAGG
	PU-O-5701	AGGCTGAAGAAAAGCGGTGCAGTCTTGACCTAGCCACCG
pGEM-7Zf(-)-TW6196E GAD-Gluc	PU-O-4631	CGGCATTTAAAGGGGCTGACTCTGTTGTGCT
	PU-O-4632	AGCACAACAGAGTCAGCCCCTTTAAATGCCG
Kernow c1p6 C483S Gluc	PU-O-5893	GGGCAGGAGTGCACCTCCTTCCTGGAGCCAGCCG
	PU-O-5894	GGCTGGCTCCAGGAAGGAGGTGCACTCCTGCCCTAAC
Kernow c1p6 C483C (TGT) Gluc	PU-O-6273	GTGCACCTGTTTCCTGGAGCCAGCCGAGG
	PU-O-6274	AGGAAACAGGTGCACTCCTGCCCTAACCACC
Kernow c1p6 Y590A Gluc	PU-O-5790	GGCCCAGAGCAGGCTGTCCTGTCATTTGACGCCTCCTG
	PU-O-5791	CAAATGACAGGACAGCCTGCTCTGGGCCATTCGCTTCAAG
Kernow c1p6 Y590F Gluc	PU-O-5891	GAATGGCCCAGAGCAGTTTGTCCTGTCATTTGACGC
	PU-O-5892	GTCAAATGACAGGACAAACTGCTCTGGGCCATTCGC
Kernow c1p6 Y590L Gluc	PU-O-5238	GAATGGCCCAGAGCAGCTTGTCCTGTCATTTGAC
	PU-O-5239	GTCAAATGACAGGACAAGCTGCTCTGGGCCATTC
Kernow c1p6 C483G Gluc	PU-O-5240	GGGCAGGAGTGCACCGGCTTCCTGGAGCCAGCC
	PU-O-5241	GGCTGGCTCCAGGAAGCCGGTGCACTCCTGCCC
Kernow c1p6 C483A (GCT) Gluc	PU-O-6060	AGCGGTGCACTCCTGCCCTAACCACCG
	PU-O-6061	CAGGAGTGCACCGCTTTCCTGGAGCCA
Kernow c1p6 C483A (GCA) Gluc	PU-O-6290	CAGGAGTGCACCGCATTCCTGGAGCCA
	PU-O-6291	TGCGGTGCACTCCTGCCCTAACCACCG
Kernow c1p6 C483A (GCG) Gluc	PU-O-6292	CAGGAGTGCACCGCGTTCCTGGAGCCA
	PU-O-6293	CGCGGTGCACTCCTGCCCTAACCACCG
Kernow c1p6 C483A (GCC) Gluc	PU-O-6288	CAGGAGTGCACCGCCTTCCTGGAGCCA
	PU-O-6289	GGCGGTGCACTCCTGCCCTAACCACCG
Kernow c1p6 Q433A C434A R435A Gluc	PU-O-7667	CAGCTCCAGTTTTATGCAGCTGCAGCTCGGTGGCTATCTG
	PU-O-7668	CAGATAGCCACCGAGCTGCAGCTGCATAAAACTGGAGCTG
Kernow c1p6 R436A W437A L438A Gluc	PU-O-7669	GCACAGTGCCGAGCTGCAGCTTCTGCAGGATTCCATCTGG
	PU-O-7670	CCAGATGGAATCCTGCAGAAGCTGCAGCTCGGCACTGTGC
Kernow c1p6 S439A A440A G441A Gluc	PU-O-7671	CACAGTGCCGACGGTGGCTAGCTGCAGCTTTCCATCTGGAC
	PU-O-7672	GTCCAGATGGAAAGCTGCAGCTAGCCACCGTCGGCACTGTG
Kernow c1p6 F442A H443A L444A Gluc	PU-O-7673	GGCTATCTGCAGGAGCTGCAGCTGACCCCAGGGTGCTTG
	PU-O-7674	CAAGCACCCTGGGGTCAGCTGCAGCTCCTGCAGATAGCC
Kernow c1p6 D445A P446A R447A Gluc	PU-O-7675	GCAGGATTCCATCTGGCTGCAGCTGTGCTTGTTTTTGATG
	PU-O-7676	CATCAAAAACAAGCACAGCTGCAGCCAGATGGAATCCTGC
Kernow c1p6 V448A L449A V450A Gluc	PU-O-7677	CAGGATTCCATCTGGACCCCAGGGGCTGCAGCTTTTGATGAATCAG
	PU-O-7678	CTGATTCATCAAAAGCTGCAGCCCTGGGGTCCAGATGGAATCCTG
Kernow c1p6 F451A D452A E453A Gluc	PU-O-7679	CCCCAGGGTGCTTGTTGCTGCAGCTTCAGTGCCATGTCG
	PU-O-7680	CGACATGGCACTGAAGCTGCAGCAACAAGCACCCTGGGG
Kernow c1p6 S454A V455A P456A Gluc	PU-O-7681	GGGTGCTTGTTTTTGATGAAGCTGCAGCTTGTCGTTGTAGG
	PU-O-7682	CCTACAACGACAAGCTGCAGCTTCATCAAAAACAAGCACCC
Kernow c1p6 C457A R458A C459A Gluc	PU-O-7683	GATGAATCAGTGCCAGCTGCAGCTAGGACGTTCCTGAAG
	PU-O-7684	CTTCAGGAACGTCCTAGCTGCAGCTGGCACTGATTCATC
Kernow c1p6 R460A T461A F462A Gluc	PU-O-7685	CAGTGCCATGTCGTTGTGCTGCAGCTCTGAAGAAAGTCGC
	PU-O-7686	GCGACTTTCTTCAGAGCTGCAGCACAACGACATGGCACTG
Kernow c1p6 L463A K464A K465A Gluc	PU-O-7687	CGTTGTAGGACGTTCGCTGCAGCTGTCGCGGGTAAATTC
	PU-O-7688	GAATTTACCCGCGACAGCTGCAGCGAACGTCCTACAACG
Kernow c1p6 V466A G468A Gluc	PU-O-7689	CCTGAAGAAAGCTGCAGCTAAATTCTGCTGTTTTATGCGG
	PU-O-7690	CCGCATAAAACAGCAGAATTTAGCTGCAGCTTTCTTCAGG
Kernow c1p6 K469A F470A C471A Gluc	PU-O-7691	GAAGAAAGTCGCGGGTGCTGCAGCTTGTTTTATGCGGTGG
	PU-O-7692	CCACCGCATAAAACAAGCTGCAGCACCCGCGACTTTCTTC
Kernow c1p6 C472A F473A M474A Gluc	PU-O-7693	GAAAGTCGCGGGTAAATTCTGCGCTGCAGCTCGGTGGTTAG
	PU-O-7694	CTAACCACCGAGCTGCAGCGCAGAATTTACCCGCGACTTTC
Kernow c1p6 R475A W476A L477A Gluc	PU-O-7695	CTGCTGTTTTATGGCTGCAGCTGGGCAGGAGTGCACCTGC
	PU-O-7696	GCAGGTGCACTCCTGCCCAGCTGCAGCCATAAAACAGCAG
Kernow c1p6 G478A Q479A E480A Gluc	PU-O-7697	GTTTTATGCGGTGGTTAGCTGCAGCTTGCACCTGCTTCC
	PU-O-7698	GGAAGCAGGTGCAAGCTGCAGCTAACCACCGCATAAAAC
Kernow c1p6 C481A T482A C483A Gluc	PU-O-7699	GGTTAGGGCAGGAGGCTGCAGCTTTCCTGGAGCCAGCCG
	PU-O-7700	CGGCTGGCTCCAGGAAAGCTGCAGCCTCCTGCCCTAACC
Kernow c1p6 484A L485A E486A Gluc	PU-O-7701	GGGCAGGAGTGCACCTGCGCTGCAGCTCCAGCCGAGG
	PU-O-7702	CTCGGCTGGAGCTGCAGCGCAGGTGCACTCCTGCCC
Kernow c1p6 P487A E489A Gluc	PU-O-7703	GCACCTGCTTCCTGGAGGCTGCAGCTGGTTTAGTTGGCG
	PU-O-7704	CGCCAACTAAACCAGCTGCAGCCTCCAGGAAGCAGGTGC
Kernow c1p6 G490A L491A V492A Gluc	PU-O-7705	CCTGGAGCCAGCCGAGGCTGCAGCTGGCGACCATGGCC
	PU-O-7706	GGCCATGGTCGCCAGCTGCAGCCTCGGCTGGCTCCAGG
Kernow c1p6 G493A D494A H495A Gluc	PU-O-7707	CAGCCGAGGGTTTAGTTGCTGCAGCTGGCCATGACAATG
	PU-O-7708	CATTGTCATGGCCAGCTGCAGCAACTAAACCCTCGGCTG
Kernow c1p6 G496A H497A D498A Gluc	PU-O-7709	GTTTAGTTGGCGACCATGCTGCAGCTAATGAGGCTTATG
	PU-O-7710	CATAAGCCTCATTAGCTGCAGCATGGTCGCCAACTAAAC
Kernow c1p6 N499A E500A Gluc	PU-O-7711	GGCCATGACGCTGCAGCTTATGAAGGTTCTGAGGTCGACC
	PU-O-7712	GGTCGACCTCAGAACCTTCATAAGCTGCAGCGTCATGGCC

Kernow c1p6 Y502A E503A G504A Gluc	PU-O-7713	CCATGACAATGAGGCTGCTGCAGCTTCTGAGGTCGACCAG
	PU-O-7714	CTGGTCGACCTCAGAAGCTGCAGCAGCCTCATTGTCATGG
Kernow c1p6 S505A E506A V507A Gluc	PU-O-7715	GGCTTATGAAGGTGCTGCAGCTGACCAGGCTGAACCTGCC
	PU-O-7716	GGCAGGTTCAGCCTGGTCAGCTGCAGCACCTTCATAAGCC
Kernow c1p6 D508A Q509A Gluc	PU-O-7717	GAAGGTTCTGAGGTCGCTGCAGCTGAACCTGCCCATC
	PU-O-7718	GATGGGCAGGTTCAGCTGCAGCGACCTCAGAACCTTC
Kernow c1p6 E511A P512A Gluc	PU-O-7719	GAGGTCGACCAGGCTGCTGCAGCTCATCTTGATGTTTCGG
	PU-O-7720	CCGAAACATCAAGATGAGCTGCAGCAGCCTGGTCGACCTC
Kernow c1p6 H514A L515A D516A Gluc	PU-O-7721	CAGGCTGAACCTGCCGCTGCAGCTGTTTCGGGGACTTATG
	PU-O-7722	CATAAGTCCCCGAAACAGCTGCAGCGGCAGGTTCAGCCTG
Kernow c1p6 V517A S518A G519A Gluc	PU-O-7723	CCTGCCCATCTTGATGCTGCAGCTACTTATGCCGTCCACG
	PU-O-7724	CGTGGACGGCATAAGTAGCTGCAGCATCAAGATGGGCAGG
Kernow c1p6 T520A Y521A Gluc	PU-O-7725	GCCCATCTTGATGTTTCGGGGGCTGCAGCTGTCCACGGGC
	PU-O-7726	GCCCGTGGACAGCTGCAGCCCCCGAAACATCAAGATGGGC
Kernow c1p6 V523A H524A G525A Gluc	PU-O-7727	GTTTCGGGGACTTATGCCGCTGCAGCTCACCAGCTTGTAG
	PU-O-7728	CTACAAGCTGGTGAGCTGCAGCGGCATAAGTCCCCGAAAC
Kernow c1p6 H526A Q527A L528A Gluc	PU-O-7729	GCCGTCCACGGGGCTGCAGCTGTAGCCCTCTATAGGGC
	PU-O-7730	GCCCTATAGAGGGCTACAGCTGCAGCCCCGTGGACGGC
Kernow c1p6 V529A L531A Gluc	PU-O-7731	CCACGGGCACCAGCTTGCTGCAGCTTATAGGGCACTTAATG
	PU-O-7732	CATTAAGTGCCCTATAAGCTGCAGCAAGCTGGTGCCCGTGG
Kernow c1p6 Y532A R533A Gluc	PU-O-7733	CCAGCTTGTAGCCCTCGCTGCAGCTCTTAATGTCCCAC
	PU-O-7734	GTGGGACATTAAGAGCTGCAGCGAGGGCTACAAGCTGG
Kernow c1p6 L535A N536A V537A Gluc	PU-O-7735	GCCCTCTATAGGGCAGCTGCAGCTCCACATGATATTGCCG
	PU-O-7736	CGGCAATATCATGTGGAGCTGCAGCTGCCCTATAGAGGGC
Kernow c1p6 P538A H539A D540A Gluc	PU-O-7737	GGGCACTTAATGTCGCTGCAGCTATTGCCGCTCGAGC
	PU-O-7738	GCTCGAGCGGCAATAGCTGCAGCGACATTAAGTGCCC
Kernow c1p6 I541A Gluc	PU-O-7739	GTCCCACATGATGCTGCAGCTCGAGCTTCCCGATTAACGG
	PU-O-7740	CCGTTAATCGGGAAGCTCGAGCTGCAGCATCATGTGGGAC
Kernow c1p6 R544A S546A Gluc	PU-O-7741	CCCACATGATATTGCCGCTGCTGCAGCTCGATTAACGGC
	PU-O-7742	GCCGTTAATCGAGCTGCAGCAGCGGCAATATCATGTGGG
Kernow c1p6 R547A L548A T549A Gluc	PU-O-7743	GCCGCTCGAGCTTCCGCTGCAGCTGCTACTGTTGAGC
	PU-O-7744	GCTCAACAGTAGCAGCTGCAGCGGAAGCTCGAGCGGC
Kernow c1p6 T551A V552A Gluc	PU-O-7745	CCCGATTAACGGCTGCAGCTGAGCTTGTTGCAGGTCCAG
	PU-O-7746	CTGGACCTGCAACAAGCTCAGCTGCAGCCGTTAATCGGG
Kernow c1p6 E553A L554A V555A Gluc	PU-O-7747	CGGCTACTGTTGCTGCAGCTGCAGGTCCAGACCGCTTGG
	PU-O-7748	CCAAGCGGTCTGGACCTGCAGCTGCAGCAACAGTAGCCG
Kernow c1p6 G557A P558A Gluc	PU-O-7749	CGGCTACTGTTGAGCTTGTTGCTGCAGCTGACCGCTTGGAG
	PU-O-7750	CTCCAAGCGGTCAGCTGCAGCAACAAGCTCAACAGTAGCCG
Kernow c1p6 D559A R560A L561A Gluc	PU-O-7751	GCTTGTTGCAGGTCCAGCTGCAGCTGAGTGCCGCACTGTGC
	PU-O-7752	GCACAGTGCGGCACTCAGCTGCAGCTGGACCTGCAACAAGC
Kernow c1p6 E562A C563A R564A Gluc	PU-O-7753	GGTCCAGACCGCTTGGCTGCAGCTACTGTGCTCGG
	PU-O-7754	CCGAGCACAGTAGCTGCAGCCAAGCGGTCTGGACC
Kernow c1p6 T565A V566A L567A Gluc	PU-O-7755	CCGCTTGGAGTGCCGCACTGTGCTCGGTAATAAGACCTTC
	PU-O-7756	GAAGGTCTTATTACCGAGCACAGTGCGGCACTCCAAGCGG
Kernow c1p6 G568A N569A K570A Gluc	PU-O-7757	GTGCCGCACTGTGCTCGCTGCAGCTACCTTCCGGACGACG
	PU-O-7758	CGTCGTCCGGAAGGTAGCTGCAGCGAGCACAGTGCGGCAC
Kernow c1p6 T571A F572A R573A Gluc	PU-O-7759	GCTCGGTAATAAGGCTGCAGCTACGACGGTGGTTGATGGC
	PU-O-7760	GCCATCAACCACCGTCGTAGCTGCAGCCTTATTACCGAGC
Kernow c1p6 T574A T575A V576A Gluc	PU-O-7761	GTAATAAGACCTTCCGGGCTGCAGCTGTTGATGGCGCCC
	PU-O-7762	GGGCGCCATCAACAGCTGCAGCCCGGAAGGTCTTATTAC
Kernow c1p6 V577A D578A G579A Gluc	PU-O-7763	GTAATAAGACCTTCCGGACGACGGTGGCTGCAGCTGCCCATCTTG
	PU-O-7764	CAAGATGGGCAGCTGCAGCCACCGTCGTCCGGAAGGTCTTATTAC
Kernow c1p6 H581A L582A Gluc	PU-O-7765	CGGTGGTTGATGGCGCTGCAGCTGAAGCGAATGGCCCAG
	PU-O-7766	CTGGGCCATTCGCTTCAGCTGCAGCGCCATCAACCACCG
Kernow c1p6 E583A N585A Gluc	PU-O-7767	GATGGCGCCCATCTTGCTGCAGCTGGCCCAGAGCAGTATG
	PU-O-7768	CATACTGCTCTGGGCCAGCTGCAGCAAGATGGGCGCCATC
Kernow c1p6 G586A P587A E588A Gluc	PU-O-7769	CCCATCTTGAAGCGAATGCTGCAGCTCAGTATGTCCTGTC
	PU-O-7770	GACAGGACATACTGAGCTGCAGCATTCGCTTCAAGATGGG
Kernow c1p6 Q589A Y590A V591A Gluc	PU-O-7771	GCGAATGGCCCAGAGGCTGCAGCTCTGTCATTTGACGCC
	PU-O-7772	GGCGTCAAATGACAGAGCTGCAGCCTCTGGGCCATTCGC
Kernow c1p6 Y590A V591A L592A Gluc	PU-O-7773	GGCCCAGAGCAGGCTGCAGCTTCATTTGACGCCTCCTGTC
	PU-O-7774	GACAGGAGGCGTCAAATGAAGCTGCAGCCTGCTCTGGGCC
Kernow c1p6 C434A Gluc	PU-O-7775	CCAGTTTTATGCACAGGCCCGACGGTGGCTATCTGCAGG
	PU-O-7776	CCTGCAGATAGCCACCGTCGGGCCTGTGCATAAAACTGG
Kernow c1p6 C457A Gluc	PU-O-7777	GATGAATCAGTGCCAGCTCGTTGTAGGACGTTCCTGAAG
	PU-O-7778	CTTCAGGAACGTCCTACAACGAGCTGGCACTGATTCATC
Kernow c1p6 C459A Gluc	PU-O-7779	GAATCAGTGCCATGTCGTGCTAGGACGTTCCTGAAG
	PU-O-7780	CTTCAGGAACGTCCTAGCACGACATGGCACTGATTC
Kernow c1p6 C471A Gluc	PU-O-7781	GAAAGTCGCGGGTAAATTCGCCTGTTTTATGCGGTGGTTAG
	PU-O-7782	CTAACCACCGCATAAAACAGGCGAATTTACCCGCGACTTTC
Kernow c1p6 C472A Gluc	PU-O-7783	GTCGCGGGTAAATTCTGCGCTTTTATGCGGTGGTTAGGG
	PU-O-7784	CCCTAACCACCGCATAAAAGCGCAGAATTTACCCGCGAC
Kernow c1p6 C481A Gluc	PU-O-7785	GCGGTGGTTAGGGCAGGAGGCCACCTGCTTCCTGGAGCC
	PU-O-7786	GGCTCCAGGAAGCAGGTGGCCTCCTGCCCTAACCACCGC
Kernow c1p6 C563A Gluc	PU-O-7787	GCAGGTCCAGACCGCTTGGAGGCCCGCACTGTGCTCGG
	PU-O-7788	CCGAGCACAGTGCGGGCCTCCAAGCGGTCTGGACCTGC
Kernow c1p6 D248A Gluc	PU-O-10292	CTACTAAAATAGTTGGTGCCCACCCGTTGGTTATAGAGCGTG
	PU-O-10293	CCAACGGGTGGGCACCAACTATTTTAGTAGTACGGATCC
Kernow c1p6 H249A Gluc	PU-O-10294	CGTACTACTAAAATAGTTGGTGACGCCCCCGTTGGTTATAGAGC
	PU-O-10295	GCCCGCACACGCTCTATAACCAACGGGGCGTCACCAACTATTTTAG
AAAPG-HA tag-AAAPG from pLVX ORF2-HA	PU-O-9180	GCTGCCCCGGGGGCGGCCGCACCAGGAtacccatacgat
	PU-O-9183	GGTAGGGTGGGGCAGTCCTGGTGCGGCCGCagcgtaatctg
PCR linearize p6/BSR-2A-ZsGreen insert AAAPG-HA tag-AAAPG	PU-O-9229	CTGCCCCACCCTACCCCGCCTGCTAGTGATATTTGGGCGTTAC
	PU-O-9228	CCCCCGGGGCAGCAGCCGAAGCAGGGGCGG
PCR linearize Kernow C1/p6 Gluc insert AAAPG-HA tag-AAAPG	PU-O-9229	CTGCCCCACCCTACCCCGCCTGCTAGTGATATTTGGGCGTTAC
	PU-O-9228	CCCCCGGGGCAGCAGCCGAAGCAGGGGCGG
ORF1 AAAPG-HA tag-AAAPG	PU-O-9907	CGGTGAATTCCTCGAGGGCAGACCACGTATGTGGTCGATGG
	PU-O-9908	CCGCGGCCGCTCTAGAGGTGATCCCATGGGCGATGCA

	Residue 1:	Residue 2:	Length: (Å)
ORF1 WT	C471	C472	4.233
	C471	C481	4.808
	C471	C483	3.327
	C472	C481	8.305
	C472	C483	7.193
	C481	C483	2.047
	C457	C459	2.047
	C457	D248	5.704
	C457	H249	4.623
	C459	H249	4.599
	C459	D248	3.738
	D248	H249	5.705
ORF1 C483A			
	C471	C472	5.851
	C471	C481	5.286
	C472	C481	10.734
	C457	C459	2.048
	C457	D248	5.376
	C457	H249	4.577
	C459	H249	4.453
	C459	D248	3.496
	D248	H249	6.088
ORF1 C563A			
	C471	C472	5.263
	C471	C481	3.525
	C471	C483	3.676
	C472	C481	7.819
	C472	C483	7.976
	C481	C483	2.052
	C457	C459	2.045
	C457	D248	7.162
	C457	H249	4.571
	C459	H249	3.914
	C459	D248	5.239
	D248	H249	6.056

	Residue 1:	Residue 2:	Length: (Å)
ORF1 D248A			
	C471	C472	5.168
	C471	C481	2.987
	C471	C483	3.178
	C472	C481	7.042
	C472	C483	6.774
	C481	C483	3.759
	C457	C459	2.063
	C457	H249	5.053
	C459	H249	4.824
ORF1 H249A			
	C471	C472	5.588
	C471	C481	3.426
	C471	C483	4.582
	C472	C481	5.117
	C472	C483	4.699
	C481	C483	2.079
	C457	C459	2.055
	C457	D248	4.446
	C459	D248	3.867

Supplementary Table 2: Predicted bond lengths (Å) of cysteines within HEV hexa-cysteine motif and D248/H249 of upstream Y-domain. Residue distances between the sulfur atom of the side chains of the cysteine, the 2nd oxygen of the aspartic acid side chain, or the 2nd nitrogen on the side chain of the histidine. Atomic distances were calculated with UCSF Chimera (Pettersen et al., 2004) structure analysis functionality. Orange – bond length beyond biological relevance (residues were considered interacting if within 3Å, as determined by (Zheng et al., 2008)).

Supplementary Movies:

All supplementary movies show the best ranked model of their respective ORF1 (WT or mutant) predicted with the AlphaFold algorithm. Video 1 exhibits the ORF1-WT AlphaFold prediction, then transitions to a visual representation of the alanine scanning mutagenesis data across the pPCP. Color gradation corresponds fold change of WT of alanine triplet. Below (-2) fold change is orange. Between (-2) and (-1) fold change is sea green. Between (-1) and 0 fold change is cornflower blue, Above WT replication levels is dark blue. Videos 2-6 exhibits the conserved hexa-cysteine motif (magenta), interacting Y-domain residues D248 and H249 (cornflower blue), methyltransferase domain (orange), putative PCP (green), helicase (yellow), RNA dependent RNA polymerase (red), and putative intracellular membrane association site (cyan). Hatched yellow lines – measured distances between sulfur atoms of cysteines, 2nd oxygen of aspartic acid, or second nitrogen of histidine. Orientation of each video is with the putative membrane association site pointing up, with the pseudo zinc-finger comprised of cysteines within the hexa-cysteine motif on the right.

Supplementary References:

Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*, *28*(1), 45-48. https://doi.org/10.1093/nar/28.1.45

Bergmann, E. M., Cherney, M. M., McKendrick, J., Frormann, S., Luo, C., Malcolm, B. A., Vederas, J. C., & James, M. N. (1999). Crystal structure of an inhibitor complex of the 3C proteinase from hepatitis A virus (HAV) and implications for the polyprotein processing in HAV. *Virology*, *265*(1), 153-163. https://doi.org/10.1006/viro.1999.9968

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., & Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, *31*(1), 365-370. https://doi.org/10.1093/nar/gkg095

de Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., & Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res*, *34*(Web Server issue), W362-365. https://doi.org/10.1093/nar/gkl124

Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X., & Canard, B. (2009). The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. *J Virol*, *83*(13), 6534-6545. https://doi.org/10.1128/JVI.00189-09

O'Donovan, C., Martin, M. J., Gattiker, A., Gasteiger, E., Bairoch, A., & Apweiler, R. (2002). High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform*, *3*(3), 275-284. https://doi.org/10.1093/bib/3.3.275

Oechslin, N., Da Silva, N., Szkolnicka, D., Cantrelle, F. X., Hanoulle, X., Moradpour, D., & Gouttenoire, J. (2022). Hepatitis E virus RNA-dependent RNA polymerase is involved in RNA replication and infectious particle production. *Hepatology*, *75*(1), 170-181. https://doi.org/10.1002/hep.32100

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, *25*(13), 1605-1612. https://doi.org/10.1002/jcc.20084

Proudfoot, A., Hyrina, A., Holdorf, M., Frank, A. O., & Bussiere, D. (2019). First Crystal Structure of a Nonstructural Hepatitis E Viral Protein Identifies a Putative Novel Zinc-Binding Protein. *J Virol*, *93*(13). https://doi.org/10.1128/JVI.00170-19

Shin, G., Yost, S. A., Miller, M. T., Elrod, E. J., Grakoui, A., & Marcotrigiano, J. (2012). Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *Proc Natl Acad Sci U S A*, *109*(41), 16534-16539. https://doi.org/10.1073/pnas.1210418109

Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., & Minor, W. (2008). Data mining of metal ion environments present in protein structures. *J Inorg Biochem*, *102*(9), 1765-1776. https://doi.org/10.1016/j.jinorgbio.2008.05.006