Hepatitis C virus infection protein network

B de Chassey1,2, V Navratil2,3,9, L Tafforeau1,2,9, MS Hiet1,2,10, A Aublin-Gex1,2, S Aagugué1,2,11, G Meiffren1,2, F Pradeyronski1,2, BF Farès1,2, T Chantier1,2, M Le Breton1,2, J Pellet1,2, N Davoust1,2, PE Mangeot1,2, A Chaboud2,4, F Penin2,4, Y Jacob5, PO Vidailhet6, M Vidal7, P André1,2,8, C Rabourdin-Combe1,2 and V Lotteau1,2,8,*

1 IMAP Team, Inserm Unit 851, Lyon, France, 2 IFR128 BioSciences Lyon-Gerland, Université de Lyon, Lyon, France, 3 INRA UMR 754, rétrovirus et pathologie comparée, Lyon, France, 4 Institut de Biologie et Chimie des Protéines, CNRS UMR 5086, Lyon, France, 5 Unité Postulant de Génétique, Papillomavirus et Cancer Humain, Institut Pasteur, Paris, France, 6 Laboratoire de Génomique Virale et Vaccination, Institut Pasteur, CNRS URA 3015, Paris, France, 7 Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA and 8 Hospices Civils de Lyon, Hôpital de la Croix-Rousse, Laboratoire de virologie, Lyon, France
9 These authors contributed equally to this work
10 Present address: Department for Molecular Virology, University of Heidelberg, Heidelberg 69120, Germany.
* Corresponding author. IMAP Team, Inserm Unit 851, 21, Av. T. Garnier, Lyon 69007, France. Tel.: +33 437 282 412; Fax: +33 437 282 341; E-mail: vincent.lotteau@inserm.fr

Received 22.4.08; accepted 30.9.08

Introduction

Hepatitis C virus (HCV) infection is characterized by a high rate of chronicity and concerns 170 millions of individuals worldwide. Chronicity infected patients present liver injury essentially mediated by immune mechanisms and metabolic disorders associated with hepatic steatosis, fibrogenesis and insulin resistance to various extent (Negro, 2006; Moradpour et al., 2007). Long-term infected patients have a high risk of developing cirrhosis and hepatocarcinoma, but despite considerable efforts, molecular basis of HCV pathology remains poorly understood. HCV genome is a positive-strand RNA of 9.6 kb encoding a polyprotein that is post-translationally processed into structural (CORE, E1, E2 and p7) and non-structural (NS2, NS3, NS4A, NS4B, NS5A and NS5B) proteins (Appel et al., 2006). HCV variants have been classified into six genotypes with biological and antigenic differences. Whereas infection by all genotypes is associated with insulin resistance and fibrosis, a correlation between hepatic steatosis severity and viral replication is preferentially observed for genotype 3. Genotypic differences also correlate with interferon sensitivity, with genotypes 2 and 3 responding better to combined interferon and ribavirine therapy. We focused here on HCV genotype 1b, which is associated with insulin resistance, fibrosis, mild steatosis and poor sensitivity to treatment (Lonardo et al., 2004; Strader et al., 2004).

The rapidly growing knowledge of protein–protein interaction (PPI) networks (interactome) for human, model organisms and host–pathogen begins to provide network-based models for diseases. In a network approach, viral pathogenesis can be viewed as the expression of new constraints on the protein network imposed by the virus when connecting to the cellular interactome. Identification of topological and functional properties that are lost or deregulated, or that emerged in the ‘infection network’, becomes a major challenge for a systems understanding of viral infection (Tan et al., 2007).

High-throughput yeast two-hybrid (Y2H) screens of human cDNA library (Calderwood et al., 2007) and computation-based...
analysis (Uetz et al., 2006) have been used previously to study Epstein–Barr virus (EBV), Kaposi sarcoma herpes virus and varicella zoster virus interactions with host cell factors. Analysis of virus–human protein inter-interactome network revealed that host interactors tend to be enriched in proteins that are highly connected in the cellular network (Calderwood et al., 2007; Dyer et al., 2008). These hub proteins are thought to be essential for the normal cell functioning and during pathogenesis.

Several laboratories have joined their efforts to develop infection mapping project (I-MAP). The goal of I-MAP is to provide a comprehensive view of viral infections at the protein level by mapping the interactions of a large number of viral proteins with host proteins. Screening and mapping have been designed to address specific questions, such as virulence/attenuation, species barrier, identification of therapeutic targets, chronicity and the risk of cancer development.

Here, a proteome-wide mapping approach of interactions between HCV and cellular proteins was performed to provide a comprehensive view of viral infection (Figure 1A). A viral ORFeome was first generated that included ORFs encoding all full-length mature proteins and several protein domains of genotype 1b strain (Supplementary Figure S1). These viral baits were screened against human cDNA libraries using a highly stringent Y2H assay (IMAP Y2H data set). Together with interactions extensively mined and curated from the literature (IMAP LCI data set), this comprehensive host–virus infection network was integrated into a reconstructed human protein–protein interactome. Analysis of the ‘infection network’ (V-HHCV; Figure 1A) revealed topological features of cellular interactors and identified functional pathways related to viral biology and pathogenesis.

Results and discussion

Construction of an HCV–human interactome map

A comprehensive interactome map between HCV and cellular proteins was generated by Y2H screens. Twenty-seven constructs encoding full-length HCV mature proteins or discrete domains were cloned using a recombination-based cloning system (Walhout et al., 2000) (Supplementary Figure S1).

**Results and discussion**

**Construction of an HCV–human interactome map**

A comprehensive interactome map between HCV and cellular proteins was generated by Y2H screens. Twenty-seven constructs encoding full-length HCV mature proteins or discrete domains were cloned using a recombination-based cloning system (Walhout et al., 2000) (Supplementary Figure S1).
Four independent screens were performed with each HCV bait protein, probing two distinct human cDNA libraries, either by mating (IMAP1 screens) or by transformation (IMAP2 screens; see Materials and methods). Fetal brain and spleen cDNA libraries were used instead of a liver library because the liver is known to overexpress a large number of secreted proteins, which could interfere with the quality of the screens. Comparing EST data from fetal brain and spleen with EST data from liver revealed that 87% of genes expressed in the liver are also expressed in brain or spleen (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene). A total of 314 HCV–human PPIs were identified, involving 278 human proteins (Figure 1B, IMAP Y2H data set in Supplementary Table S1). More than 90% of the cellular interactors identified are expressed in the liver. Pairwise interactions between HCV and human proteins were also extracted from the literature by automatic text mining and checked by expert curation (see Materials and methods; IMAP LCI data set in Supplementary Table S1). A total of 135 PPI were extracted from Pubmed and 89 were extracted from BIND database (Bader et al., 2003) (Figure 1B). The resulting HCV–human interactome is thus composed of 481 PPIs with 65% new interactions, involving 11 HCV proteins and 421 distinct human proteins (Figure 1B). IMAP1 and IMAP2 screens share 22 interactions (7% of IMAP Y2H data set). This overlap is in the range of previously reported data using two Y2H high-throughput screening methods (Lim et al., 2006) and suggests that despite screens characteristics, summarized in Supplementary Table SII, saturation has not been reached. Differences in the screening methods, such as sensitivity of the different yeast strains to selective drugs, differential growth rate of colonies and low penetration of interaction phenotype, could account for this observation. The low redundancy between IMAP Y2H and IMAP LCI data sets may also emphasize a higher false-negative rate of the Y2H system, which would be in agreement with recent studies (Rual et al., 2005; Huang et al., 2007). An interesting hypothesis is that different methods of screening may lead to the exploration of different spaces of the HCV–human interactome. As false-positives may also contribute to the weak overlap of IMAP1 and IMAP2, two validation methods were used to assess the confidence of the IMAP Y2H data set. Two-thirds of the data set were retested by direct Y2H between viral protein baits and cellular protein prey identified by our Y2H screens (Y2H pairwise matrices). From the remaining interactions, 26 PPIs (25%) were retested by co-affinity purification and 22 PPIs could be validated (validation rate: 85%; Figure 1C and Supplementary Table S1). This Y2H data set was thus of very high confidence for further analysis at the topological and functional levels. In Table I, the top 21 new interactions validated by GST pull-down experiments and identified in one or two screens are shown. Analysis of the HCV-infection network (V-HHCV, Figure 1A) showed that NS3, NS5A and CORE are the most connected proteins, with 214, 96 and 76 cellular partners, respectively, highlighting the potential multifunctionality of these proteins during infection (Supplementary Table S1, Figure 1D). Highly interacting proteins are known to be significantly more disordered than low-degree (LD) proteins (Haynes et al., 2006). Interestingly, NS3, NS5A and CORE are the only HCV proteins predicted to contain at least one intrinsic disordered region, according to DISOPRED2 (Ward et al., 2004) (prediction of protein disorder server; data not shown). This correlates well with the high degree (HD) of

<table>
<thead>
<tr>
<th>HCV protein</th>
<th>Human gene name</th>
<th>Human protein official full name</th>
<th>Biological process</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS3</td>
<td>CALCOCO2</td>
<td>Calcium binding and coiled-coil domain 2</td>
<td>May have a function in viral life cycles</td>
</tr>
<tr>
<td>NS3</td>
<td>EIF4ENIF1</td>
<td>Eukaryotic translation initiation factor 4E nuclear import factor 1</td>
<td>Protein nuclear import</td>
</tr>
<tr>
<td>NS3</td>
<td>FRS3</td>
<td>Fibroblast growth factor receptor substrate 3</td>
<td>FGF receptor signalling pathway</td>
</tr>
<tr>
<td>NS3</td>
<td>CCDC21</td>
<td>Coiled-coil domain containing 21</td>
<td>Unknown</td>
</tr>
<tr>
<td>NS3</td>
<td>BCKD2</td>
<td>Branched-chain ketoacid dehydrogenase kinase</td>
<td>Branched-chain family amino-acid catabolic process</td>
</tr>
<tr>
<td>NS3</td>
<td>GBP2</td>
<td>Guanylate-binding protein 2, interferon-inducible</td>
<td>Immune response</td>
</tr>
<tr>
<td>NS3</td>
<td>KPN1</td>
<td>Karyopherin alpha 1 (importin alpha 5)</td>
<td>NLS-bearing substrate import to the nucleus</td>
</tr>
<tr>
<td>NS3</td>
<td>PSMB9</td>
<td>Proteasome subunit, beta type, 9</td>
<td>Ubiquitin-dependent protein catabolic process</td>
</tr>
<tr>
<td>NS3</td>
<td>RASL2</td>
<td>Ras protein activator-like</td>
<td>Regulation of small GTPase-mediated signal transduction</td>
</tr>
<tr>
<td>NS3</td>
<td>SMURF2</td>
<td>SMAD-specific E3 ubiquitin ligase 2</td>
<td>Regulation of TGFβ receptor signalling pathway</td>
</tr>
<tr>
<td>NS5A</td>
<td>EFEMP1</td>
<td>EGF-containing fibulin-like extracellular matrix protein 1</td>
<td>Integrity of fascia connective tissues</td>
</tr>
<tr>
<td>NS5A</td>
<td>GOLGA2</td>
<td>Golgi autoantigen, golgin subfamily a, 2</td>
<td>Vesicular transport</td>
</tr>
<tr>
<td>NS5A</td>
<td>GPS2</td>
<td>G protein pathway suppressor 2</td>
<td>JNK cascade</td>
</tr>
<tr>
<td>NS5A</td>
<td>ITGAL</td>
<td>Integrin, alpha L (antigen CD11A (p180), lymphocyte function-associated antigen 1; alpha polypeptide)</td>
<td>Integrin-mediated signalling pathway</td>
</tr>
<tr>
<td>NS5A</td>
<td>MOBK1B</td>
<td>MOB1, Mps one binder kinase activator-like 1 (yeast)</td>
<td>Cell cycle progression</td>
</tr>
<tr>
<td>NS5A</td>
<td>NAP1L2</td>
<td>Nucleosome assembly protein 1-like</td>
<td>Nucleosome assembly</td>
</tr>
<tr>
<td>NS5A</td>
<td>PPP1R13 L</td>
<td>Protein phosphatase 1, regulatory (inhibitor) subunit 13-like</td>
<td>Regulation of transcription, DNA-dependent</td>
</tr>
<tr>
<td>NS5A</td>
<td>PSMB9</td>
<td>Proteasome subunit, beta type, 9</td>
<td>Ubiquitin-dependent protein catabolic process</td>
</tr>
<tr>
<td>NS5A</td>
<td>SORBS2</td>
<td>Sorbin and SH3 domain containing 2</td>
<td>Cytoskeletal adaptor activity</td>
</tr>
<tr>
<td>NS5A</td>
<td>TXNDC11</td>
<td>Thioredoxin domain containing 11</td>
<td>Cell redox homeostasis</td>
</tr>
<tr>
<td>NS5A</td>
<td>VPS52</td>
<td>Vacuolar protein sorting 52 homolog (S. cerevisiae)</td>
<td>Vesicular transport</td>
</tr>
</tbody>
</table>

HCV proteins are referenced according to their NCBI mature peptide product name (column 1). Human proteins are referenced with their cognate NCBI gene name or their protein official full name with their main biological function (column 2, 3 and 4, respectively). The 21 interactions have been validated by GST pull-down. In the first three entries, interactions were identified in two screens. In the rest of the entries, interactions were identified in one screen. Interaction Btn1–NS5A has been removed from this list as already described in literature.
these proteins. In addition, 45 cellular proteins are targeted by more than one viral protein, suggesting their essentiality for virus biology (Calderwood et al., 2007) (Supplementary Table SIII).

A human PPI network (H–H network; Figure 1A) was reconstructed from eight databases (Gandhi et al., 2006) (see Materials and methods). This network is composed of 44,223 non-redundant PPIs between 9,520 different proteins (Figure 2A, complete list of PPIs in Supplementary Table SIV), corresponding to 30% of the human proteome (the remaining proteins have no known cellular partners and can therefore not be included in this network). Interestingly, human proteins targeted by HCV (HHCV) are clearly over-represented in this H–H network (IMAP Y2H data set: 76%; IMAP LCI data set: 88%, exact Fisher test, \( P \text{-value} < 2.2 \times 10^{-10} \)). This suggests that HCV preferentially targets host proteins already known to be engaged in protein–protein interactions (Rual et al., 2005; Stelzl et al., 2005). For the IMAP LCI data set, the higher percentage of HHCV integrated in the human interactome may be explained by inspection bias of well-studied proteins and biological pathways. Analysis of HHCV–HHCV subnetwork (all connected HHCV proteins) showed that cellular proteins interacting with HCV are significantly more interconnected than expected for random subnetworks (Figures 1A and 2B, Supplementary methods). Indeed, the 338 HHCV integrated into the human interactome are distributed into 131 connected components (versus 276 expected by random subnetworks; \( z \)-score-based test \( P \text{-value} < 10^{-10} \), Supplementary Table SV). The largest one is composed of 196 HHCV (versus 18 expected by random subnetworks; \( z \)-test, \( P \text{-value} < 10^{-10} \)) and 127 are disconnected proteins. The three remaining connected components comprised two proteins. Two contained functionally related proteins (CLEC4M and CD209 are lectins involved in viral entry (Lozach et al., 2003); MVP and PARP4 are involved in Vault complex (Kedersha and Rome, 1986)) and one contained proteins not known to be functionally linked (KIAA1549 and CADPS).

**Topological analysis of the HCV–human interaction network**

To assess how HCV proteins interplay with the cellular protein network, we next focused on the centrality measures of HHCV proteins integrated into the H–H interactome. Local (degree) and global (shortest path length and betweenness) centrality measures were calculated. Briefly, the degree \( k \) of a protein in a network corresponds to its number of direct partners and is therefore a measure of local centrality. Betweenness \( b \) is a global measure of centrality, as it measures the number of shortest paths (the minimum distance between two proteins in the network, \( l \)) that pass through a given protein. To provide an unbiased analysis, calculations were done on the basis of the 213 HHCV from the IMAP Y2H data set integrated in the human interactome. The average degree, betweenness and shortest path length of the H–H network are 9.3, 1.6 \( \times 10^{-4} \) and 4.04,

---

**Figure 2** Graphical representation of the HCV–human interaction network. (A) Graphical representation of H–H network. Each node represents a protein and each edge represents an interaction. Red and blue nodes are HHCV and HNot-HCV, respectively. (B) Graphical representation of V–HHCV interaction network. Black node: viral protein; red node: human protein; red edge: interaction between human and viral proteins (V–HHCV); blue edge: interaction between human proteins (HHCV–HHCV). The largest component containing 196 proteins is represented in the middle of the network. Names of cellular proteins belonging to the three other connected components are also represented.
HHCV (red) is given for LD and HD classes. Bottom: mean degree of H Not-HCV (blue) and HHCV (red) is given for LD and HD classes. The conventional standard error the average degree cutoff. (Supplementary Table SV). To ensure that the preferential attachment to central HHCV was not due to inherent bias associated to false positive in the H–H interactome, we performed the same analysis with a high-confidence, but less comprehensive, human interactome (Supplementary Table SV). This trend was maintained with this data set, confirming that HHCV are highly central within the human interactome, both locally and globally, and appear relatively close to each other in this network. For comparative analysis of HCV and EBV, the centrality measures were also computed for H HCV infection protein network

respectively, which is in good agreement with previous reports (Ramirez et al, 2007) (Figure 3A). As the distribution of properties such as node degree and node betweenness in PPI networks appear to follow a power law, summarizing values by their distributions appears more appropriated for comparative analysis (Goh et al, 2002; Joy et al, 2005). The degree distribution of H HCV and of the human interactome are significantly distinct (U-test P-value < 10^{-3}), with an average degree of H HCV higher than the average degree of the human interactome (15.6 versus 9.3). The comparison of degree probability distribution reveals that H HCV are preferentially represented in all class above the mean degree (Figure 3B, left). This indicates that HCV proteins have a strong tendency to interact with highly connected cellular proteins. However, as degree measures only local connectivity of proteins, global characteristics that could reflect information exchange and propagation in the network were investigated (Hernandez et al, 2007). At a global scale, the betweenness distribution of H HCV and of the human interactome are significantly distinct (U-test P-value < 10^{-3}), with an average betweenness of H HCV higher than the average betweenness of the human interactome (3.8 × 10^{-4} versus 1.6 × 10^{-4}). As for the degree, the comparison of betweenness probability distribution shows an excess of H HCV in all class above the mean betweenness (Figure 3B, right). In addition, the shortest path length distribution of H HCV and of the human interactome were found significantly distinct (U-test P-value < 10^{-5}), with an average shortest path length of H HCV lower than average shortest path length of the human interactome (3.50 versus 4.04) revealing the topological proximity of H HCV. Both local and global centrality of H HCV from the IMAP LCI data set were higher than for the IMAP Y2H data set, emphasizing the problem of literature inspection bias and reinforcing the unbiased approach of Y2H screening (Supplementary Table SV). To ensure that the preferential attachment to central H HCV was not due to inherent bias associated to false positive in the H–H interactome, we performed the same analysis with a high-confidence, but less comprehensive, human interactome (Supplementary Table SV). This trend was maintained with this data set, confirming that H HCV are highly central within the human interactome, both locally and globally, and appear relatively close to each other in this network. For comparative analysis of HCV and EBV, the centrality measures were also computed for H HCV infection protein network

<table>
<thead>
<tr>
<th>Data set</th>
<th>Proteins</th>
<th>ppl</th>
<th>k</th>
<th>b (10^{-4})</th>
<th>l</th>
</tr>
</thead>
<tbody>
<tr>
<td>H in H–H network</td>
<td>9520</td>
<td>44223</td>
<td>9.3</td>
<td>1.6</td>
<td>4.04</td>
</tr>
<tr>
<td>H HCV in H–H network</td>
<td>213</td>
<td>72</td>
<td>15.6</td>
<td>3.8</td>
<td>3.50</td>
</tr>
</tbody>
</table>

Figure 3  Topological analysis of the HCV–human interaction network. (A) Topological analysis of H and H HCV in H–H network. Average values of degree (k), betweenness (b) and shortest path length (l) for all human proteins and for H HCV from the IMAP Y2H data set. (B) Degree and betweenness distribution of H and H HCV proteins in H–H network. P(k) is the probability of a node to connect k other nodes in the network. P(b) is the probability of a node to have a betweenness equal to b in the network. Normalized log degree (left) and log betweenness (right) distribution of H (blue) and H HCV proteins (red). Solid line represents linear regression fit. Vertical dashed lines give mean degree and betweenness values. Each class is represented with conventional standard error. (C) Degree and betweenness correlation of H in H–H network. Normalized log degree (x-axis) and log betweenness (y-axis) of H proteins into H–H network. Black solid line represents the linear regression fit (R^2 = 0.56). Horizontal and vertical dashed lines give the mean degree and betweenness values, respectively. Low-degree (LD) and high-degree (HD) classes were defined by using the average degree cutoff. (D) Mean degree and betweenness of H HCV and H HCV for LD and HD proteins. Top: mean betweenness (log scale) of H HCV (blue) and H HCV (red) is given for LD and HD classes. Bottom: mean degree of H HCV (blue) and H HCV (red) is given for LD and HD classes. The conventional standard error threshold and the U-test P-value are represented (***, P-value < 10^{-10}, NS: not significant).
Table II KEGG pathway enrichment for H_\text{HCV}

<table>
<thead>
<tr>
<th>KEGG name</th>
<th>CORE</th>
<th>E1</th>
<th>E2</th>
<th>NS3</th>
<th>NS5A</th>
<th>NS5B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell–cell and cell–ECM interactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherens junction</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell communication</td>
<td>6 (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell adhesion molecules (CAMs)</td>
<td></td>
<td></td>
<td></td>
<td>3 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECM–receptor interaction</td>
<td></td>
<td></td>
<td></td>
<td>2 (1)</td>
<td>6 (6)</td>
<td></td>
</tr>
<tr>
<td>Focal adhesion</td>
<td></td>
<td></td>
<td></td>
<td>10 (9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gap junction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8 (2)</td>
<td></td>
</tr>
<tr>
<td>Tight junction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Signalling pathways</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGFβ signalling pathway</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jak–STAT signalling pathway</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adipocytokine signalling pathway</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPK signalling pathway</td>
<td>3 (2)</td>
<td>2</td>
<td></td>
<td>7 (3)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Phosphatidylinositol signalling system</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wnt signalling pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Insulin signalling pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>B cell receptor receptor signalling pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>T cell receptor signalling pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Over-represented KEGG pathways were identified after multiple testing adjustments (adjusted P-value < 5 × 10^{-5}) and are listed by viral protein. For each pathway, number of H_\text{HCV} is given, with the relative contribution of IMAP Y2H dataset in brackets. Shaded entries denote discussed pathways.

Degree, betweenness and shortest path followed the same tendency with H_\text{EBV} proteins (Supplementary Table SV and Supplementary Figure S2) and were in good agreement with a previous report (Calderwood et al, 2007). These results indicate that preferential attachment on central proteins may be a general hallmark of viral proteins as recently suggested by analysis of the literature (Dyer et al, 2008). The high centrality of proteins was previously shown to correlate with their functional essentiality for the yeast model organism (Jeong et al, 2001; Ekman et al, 2006). In mammals, lethal and disease-related proteins were found enriched in central proteins (Wachi et al, 2005; Stark et al, 2006; Goh et al, 2007; Hernandez et al, 2007). This suggests that HCV proteins interacted with essential proteins in the cell.

To determine which of the degree or the betweenness most influences the probability of interaction between viral and cellular proteins, we used a generalized linear model to test the separate and additive effects of both measures (Supplementary methods). This analysis revealed that betweenness better explains the probability of interaction between viral and human proteins (ANOV A P-value < 10^{-3}). Figure 3C shows a partial correlation between k and b centrality measures (R^2=56%, P-value < 10^{-16}), explained by the high variability of betweenness at LD values. We thus asked whether this high variability observed at LD could explain the preponderant effect of betweenness. For this purpose, the data sets were split in LD and HD protein classes according to the average degree of the human interactome. For cellular proteins included in LD class, HCV interacts preferentially with proteins of high-betweenness independently of their degree property (Figure 3D). Within the HD class, interaction with HCV proteins is dependent on both betweenness and degree of cellular proteins. On the basis of a recent study in yeast (Joy et al, 2005), it can be extrapolated that LD high-betweenness H_\text{HCV} proteins could exert an effect as connectors or bottlenecks between cellular modules and may thus be essential for the infection.

Functional analysis of the HCV–human interaction network

To better understand biological functions targeted by HCV, we next tested the enrichment of specific pathways for all interactors of a given viral protein. This was done by analysing the H_\text{HCV} proteins with regard to the KEGG functional annotation pathways (Table II, Materials and methods). Although this approach is not totally unbiased because functions have not yet been attributed to all proteins, it remains a powerful way of incorporating conventional biology in system-level data sets. This analysis showed enrichment for three pathways associated with HCV clinical syndromes (insulin, TGFβ and Jak/STAT pathways) and identified focal adhesion as a novel pathway affected by HCV.

IJT network (insulin–Jak/STAT–TGFβ network)

Chronic infection by HCV is associated with an increased risk for metabolic disorders with the development of steatosis. Insulin resistance is a common feature of this process. It also contributes to liver fibrosis and is a predictor of a poor response to interferon-α (IFN-α) anti-viral therapy (D’Souza et al, 2005; Romero-Gomez et al, 2005). Conversely, IFN-α can prevent fibrosis progression (Poynard et al, 2002). TGFβ has a crucial function in maintaining cell growth and differentiation in the liver. It is a strong profibrogenic cytokine whose production is frequently enhanced during infection. Impaired TGFβ response is also observed during HCV infection (Schuppman et al, 2003). Although insulin, TGFβ and Jak/STAT pathways have been suspected to be involved in these clinical features (Romero-Gomez, 2006), their closely related perturbation during HCV infection remains largely unexplained. We thus used a network approach to identify cellular proteins targeted by HCV and localized at the interface of these pathways. The resulting interaction map was constructed to form the IJT network (insulin–Jak/STAT–TGFβ network,
Figure 4A; Supplementary methods). Sixty-six HHCV proteins are connecting two pathways, whereas 30 HHCV proteins are connecting the three pathways. Interaction of these proteins with HHCV proteins may thus induce functional perturbations that could expand to adjacent pathways. One of these proteins is PLSCR1 (Scramblase 1), connecting insulin and Jak/STAT pathways. Known to be involved in the redistribution of plasma membrane phospholipids (Sahu et al, 2007), this protein is also a potential activator of genes in response to interferon, and its knockdown with siRNA favours viral replication (Dong et al, 2004). Interestingly, PLSCR1/C0/C0 mice also exhibit an onset of insulin resistance (Wiedmer et al, 2004). Although not annotated in the insulin or Jak/STAT pathways, PLSCR1 thus appears essential for the functionality of these pathways. By interacting with PLSCR1, CORE could therefore interfere with both Jak/STAT and insulin pathways. Another example is the nuclear factor Yin Yang 1 (YY1), which exhibits a more central position in the IJT network as it connects the three pathways. HCV CORE interaction with YY1 has been previously shown to be functional relieving NPM1 expression. This observation could be extrapolated to PPARδ expression and SMADs transcriptional activity in support of insulin and TGFβ pathway modulation (Kurisaki et al, 2003; Mai et al, 2006; He et al, 2008). Interestingly, BCL6, targeted by NS5A, is another transcriptional repressor at the interface of the three pathways that inhibit Smad signalling (Wang et al, 2008). It also exerts an effect as a corepressor of PPARδ (Lee et al, 2003) and it regulates the expression of a subset of Jak/STAT pathway target genes (Arbouzova et al, 2006). Thus, perturbation of these pathways can reasonably be expected as a consequence of BCL6 or YY1 targeting by HCV. Also central in the IJT network, NOTCH1 has been reported to interfere functionally with the three pathways. Literature analysis revealed that many of the proteins at the interface are actually known to have an important function in the regulation of one, two or the three pathways without being annotated in the KEGG database. These are only illustrative examples of cellular targets most likely to be involved in HCV-induced phenotypes. Although this molecular approach of the pathology is applicable to basal element of a system (proteins in this work) some of the clinical phenotypes observed in chronic HCV infection are most likely
to result from the integrative effect of protein interactions depicted in the IJT network. In addition, the robustness property of a network can confer its ability to remain functional in face of different perturbations despite the deregulation of a single protein.

Another issue that became apparent in the IJT network is that CORE protein mediates proportionally more interactions than the other HCV proteins (Figure 4B and C). Indeed, preferential interaction with IJT network was observed only with CORE (51.3%, Supplementary Table SVI). As a consequence, CORE makes 27.7% of the interactions in the IJT network, corresponding to a significant enrichment (exact Fisher test \( P \text{-value} < 10^{-4} \)). More precisely, this CORE’s interactors are over-represented in Jak-STAT and TGFβ pathways (exact Fisher test \( P \text{-value} < 0.05 \)) and in HCV connecting insulin–Jak/STAT and insulin–TGFβ pathways (exact Fisher test \( P \text{-value} < 0.05 \), Supplementary Table SVI). CORE thus appears as a major perturbator of the IJT network. Interestingly, transgenic mice expressing CORE develop insulin resistance (Shintani et al., 2004; Pazienza et al., 2007). A proposed mechanism was that CORE-induced SOCS3 promotes proteasomal degradation of IRS1 and IRS2 through ubiquitination (Kawaguchi et al., 2004). As SOCS3 is also a negative regulator of Jak/STAT pathway, this could explain the occurrence of IFN-α resistance. Clearly, the IJT network indicates that the action of CORE is most likely to be much more complex that previously thought. Although the IJT network cannot yet be analysed dynamically, it remains that it provides a unique way of deciphering some of the complex disorders associated with chronicity. It is also worth considering that the IJT network may identify a series of genes involved in diseases, such as steatosis and fibrogenesis, in the absence of viral infection.

**Focal adhesion**

Focal adhesion was over-represented as a new function targeted by NS3 and NS5A proteins, with a major contribution
of data generated by IMAP Y2H screens (Table II). Integrin-linked focal adhesion complexes control cell adhesion to extracellular matrix (ECM) and association of these complexes with actin-cytoskeleton has an important function in cell migration. Upon binding to the ECM, both α and β integrin subunits recruit proteins establishing a physical link between the actin-cytoskeleton and signal transduction pathways. When deregulated, this functional process can lead to perturbation of cell mobility, detachment from the ECM and tumour initiation and progression. Figure 5A shows KEGG focal adhesion pathway with proteins targeted by HCV, mainly NS3 and NS5A proteins. Impact of single expression of NS3, NS3/4A or NS5A on focal adhesion functionality was assessed using a cellular adhesion assay on fibronectin and poly-L-lysine. These viral proteins significantly inhibited cell adhesion to fibronectin compared with NS2-expressing cells (an HCV protein with no interactor in the focal adhesion pathway) or mock-transfected cells, with 40% increase of FA50 (matrix concentration for half maximum adhesion) (Figure 5B left, Student’s t-test P-value <0.05). By contrast, adhesion to poly-L-lysine, which does not engage integrins, was not affected (Figure 5B right). The same inhibition level was observed for NS3/4A and NS3, suggesting that the enzyme activity of this protease does not have a major effect on focal adhesion perturbation. In addition to initiation and progression of cancer, the engagement of focal adhesion by HCV could have consequences on viral spreading. Interference with several steps of the actin-cytoskeleton remodelling has been described for retroviruses, which can exploit this process to surf along cellular protrusions of target cells to reach the entry site (Lehmann et al., 2005). It is conceivable that a related process, involving binding of the viral envelop to integrins, could be exploited by HCV to favour its transmission. This intriguing hypothesis will, however, be difficult to test until an efficient infection system of polarized cell is available.

Conclusion

In a network approach of HCV infection, the interaction map identifies all connections potentially needed for the virus to replicate and escape host defence. Whether all interactions really occur and have functional consequences is the open question of all interactome studies. The answer to this question necessitates the integration of system-level data sets of different origins that will set the stage for complex systems analysis of the infection. In a complex biological system, function cannot be predicted without understanding the component parts and their interactions and will result from the combination of theoretical knowledge of the cellular network with biological measurement of the interactions. Biological measurement, however, is still in the realm of low-throughput biology and needs major experimental improvement before prediction becomes the rule rather than the exception. Another fascinating challenge of this approach is to identify molecular signatures common to several viruses at the protein network level to develop original large-spectrum anti-viral molecules. A major step towards this goal is the high-throughput screening of a large variety of viruses, which is the aim of I-MAP.

Materials and methods

Construction of the HCV ORFeome

All HCV protein sequences were cloned in full length and domains except NS4B, for which no domain has been designed, using the euHCVdb facilities (http://euhcvdb.ibcp.fr; Combet et al., 2007) (Supplementary Figure S1). NS4A–NS3 fusion protein, as well as NS4A–NS3 protease domain were constructed (Kim et al., 1996; Taremi et al., 1998). All 27 ORFs from the HCV genotype 1b, isolate conl (AJ238799) (Lohmann et al., 1999), were cloned in a Gateway recombinational cloning system (Wallbott et al., 2000). Each ORF was PCR-amplified (with KOD polymerase, Novagen) using attB1.1 and attB2.1 recombination sites fused to forward and reverse primers, then cloned into pDONR223 (Rual et al., 2004). All entry clones were sequence-verified.

Yeast Two-hybrid (Y2H) screens

HCV ORFs were transferred from pDONR223 into bait vector (pPC97) to be expressed as Gal4–DB fusions in yeast. Two different screening methods were used (IMAP1 and IMAP2). For IMAP1, bait vectors were introduced in MAV203 yeast strain, and both human spleen and fetal brain AD-cDNA libraries (Invitrogen) were screened by transformation as described (Li et al., 2004). All primary positive clones (selected on SD–W–L–H + 3–AT) were tested by further phenotypic assay using two additional reporter genes: LacZ (β-Gal colorimetric assay) and URA3 (growth assay on S-FOA supplemented medium). Positive clones that displayed at least two out of three positive phenotypes were restested in fresh yeasts: bait vectors were transformed into MAV203 and each prey cDNA (obtained by colony PCR, see below) were transformed in combination with linearized prey vector (gap repair; Wallbott and Vidal, 2001). Clones that did not retest were discarded. AD-cDNA were PCR-amplified and inserts were sequenced to identify interactors. IMAP2 screens were performed by yeast mating, using AH109 and Y187 yeast strains (Chlontech; Albers et al., 2005). Bait vectors were transformed into AH109 (bait strain), and human spleen and fetal brain AD-cDNA libraries (Invitrogen) were transformed into Y187 (prey strain). Single bait strains were mated with prey strain, then diploids were plated on SD–W–L–H + 3–AT medium. Positive clones were maintained onto this selective medium for 15 days to eliminate any contaminant AD-cDNA plasmid (Vidalain et al., 2004). AD-cDNAs were PCR-amplified and inserts were sequenced.

Text-mining of interactions between HCV and human proteins

Language-curated interactions (LCI), describing binary interactions between cellular and HCV proteins, were extracted from BIND database and PubMed (publications before August 2007) by using an automatic text-mining pipeline completed by expert curation process. For the text-mining approach, all abstracts related to ‘HCV’ and ‘viral protein interactions’ keywords were retrieved, subjected to a sentencizer (sentence partition) and a part-of-speech tagger for gene name (based on NCBI gene name and aliases) and interaction verbs (Rebolitz-Schuhmann et al., 2008) (interact, bind, attach and so on). Sentences presenting co-occurrences of at least one viral gene name, one viral gene name and one interaction term were prioritized to curation by human expert.

Validation by co-affinity purification

Cellular ORFs (interacting domains found in Y2H screens) were cloned by recombinational cloning from a pool of human cDNA library or the MGC cDNA plasmids using KOD polymerase (Toyobo) into pDONR207 (Invitrogen). After validation by sequencing, these ORFs were transferred into pC-Neo-3×FLAG gateway-converted. HCV ORFs were transferred into pDEST27 (GST fusion in N-term). A total of 4×10⁶ HEK-293T cells were then co-transfected (6 μg JetPei, Polyplus) with 1.5 μg of each pair of plasmid. Controls are GST-alone against 3×FLAG-tagged prey. Two days after transfection, cells were
harvested and lysed (0.5% NP-40, 20 mM Tris–HCl (pH 8.0), 180 mM NaCl, 1 mM EDTA and Roche complete protease inhibitor cocktail). Cell lysates were cleared by centrifugation for 20 min at 13 000 r.p.m. at 4 °C and soluble protein complexes were purified by incubating 300 μg of cleared cell lysate with 40 μl glutathione sepharose 4B beads (GE Healthcare). Beads were then washed extensively with lysis buffer and proteins were separated on SDS–PAGE and transferred to nitrocellulose membrane. A total of 50 μg of cleared cell lysate was analysed by western blot to check the amount of 3 × FLAG-tagged cell protein. GST-tagged viral proteins and 3 × FLAG-tagged cellular proteins were detected using standard immunoblotting techniques using anti-GST (Covance) and anti-FLAG M2 (Sigma) monoclonal antibodies.

**Integrated human interactome network (H–H network)**

Only physical and direct binary protein–protein interactions were retrieved from BIND (Bader et al., 2003), BioGRID (Stark et al., 2006), DIP (Xenarios et al., 2002), GeneRIF (Lu et al., 2007), HPRD (Peri et al., 2004), IntAct (Kerrien et al., 2007), MINT (Chatr-aryamontri et al., 2007) and Reactome (Vastrik et al., 2007). NCBI official gene names were used to unify protein ACC, protein ID, gene name, symbol or alias defined in different genome reference databases (i.e ENSEMBL, UNIPROT, NCBI, INTACT, HPRD and so on) and to eliminate interaction redundancy due to the existence of different protein isoforms for a single gene. Thus, the gene name was used in the text to identify the proteins. Finally, only non-redundant protein–protein interactions were retained for building the human interactome data set.

**Topological analysis**

The R (http://www.r-project.org/) statistical environment was used to perform statistical analysis and the igraph R package (http://cneurocvs.rmi.kit.edu/igraph/) to compute network connected components, centrality (degree, betweenness) and shortest path measures.

The Wilcoxon–Mann–Whitney rank sum test (the U-test) was chosen to statistically challenge observed differences. The U-test is a non-parametric alternative to the paired Student’s t-test for the case of two related samples or repeated measurements on a single sample. The generalized linear model and ANOVA analysis was used to respectively model and test the separate and additive effects of degree and betweenness on the probability that HCV proteins interact with human proteins.

**Functional analysis using KEGG annotations**

Cellular pathway data were retrieved from KEGG (Aoki-Kinoshita and Kanehisa, 2007) and the Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/) and were used to annotate NCBI gene functions. For each viral–host protein interactors, the enrichment of specific KEGG pathway was tested by using an exact Fisher test (P-value < 5 × 10−5) followed by the Benjamini and Hochberg multiple test correction (Benjamini et al., 2001) to control false discovery rate.

**Cell-adhesion assay**

Serial dilutions (from 20 to 0.04 μg/ml) of fibronectin or poly-l-lysine in PBS were coated on 96-well microplates overnight at 4 °C. Non-specific binding sites were saturated at room temperature with PBS 1% BSA for 1 h. HEK 293T cells were transfected with PCI-neo-3 × FLAG NS2, NS3, NS4A or NNSA (JetPei, Polyclus), collected 2 days later with 2 mM EDTA in PBS, spread in triplicate at 1 × 10^5 cell per well in serum-free medium with 0.1% BSA and incubated for 30 min at 37 °C. Non-adherent cells were washed away and adherent cells were fixed with 3.7% paraformaldehyde. Cells were stained with 0.5% crystal violet in 20% methanol for 20 min at room temperature and washed five times in H2O. Staining was extracted from 50% ethanol in 50 mM sodium citrate, pH 4.5, and the absorbance was read at 590 nm on an ELISA reader (MRX microplate reader, Dynatech Laboratories). Values were normalized to 100% adhesion at 10 μg/ml. The percentage of adhesion was determined for each cell type at each matrix concentration. 50% of maximum adhesions (FASU) were calculated from the curves (Supplementary Figure S3) (adapted from Miao et al., 2000).

**Supplementary information**

Supplementary information is available at the Molecular Systems Biology website (www.nature.com/msb).

**Acknowledgements**

This work was funded by ANRS, INSERM and the French Ministry of Industry. We acknowledge L Meyniel for critical reading of the manuscript. We also acknowledge Lyon Biopôle. V Navratil is supported by a grant from INRA.

**Conflict of interest**

The authors declare that they have no conflict of interest.

**References**


© 2008 EMBO and Macmillan Publishers Limited


W de Chassey et al.

Molecular Systems Biology is an open-access journal published by European Molecular Biology Organization and Nature Publishing Group.

This article is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Licence.