Intrinsic Disorder Mediates Hepatitis C Virus Core – Host Cell Protein Interactions

Patrick T. Dolan, Andrew P. Roth, Bin Xue, Ren Sun, A. Keith Dunker, Vladimir N. Uversky, and Douglas J. LaCount

SUPPLEMENTARY MATERIAL

Introduction

The Supplementary Material includes additional discussions of the current status of computational methods for identifying IDR and MoRF and the results of the phylogentic analysis of the HCV MoRFs. We also include an expanded methods section, a table that summarizes the functions of the human proteins that interacted with HCV Core in this study, a list of primers used to clone HCV Core and NS5A fragments, and Supplementary Figure 1.

C*omputational algorithms for IDR and MoRF identification* – Efforts to assess the prevalence and function of intrinsic disorder and MoRFs within proteomes have yielded over 50 bioinformatic predictors of disorder (reviewed in [1]) and several MoRF predictors [2-5]. The earlier predictors [2] focused on MoRFs that tend to form helices upon binding (called α-MoRFs). However, MoRFs of various secondary structures and even combinations of secondary structures have been observed [6]. General predictors have been developed to identify sequences that form irregular or "random coil" forms upon binding as well as those that form helix, sheet, or combinations thereof [5,7]. Other predictors of disorder [7] and protein binding sites have also been developed, including several others that use linear sequence motifs for identifying binding sites [7-10]. These motif-based methods do not explicitly consider structure of the conserved motifs, but these linear binding motifs [11] are mostly located in regions of predicted disorder [12,13].

To identify IDRs and MoRFs within the HCV polyprotein, we employed the Predictors of Naturally Disordered Regions (PONDR®), PONDR®-VLXT [14] and the "meta-predictor", PONDR®-FIT [15]. The PONDR®-FIT meta-predictor combines three predictors from the PONDR® family with three additional predictors of disorder, namely IUPred [16], FoldIndex [17], and TOP-IDP [18]. This meta-predictor performs

better than any of the individual predictors, providing one of the best predictions of disorder across a protein sequence [15]. PONDR®-VLXT yields better sensitivity in predicting alpha-helical propensity and, in combination with the α-MoRF-Pred computational tool, can identify putative α-MoRFs [4]. Although several additional predictors have been developed to identify likely MoRFs within longer regions of disorder, we chose this particular predictor because of previous success with this approach.

Despite the abundance of computational tools, very few studies have integrated binding site predictions to help identify and characterize protein-protein interaction binding interfaces. When PONDR®-VLXT disorder prediction was applied to the RNA degradosome, the C-terminal half of the protein was strongly predicted to be disordered but with three different short predicted regions of structure, called "regions of increased structural propensity." These three regions were each shown to bind to different protein partners (12). Consistent with the computational prediction, one of these short regions was partially helical in the X-ray crystal structure of the complex with enolase [19]. In a second study, computational analyses of amino acid sequences of the proteins involved in protein-protein interaction networks were used to identify likely linear motifs [20]. Two previously untested linear motifs were then shown experimentally to bind with their indicated partners [20]. Neither of these prior studies carried out experiments to test whether the same MoRF or linear motif could bind to more than one partner as shown herein.

Core MoRFs are undergoing purifying selection – Observing stronger negative, or purifying, selection within the predicted MoRFs would provide independent evidence of a functional role for these features in HCV biology. To assess the levels of purifying selection acting on specific partitions of the HCV Core protein, we used Bayesian phylogenetic inference to estimate the site-specific ratios of non-synonymous and synonymous substitution rates, dN/dS, from Core coding sequences from 86 human HCV isolates representing genotypes 1-7. Each sequence was divided into five partitions corresponding to the two MoRFs, the non-MoRF IDR, the entire IDR (domain I), and the structured C-terminal region (domain II). Mean dN/dS values were calculated for each sequence partition from each sampled tree and the

distributions of the means were used to determine the 95 and 99% Bayesian confidence intervals (Table 1). MoRF1 and 2 had significantly lower dN/dS values than the surrounding IDR ($P<0.01$ for both), indicating that they have undergone stronger purifying selection (Fig. 2C). This analysis also revealed that the structured C-terminus of Core, which mediates the interaction with cytoplasmic lipid droplets (residues 118-191 in JFH1 Core), has undergone stronger purifying selection than the Core IDR (residues 1-117 in HCV 2a JFH1 Core).

The picture for NS5A is more complicated than for Core. Overall, there is greater sequence variation in the NS5A IDRs, with many insertions and deletions that complicate the alignment (Fig. 2B). As with Core, 76 NS5A sequences from all 7 HCV genotypes were divided into MoRF and non-MoRF partitions. Estimates of the dN/dS ratios for the NS5A MoRFs showed that MoRFs 1 and 3 were significantly lower than the flanking IDR ($P<0.01$ for both) (Fig. 2D and Table 2). The dN/dS estimates for NS5A MoRF2 and 5 suggested that they are undergoing purifying selection ($P<0.05$ for both) but were not significantly lower than the dN/dS ratio estimate for the flanking IDR. Surprisingly, the dN/dS estimates for the MoRF 4 partition indicated that it is undergoing diversifying selection ($dN/dS >1$, $P<0.01$).

Supplementary Methods

Yeast two-hybrid assays – Yeast two-hybrid library screens were performed as described in [21-24]. Briefly, MoRF-containing fragments of the HCV JFH1 polyprotein (UNIPROT: Q99IB8) were cloned into the yeast two-hybrid DNA-binding domain (DBD) plasmid pOBD2 by homologous recombination in the yeast strain R2HMet (*MAT*a *ura3-52 ade2-101 trp1-901 leu2-3,112 his3-200 met2*Δ*::hisG gal4*Δ *gal80*Δ) [24,25]. The primers used to generate the fragments are shown in Table 1. Expression of all DBD constructs generated in this study was confirmed by western blot as described in [26]. All constructs were verified by PCR and sequencing. Yeast two-hybrid assays were performed by mating yeast expressing the DBD fusion proteins with yeast strain BK100 (*MATa ura3-52 ade2-101 trp1-901 leu2-3,112 his3-200 gal4*Δ *gal80*Δ *GAL2-ADE2 LYS2::GAL1-HIS3 met2::GAL7-lacZ*, a derivative of PJ69-4A) [24,27,28] that contained a human gene fragment cloned into the activation domain (AD) plasmid pOAD.103 [21,29]. Diploid yeast

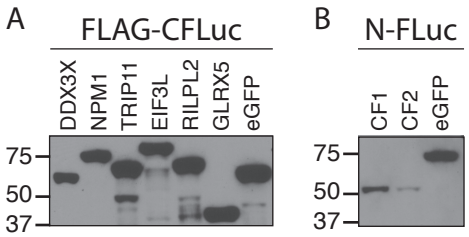
were selected on synthetic dropout (SD) media lacking tryptophan and leucine (SD-TL), then plated on SD media lacking tryptophan, leucine, uracil, and histidine, plus 3-amino-1,2,4-triazole at concentrations empirically determined to inhibit background growth (Y2H selection media).

Split-luciferase assays – HCV Core fragments 1 and 2 were cloned into p424-BYDV-NFLuc, whereas each cellular binding partner of Core MoRF1 and MoRF2 were cloned into p424-BYDV-C-Fluc-FLAG [30]. EGFP was cloned into both vectors and served as a negative control to determine the background in the assay for each fusion protein. Fusion proteins were expressed in TNT® SP6 High-Yield Wheat Germ Lysate (Promega), subjected to SDS PAGE and western blotting with anti-FLAG (Sigma, F1804) or anti-NFLuc (Santa Cruz Biotechnology, sc-57603) antibodies. Based on the abundance of the proteins on the western blots, the concentrations of the samples were normalized to the EGFP controls. For split-luciferase assays, MoRF1, MoRF2, and EGFP were combined with the cellular binding partners or EGFP as a negative control. After incubating 1 hour at 4°C, binding reactions were combined with luciferase substrate solution and luminescence was assessed. Luminescence for each reaction was compared to that of each control. For an interaction to be considered significant, the luminescent signal must be higher than both control reactions (NFLuc-Test + CFLuc-EGFP and NFLuc-EGFP + CFLuc-human protein) as determined by a one-tailed t-test ($P<0.05$).

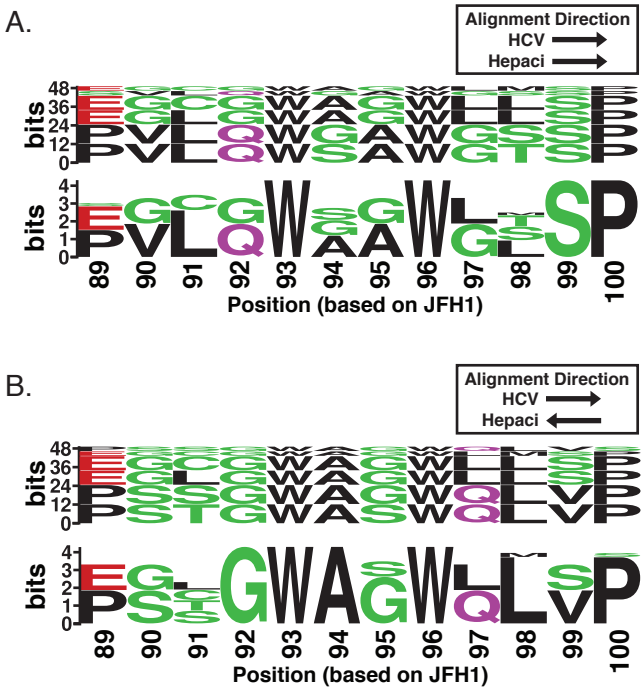
Error-prone mutagenesis – Error-prone mutagenesis of the MoRF-containing fragments of the HCV Core protein were generated as described in [31,32]. Nucleotide analogs 8-oxo-2' deoxyguanosine (8-oxo-dGTP) and 6-(2-deoxy-b-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-c][1,2]oxazin-7-one (dP) were added to standard GoTaq PCR reactions (Promega) at concentrations of 1 and 10uM. The mutagenic reactions were run for five cycles, after which the products were purified to remove unincorporated analogs and used as templates for another 25 cycles. PCR products were inserted into pOBD2 by homologous recombination.

Supplementary Table 1. The human binding partners of HCV Core. Description of the functions of Core-interacting proteins and any previously identified roles in HCV infection. Start and end refer to the amino acid positions of the fragments identified in the yeast two-hybrid screens.

Core Fragment	Human Protein	Protein Name	Start	End	Function
HCV Core 1 (Res. 1-76)	**DDX3X**	Homo *sapiens DEAD* (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked (DDX3X)	465	632	ATP-dependent RNA helicase implicated in transcription, translation, nuclear shuttling and antiviral responses. Binds to Core and required for HCV infection [33-39].
	NPM1	*Homo sapiens* nucleophosmin (nucleolar phosphoprotein B23, numatrin) (NPM1)	95	206	Multimeric chaperone primarily found in the nucleolus. Involved in ribosome biogenesis, centrosome duplication, histone assembly, cell proliferation, and the regulation of tumor suppressors TP53/p53 and ARF. Binds HCV Core [40].
	TRIP11	*Homo sapiens* thyroid hormone receptor interactor 11 (TRIP11)	1533	1757	Associated with the cis-Golgi. Proposed role in Golgi ribbon formation and cis-cisternae fusion.
HCV Core 2 (Res. 36-117)	**EIF3L**	*Homo sapiens* eukaryotic translation initiation factor 3, subunit L (EIF3L)	123	301	Component of the EIF3 complex that regulates translation. Specifically involved in IRES-mediated translation. EIF3L-containing EIF3 binds to the HCV IRES [41].
	GLRX5	*Homo sapiens* glutaredoxin 5 (GLRX5)	41	154	Iron-sulfur cluster assembly protein involved in cellular iron homeostasis [42].
	RILPL2	*Homo sapiens* Rab interacting lysosomal protein-like 2 (RILPL2)	1	196	Involved in cytoplasmic trafficking and cell morphology. Adapter protein for non-muscle Myosin Va [43-45]. Required for HCV infection [22].



Supplementary Fig. 1. Western analysis of split-luciferase fusion proteins. FLAG- and CFLuc tagged human proteins (A) and NFLuc-HCV Core (B) fusion proteins were in vitro translated in wheat germ extracts (Promega). Equal volumes of each sample were subjected to SDS PAGE followed by immunoblotting with anti-FLAG or anti-NFLuc antibodies. The concentrations of fusion proteins were adjusted based on the western blot results so that the test protein concentrations were equal to (or slightly less) than the GFP negative controls.



Supplementary Fig. 2. A potential inverted linear motif in Core MoRF2. Alignment of the reversed NPHV sequences with HCV MoRF2 yielded the consensus sequence GWAxWxLxP, which may represent an inverted protein-binding site in Core. Such sites, which have been termed retro-MoRFs, have been proposed [46], but to date very few such retro-MoRFs have been functionally characterized [47]. **(**A) Block Logo and Web Logo of Core residues 89-100 (based on HCV JFH1 numbering) from HCV genotypes 1-6 (GTs) and 8 Hepacivirus isolates. The Block Logo (A, top) shows the frequency of specific peptide sequences, with the relative height of each peptide reflecting its frequency in the input alignment. In contrast, the Web Logo shows the conservation of residues at each position. (B) Block Logo and Web Logo of Core residues 89-100 with the same region from the Hepaciviruses inverted. The Web Logo of the HCV and the inverted Hepacivirus sequences revealed a conserved linear motif at position 92-100, GWAxWxLxP.

SUPPLEMENTAL REFERENCES:

1. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK. Predicting intrinsic disorder in proteins: an overview. Cell Res 2009;19(8):929-49.
2. Garner E, Romero P, Dunker AK, Brown C, Obradovic Z. Predicting Binding Regions within Disordered Proteins. Genome Inform Ser Workshop Genome Inform 1999;10:41-50.
3. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry 2007;46(47):13468-77.
4. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. Biochemistry 2005;44(37):12454-70.
5. Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, Kurgan L. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics 2012;28(12):i75-83.
6. Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, Uversky VN. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006;362(5):1043-59.
7. Meszaros B, Simon I, Dosztanyi Z. Prediction of protein binding regions in disordered proteins. PLoS Comput Biol 2009;5(5):e1000376.
8. Davey NE, Cowan JL, Shields DC, Gibson TJ, Coldwell MJ, Edwards RJ. SLiMPrints: conservation-based discovery of functional motif fingerprints in intrinsically disordered protein regions. Nucleic Acids Res 2012;40(21):10628-41.
9. Obenauer JC, Yaffe MB. Computational prediction of protein-protein interactions. Methods Mol Biol 2004;261:445-68.
10. Obenauer JC, Cantley LC, Yaffe MB. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 2003;31(13):3635-41.
11. Dinkel H, Van Roey K, Michael S, Davey NE, Weatheritt RJ, Born D, Speck T, Kruger D, Grebnev G, Kuban M and others. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res 2014;42(Database issue):D259-66.
12. Fuxreiter M, Tompa P, Simon I. Local structural disorder imparts plasticity on linear motifs. Bioinformatics 2007;23(8):950-6.
13. Meszaros B, Dosztanyi Z, Simon I. Disordered binding regions and linear motifs--bridging the gap between two models of molecular recognition. PLoS One 2012;7(10):e46829.
14. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins 2001;42(1):38-48.
15. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 2010;1804(4):996-1010.
16. Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005;21(16):3433-4.
17. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, Sussman JL. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 2005;21(16):3435-8.
18. Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, Dunker AK. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. Protein Pept Lett 2008;15(9):956-63.
19. Chandran V, Luisi BF. Recognition of enolase in the Escherichia coli RNA degradosome. J Mol Biol 2006;358(1):8-15.

20. Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB. Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol 2005;3(12):e405.
21. Khadka S, Vangeloff AD, Zhang C, Siddavatam P, Heaton NS, Wang L, Sengupta R, Sahasrabudhe S, Randall G, Gribskov M and others. A physical interaction network of dengue virus and human proteins. Mol Cell Proteomics 2011;10(12):M111 012187.
22. Dolan PT, Zhang C, Khadka S, Arumugaswami V, Vangeloff AD, Heaton NS, Sahasrabudhe S, Randall G, Sun R, LaCount DJ. Identification and comparative analysis of hepatitis C virus-host cell protein interactions. Mol Biosyst 2013;9(12):3199-209.
23. LaCount DJ. Interactome mapping in malaria parasites: challenges and opportunities. Methods Mol Biol 2012;812:121-45.
24. LaCount DJ, Vignali M, Chettier R, Phansalkar A, Bell R, Hesselberth JR, Schoenfeld LW, Ota I, Sahasrabudhe S, Kurschner C and others. A protein interaction network of the malaria parasite Plasmodium falciparum. Nature 2005;438(7064):103-7.
25. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P and others. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 2000;403(6770):623-7.
26. Kushnirov VV. Rapid and reliable protein extraction from yeast. Yeast 2000;16(9):857-60.
27. James P, Halladay J, Craig EA. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. Genetics 1996;144(4):1425-36.
28. Vignali M, McKinlay A, LaCount DJ, Chettier R, Bell R, Sahasrabudhe S, Hughes RE, Fields S. Interaction of an atypical Plasmodium falciparum ETRAMP with human apolipoproteins. Malar J 2008;7:211.
29. Lee S, Salwinski L, Zhang C, Chu D, Sampankanpanich C, Reyes NA, Vangeloff A, Xing F, Li X, Wu TT and others. An integrated approach to elucidate the intra-viral and viral-cellular protein interaction networks of a gamma-herpesvirus. PLoS Pathog 2011;7(10):e1002297.
30. Brown HF, Wang L, Khadka S, Fields S, LaCount DJ. A densely overlapping gene fragmentation approach improves yeast two-hybrid screens for Plasmodium falciparum proteins. Mol Biochem Parasitol 2011;178(1-2):56-9.
31. Rasila TS, Pajunen MI, Savilahti H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment. Anal Biochem 2009;388(1):71-80.
32. Zaccolo M, Williams DM, Brown DM, Gherardi E. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J Mol Biol 1996;255(4):589-603.
33. Mamiya N, Worman HJ. Hepatitis C virus core protein binds to a DEAD box RNA helicase. J Biol Chem 1999;274(22):15751-6.
34. Owsianka AM, Patel AH. Hepatitis C virus core protein interacts with a human DEAD box protein DDX3. Virology 1999;257(2):330-40.
35. You LR, Chen CM, Yeh TS, Tsai TY, Mai RT, Lin CH, Lee YH. Hepatitis C virus core protein interacts with cellular putative RNA helicase. J Virol 1999;73(4):2841-53.
36. Li Q, Brass AL, Ng A, Hu Z, Xavier RJ, Liang TJ, Elledge SJ. A genome-wide genetic screen for host factors required for hepatitis C virus propagation. Proc Natl Acad Sci U S A 2009;106(38):16410-5.
37. Angus AG, Dalrymple D, Boulant S, McGivern DR, Clayton RF, Scott MJ, Adair R, Graham S, Owsianka AM, Targett-Adams P and others. Requirement of cellular DDX3 for hepatitis C virus replication is unrelated to its interaction with the viral core protein. J Gen Virol 2010;91(Pt 1):122-32.
38. Ariumi Y, Kuroki M, Abe K, Dansako H, Ikeda M, Wakita T, Kato N. DDX3 DEAD-box RNA helicase is required for hepatitis C virus RNA replication. J Virol 2007;81(24):13922-6.

39. Randall G, Panis M, Cooper JD, Tellinghuisen TL, Sukhodolets KE, Pfeffer S, Landthaler M, Landgraf P, Kan S, Lindenbach BD and others. Cellular cofactors affecting hepatitis C virus infection and replication. Proc Natl Acad Sci U S A 2007;104(31):12884-9.
40. Mai RT, Yeh TS, Kao CF, Sun SK, Huang HH, Wu Lee YH. Hepatitis C virus core protein recruits nucleolar phosphoprotein B23 and coactivator p300 to relieve the repression effect of transcriptional factor YY1 on B23 gene expression. Oncogene 2006;25(3):448-62.
41. Zhou M, Sandercock AM, Fraser CS, Ridlova G, Stephens E, Schenauer MR, Yokoi-Fong T, Barsky D, Leary JA, Hershey JW and others. Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. Proc Natl Acad Sci U S A 2008;105(47):18139-44.
42. Wingert RA, Galloway JL, Barut B, Foott H, Fraenkel P, Axe JL, Weber GJ, Dooley K, Davidson AJ, Schmid B and others. Deficiency of glutaredoxin 5 reveals Fe-S clusters are required for vertebrate haem synthesis. Nature 2005;436(7053):1035-39.
43. Lise MF, Srivastava DP, Arstikaitis P, Lett RL, Sheta R, Viswanathan V, Penzes P, O'Connor TP, El-Husseini A. Myosin-Va-interacting protein, RILPL2, controls cell shape and neuronal morphogenesis via Rac signaling. J Cell Sci 2009;122(Pt 20):3810-21.
44. Schaub JR, Stearns T. The Rilp-like proteins Rilpl1 and Rilpl2 regulate ciliary membrane content. Mol Biol Cell 2013;24(4):453-64.
45. Wei Z, Liu X, Yu C, Zhang M. Structural basis of cargo recognitions for class V myosins. Proc Natl Acad Sci U S A 2013;110(28):11314-9.
46. Xue B, Dunker AK, Uversky VN. Retro-MoRFs: identifying protein binding sites by normal and reverse alignment and intrinsic disorder prediction. Int J Mol Sci 2010;11(10):3725-47.
47. Feng S, Chen JK, Yu H, Simon JA, Schreiber SL. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. Science 1994;266(5188):1241-7.