Bioinformatics, 31(7), 2015, 1025–1033
doi: 10.1093/bioinformatics/btu763
Advance Access Publication Date: 21 November 2014
Original Paper

Systems biology

Molecular principles of human virus protein–protein interactions

Rachita Ramachandra Halehalli[1,2] and Hampapathalu Adimurthy Nagarajaram[1,*]

[1]Laboratory of Computational Biology, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, Telangana, 500001, India and [2]Graduate School, Manipal University, Manipal, 576104, Karnataka, India

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on June 25, 2014; revised on October 22, 2014; accepted on November 12, 2014

Abstract

Motivation: Viruses, from the human protein–protein interaction network perspective, target hubs, bottlenecks and interconnected nodes enriched in certain biological pathways. However, not much is known about the general characteristic features of the human proteins interacting with viral proteins (referred to as hVIPs) as well as the motifs and domains utilized by human-virus protein–-protein interactions (referred to as Hu-Vir PPIs).

Results: Our study has revealed that hVIPs are mostly disordered proteins, whereas viral proteins are mostly ordered proteins. Protein disorder in viral proteins and hVIPs varies from one subcellular location to another. In any given viral-human PPI pair, at least one of the two proteins is structurally disordered suggesting that disorder associated conformational flexibility as one of the characteristic features of virus–host interaction. Further analyses reveal that hVIPs are (i) slowly evolving proteins, (ii) associated with high centrality scores in human-PPI network, (iii) involved in multiple pathways, (iv) enriched in eukaryotic linear motifs (ELMs) associated with protein modification, degradation and regulatory processes, (v) associated with high number of splice variants and (vi) expressed abundantly across multiple tissues. These aforementioned findings suggest that conformational flexibility, spatial diversity, abundance and slow evolution are the characteristic features of the human proteins targeted by viral proteins. Hu-Vir PPIs are mostly mediated via domain–motif interactions (DMIs) where viral proteins employ motifs that mimic host ELMs to bind to domains in human proteins. DMIs are shared among viruses belonging to different families indicating a possible convergent evolution of these motifs to help viruses to adopt common strategies to subvert host cellular pathways.

Availability and implementation: Hu-Vir PPI data, DDI and DMI data for human–virus PPI can be downloaded from http://cdfd.org.in/labpages/computational_biology_datasets.html.

Contact: han@cdfd.org.in

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Viruses are the obligate intracellular parasites and hence need host cellular machinery for their genome replication and propagation enabled by interactions between viral and several host proteins during various stages of viral infection. Hence, viral-host protein–-protein interactions have been keenly studied by employing both computational (Dyer *et al.*, 2008; Meyniel-Schicklin *et al.*, 2012; Navratil *et al.*, 2011; Wuchty *et al.*, 2010) and experimental (Calderwood *et al.*, 2007; de Chassey *et al.*, 2008; Jäger *et al.*, 2012; Khadka *et al.*, 2011; Krishnan *et al.*, 2008; König *et al.*, 2008; Simonis *et al.*, 2012; Uetz *et al.*, 2006; Zhang *et al.*, 2009; Zhou *et al.*, 2008) approaches.

Human protein–protein interaction (Hu-PPI) network is robust to random attacks owing to its scale-free nature (Barabási, 1999); however, the network breaks down due to targeted attacks (Albert *et al.*, 2000). Studies have shown that viral proteins interact with hubs, bottlenecks and rich clubs in Hu-PPI network, and they also interact with the human proteins involved in signalling, cell cycle regulation, cell trafficking, transcription and translation (Calderwood *et al.*, 2007; de Chassey *et al.*, 2008; Dyer *et al.*, 2008; Navratil *et al.*, 2011; Uetz *et al.*, 2006; Wuchty *et al.*, 2010). The fact that viral proteins, although very a few in number, target hubs, bottlenecks and rich clubs indicates that viral subversion of human cellular processes is achieved by targeted attacks (Meyniel-Schicklin *et al.*, 2012). Our recent study has revealed yet another interesting facet of viral proteins that some of them act as articulation points in the human-viral-bridged PPI network by bridging unconnected components of Hu-PPI network, thereby mediating certain novel interactions to tune host machinery for their propagation (Halehalli and Nagarajaram, 2014).

In literature, one can find studies reporting properties of host proteins targeted by viral proteins. For example, it has been reported that the human proteins targeted by viruses are associated with diabetes and auto immune diseases (Navratil *et al.*, 2011). Viral immune modulators have been shown to target proteins involved in signalling pathways and cellular processes (Pichlmair *et al.*, 2012). It has been found that the targets of HIV-1 are highly conserved across primates (Jäger *et al.*, 2012).

Structural interaction network studies comprising about 50 viral-Hu-PPIs (Franzosa and Xia, 2011) have revealed that the endogenous interfaces (Hu-PPI) evolve at slower rates compared with rest of the surface, whereas the exogenous interfaces (viral targets) evolve at faster rates (Franzosa and Xia, 2011). Later studies based on another dataset comprising 954 interactions reported results (Franzosa *et al.*, 2012) contradicted these results.

Among proteins, some are referred to as intrinsically disordered proteins (IDPs) as these proteins are enriched with disordered regions lacking stable three-dimensional structures. IDPs are known to perform many biological functions such as signalling, regulatory processes, DNA binding and transcription (Babu *et al.*, 2011; Dyson and Wright, 2005; Tompa, 2002, 2009; Uversky and Dunker, 2010). IDPs usually interact with large number of proteins and hence occupy topologically important positions in protein–protein interaction networks (Dosztányi *et al.*, 2006; Kim *et al.*, 2008). They are tightly regulated (Gsponer *et al.*, 2008) in the cell and found to harbour short linear motifs [SLiMs/eukaryotic linear motifs (ELMs)] in their disordered regions that are recognized by modular protein domains such as SH3, PDZ and SH2 (Diella *et al.*, 2008; Nguyen Ba *et al.*, 2012).

Many genomes, from prokaryotes to eukaryotes, harbour several genes that code for IDPs (Ward *et al.*, 2004a) (about 2.0% in archaea, 4.2% in bacteria and 33.0% in eukaryotes). Viral genomes too code for IDPs, and the proportion of amino acids in the disordered regions of these genomes vary from 1% to 70% depending on their hosts or genome types (Pushker *et al.*, 2013; Xue *et al.*, 2012). Furthermore, some of the viruses harbour sequences that mimic ELMs in the disordered regions (Davey *et al.*, 2011; Diella *et al.*, 2008).

Some of the IDPs are associated with certain human genetic disorders, cancers, neurodegenerative and autoimmune diseases (Babu *et al.*, 2011; Midic *et al.*, 2009; Uversky *et al.*, 2008). However, role of IDPs in infectious diseases, in particular those caused by viral infections, has not been well studied.

In this study, we report a systematic analysis of characteristic features of human proteins interacting with viral proteins (hVIPs) as well as on general features of human-virus (Hu-Vir) PPI. We find that significant number of hVIPs can be classified as IDPs. Furthermore, hVIPs occupy central positions in PPI network; evolve at relatively slower rates; involve in multiple pathways, harbour a number of potential binding interfaces and linear motifs in their disordered regions. We also find that virus-Hu-PPIs are mediated by a common set of domain–motif interactions (DMIs).

2 Materials and methods

2.1 Interaction data

2.1.1 Hu-Vir PPI data

The dataset used in this study is same as the one used in our previous study (Halehalli and Nagarajaram, 2014). For the sake of continuity, we give some essential details of this dataset. The dataset comprises curated human and virus protein interaction data, which were obtained from the publically available, literature-curated databases VirusMINT (Chatr-aryamontri *et al.*, 2009) and PIG (Driscoll *et al.*, 2009). In addition, interactions predicted from yeast two-hybrid studies of human proteins with Vaccinia virus (Zhang *et al.*, 2009), Dengue virus (Khadka *et al.*, 2011), HTLV-1 and HTLV-2 (Simonis *et al.*, 2012) were also added (further details are given in Supplementary Material).

2.1.2 Hu-PPI data

Union of PPIs in HPRD (Keshava Prasad *et al.*, 2009) and IntAct (Aranda *et al.*, 2010) was used as a representative dataset of binary Hu-PPIs for studies reported here. The data consist of 49 772 interactions involving 10 446 proteins. Of the proteins in this dataset, 8887 are not represented in the Hu-Vir protein–protein interaction (Hu-Vir PPI) dataset and therefore, they are referred to as 'non-hVIPs' for the purpose of comparative analysis with hVIPs.

2.2 Disorder and subcellular localization prediction

2.2.1 Prediction of intrinsic disorder in proteins

Structural disorder in human proteins was predicted using DISOPRED2 (Ward *et al.*, 2004b) with a false-positive rate of up to 2.0, and the protein sequences used for predictions correspond to the reviewed canonical forms represented in UniProt/SwissProt (version 116) (Apweiler *et al.*, 2004). In literature, a protein is said to be disordered depending on the length of disordered region (Ward *et al.*, 2004a) or percentage of disordered residues (Gsponer *et al.*, 2008) in that protein. In this study, a protein is considered as disordered (IDP) if it has >30% of its residues in the predicted disordered regions; ordered protein (ODP) if it has ≤10% of its residues in disordered regions and moderately disordered protein (MDP) if it has 10–30% of its residues in disordered regions. While performing χ^2 tests, we have used the expected numbers of IDPs, MDPs and ODPs calculated from their % in UniProt/SwissProt.

2.2.2 Subcellular localization of human proteins and localization entropy

Subcellular localization was predicted using localization prediction tool WoLF PSORT (Horton *et al.*, 2007). Localization entropy (LE, Snel *et al.*, 2002) of a hVIP was calculated using the following formula:

$$\mathrm{LE} = -\sum \frac{T_k}{\sum_k^n T_k} \log\left(\frac{T_k}{\sum_k^n T_k}\right)$$

Where T_k is the number of times that the subcellular localization 'k' associated with all known human protein partners of a human

protein in Hu-PPI network and '*n*' is the number of distinct subcellular localizations associated with all interacting human proteins. Similarly, LE was calculated for all viral proteins based on subcellular localization of their human interacting partners in Hu-Vir PPI network.

2.3 Pathway centrality, gene expression, protein abundance, splice variants and evolutionary rate of proteins

Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa, 2000) pathway annotation data were used to find the pathways involving hVIPs and other human proteins of interest. We defined the total number of pathways a human protein is involved as its pathway centrality. Gene expression data were downloaded from BioGPS (Wu *et al.*, 2009), and median of the expression values was used as the cutoff for designating a gene as 'expressed' or 'not expressed' in a given tissue. Protein abundance information was taken from PaxDb (Wang *et al.*, 2012). Splice variant/transcript count, d*N* and d*S* values for proteins were retrieved from BioMart (Kinsella *et al.*, 2011) (Ensemble). Overall evolutionary rate (d*N*/d*S*) was quantified for hVIPs and 'non-hVIPs' of Hu-PPI network as the divergence of human protein sequences from mouse orthologous protein sequences.

2.4 Putative binding interfaces and ELM motifs in proteins

We used ANCHOR (Dosztányi *et al.*, 2009) to predict the putative binding interfaces in the disordered regions of proteins. ANCHOR also identifies known ELMs [which are listed in ELM database (Dinkel *et al.*, 2012)] present in a given protein. ELM enrichment *P* values for hVIPs were calculated using Fisher test in R, and resultant *P* values were adjusted using false discovery rate (FDR) correction method in R.

2.5 Mapping DMIs and domain–domain interactions

Domains were identified in human protein and viral protein sequences using the domain search tool InterProScan (Quevillon *et al.*, 2005). ELMs in both human and viral proteins were identified using ANCHOR. A domain interacting with an ELM is referred to as a linear motif binding domain (LMBD). Known DMIs were taken from ELM database (Dinkel *et al.*, 2012) and iELM web server (Weatheritt *et al.*, 2012). ELM-LMBD pairs in ELM database comprise 200 experimentally determined, literature-curated associations. iELM web server predicts ELM-LMBD associations in Hu-PPI pairs and hence we predicted interactions for all Hu-PPIs in the dataset using iELM and obtained 116 high confidence domain–motif pairs. Union of aforementioned sources comprising 240 iELMs were searched in Hu-Vir PPI pairs. Domain–domain interaction (DDI) data were downloaded from iPfam (Finn *et al.*, 2005), which gives known DDIs from PDB structures and from DOMINE (Yellaboina *et al.*, 2011), which is a database of predicted DDIs. Using these DDI data, Hu-Vir PPI pairs were screened for the presence of interacting domain pairs.

3 Results

Earlier studies had shown that viruses target central nodes in the Hu-PPI network (Calderwood *et al.*, 2007; de Chassey *et al.*, 2008; Dyer *et al.*, 2008; Navratil *et al.*, 2011; Uetz *et al.*, 2006). We reconfirmed these results by analyzing our dataset, which is much larger and also non-redundant (Supplementary Materials and Methods) when compared with the ones used in the earlier studies. We calculated different centrality measures (Supplementary Materials and Methods) such as degree, betweenness, transitivity/clustering coefficient, closeness, PageRank, Eigen vector and hub score for hVIPs in Hu-PPI network and compared them with non-hVIPs. It was found that hVIPs are associated with higher centrality measures ($P \ll 10^{-16}$) when compared with non-hVIPs (Supplementary Table S1), thereby reconfirming earlier observations that viruses attack centrally important nodes in Hu-PPI network. Attack on centrally important nodes is akin to targeted attacks and such attacks are known to breakdown networks (Meyniel-Schicklin *et al.*, 2012). This reconfirms earlier findings (Calderwood *et al.*, 2007; de Chassey *et al.*, 2008; Dyer *et al.*, 2008; Navratil *et al.*, 2011; Uetz *et al.*, 2006) that viruses target important proteins that allow them to gain control over information flow in Hu-PPI network, which in turn help for their replication and reproduction.

3.1 Viruses target human proteins that are rich in disordered regions

In our dataset, nearly 70% (1196 of 1735) of the hVIPs are either IDPs or MDPs (Fig. 1a), and this corresponds to a significant enrichment for disordered proteins in viral targets considering the distribution of IDPs and MDPs observed in all the known human proteins in Uniprot/Swissprot ($P \ll 10^{-16}$) (Fig.1a). Involvement of protein disorder in Hu-Vir PPIs is also reflected in the number of interacting pairs involving structurally disordered proteins. Of the 3389 pairs, about 60% of them involve either disordered human or disordered viral protein or both (Fig. 1b). Interacting pairs where both human and viral proteins are ODPs constitute mere 12% of the pairs in our dataset. Such interactions are mostly seen in extra cellular space, cell membrane and cytoplasm. The fact that a majority of Hu-Vir PPIs is mediated by protein disorder asserts one to think beyond the so called 'structural interaction networks (SINs)' as proposed by (Franzosa and Xia, 2011) for viral-Hu-PPIs.

We also found that the extent of disorder in hVIPs is negatively correlated ($r = -0.76$, $P = 0.005$) to the extent of disorder in their interacting viral proteins (Fig. 2a). This indicates that the hVIPs and their viral interacting partners complement each other in their structural disorder/order. Further, we investigated structural disorder of hVIPs and their viral proteins with respect to subcellular localization of hVIPs (Fig. 2b). The structural disorder associated with hVIPs

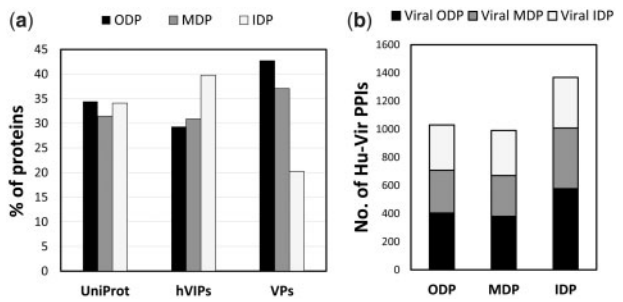


Fig. 1. Hu-Vir PPIs involve disordered proteins. (**a**) Distribution of different structural categories of proteins in human and viral genomes. Structurally ordered proteins (ODPs) (0–10% of amino acids in disordered region), MDPs (10–30% of amino acids in disordered region) and IDPs (>30% residues in disordered regions). These distributions indicate that the hVIPs are mostly disordered ($\chi^2 = 75.06$, $P \ll 10^{-16}$) and the viral proteins that interact with human proteins are mostly ordered ($\chi^2 = 41.53$, $P < 10^{-10}$). (**b**) Proportion of ODP, MDP and IDP viral proteins interacting with human ODP, MDP and IDPs. Fifty-nine percent of the Hu-Vir PPIs involve either disordered human protein or disordered viral protein or both

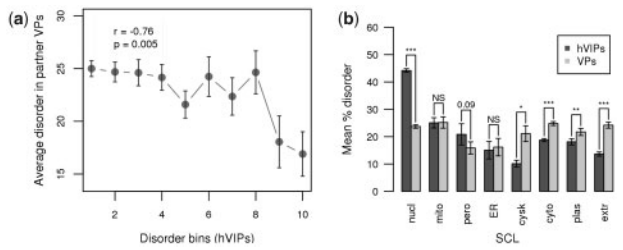


Fig. 2. Structural order–disorder complementarity in Hu-Vir PPIs. (**a**) Average disorder in VPs versus disorder in hVIPs. Each bin of hVIPs corresponds to a widow size of 10% disorder. As disorder in hVIPs increases, average disorder of their viral protein partners in Hu-Vir PPI network decreases (one-sided Pearson correlation test gives Rho value and *P* value). (**b**) Mean % disorder in hVIPs and VPs in different subcellular locations. extr, extracellular space; plas, plasma membrane; cyto, cytoplasm; cysk, cytoskeleton; ER, endoplasmic reticulum; pero, peroxisome, mito, mitochondria; nucl, nucleus. $*P < 10^{-3}$, $**P < 10^{-6}$, $***P < 10^{-15}$, NS, no significance. Error bars represent ± standard error in distribution

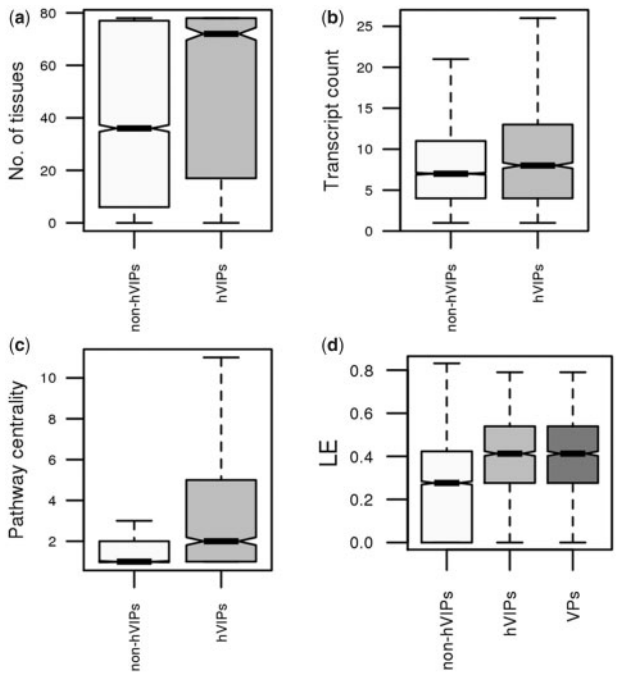


Fig. 3. hVIPs are associated with functional diversity. (**a**) Expression breadths of hVIPs and non-hVIPs. hVIPs are expressed in many tissues when compared with non-hVIPs ($P \ll 10^{-16}$). (**b**) Transcript abundance of hVIPs and non-VIPs. hVIPs are associated with higher number of splice variants when compared with non-hVIPs ($P < 10^{-12}$). (**c**) Pathway centrality of hVIPs and non-hVIPs. hVIPs are involved in higher number of pathways (KEGG) than non-hVIPs ($P \ll 10^{-16}$). (**d**) Distribution of LE values of hVIPs, non-hVIPs and viral proteins (VPs); hVIPs show higher LE than non-hVIPs and viral proteins ($P \ll 10^{-16}$)

tends to increase as we move from extracellular space to nucleus, whereas an opposite trend is observed for viral proteins where their structural disorder tends to decrease from extracellular space to nucleus (Fig. 2b). It was found that hVIPs found in extracellular space, cytoskeleton and endoplasmic reticulum are mostly ODPs, whereas those found in cytoplasm, mitochondria, peroxisome and nucleus are mostly IDPs (Supplementary Fig. S1a and b). On the other hand, the viral proteins that interact with hVIPs found in extracellular space, cell membrane and cytoskeleton are disordered, and they are mostly ordered in nucleus mitochondria and peroxisome. hVIPs localized to nucleus and cytoplasm constitute majority of Hu-Vir PPIs (Supplementary Fig. S2). This suggests that intrinsic conformational flexibility associated with protein structural disorder is one of the key characteristic features of Hu-Vir PPIs.

3.2 Viruses target universally expressed and functionally diverse proteins

To understand the spatial expression pattern of the hVIPs, we calculated their expression breadths (the number of tissues in which hVIPs are expressed) and found that hVIPs are expressed in many tissues ($P < 10^{-16}$) (Fig. 3a). We also found that hVIPs are enriched with many splice variants ($P < 10^{-12}$) (Fig. 3b). Furthermore, we counted the number of KEGG pathways involving each of the hVIPs (referred to as Pathway Centrality) and found that hVIPs when compared with non-hVIPs are associated with higher pathway centralities ($P \ll 10^{-16}$) (Fig. 3c).

To know whether the interacting partners of hVIPs and non-hVIPs in Hu-PPI network and those of viral proteins in Hu-Vir PPI network are localized to a few cellular locations or found in many diverse localizations, we calculated their LE values. It was found that hVIPs are associated with higher LEs than non-hVIPs ($P \ll 10^{-16}$) indicating that the human partners of the hVIPs are located in diverse locations than the partners of non-hVIPs. Viral proteins show smaller LE compared with hVIPs ($P \ll 10^{-16}$) indicating that the partners of viral proteins or VPs (i.e. hVIPs) are located in fewer localizations which is similar to that of non-hVIPs (Fig. 3d). However, it is interesting to note that few viral proteins such as HIV-1 and HIV-2 nef (LE = 0.73 each) and EBV BFLF2 (LE = 0.70) show very high LE. LE of hVIPs positively correlates with their degree in Hu- PPI network (Spearman rank correlation, rho=0.66, $P < 10^{-196}$), and LE of VPs strongly correlates with the number of hVIPs they interact with (rho = 0.80, $P < 10^{-325}$). Viruses target house-keeping proteins that are associated with high splice variation and involved in regulation of multiple pathways and hence are able to manipulate diverse cellular processes.

3.3 Viruses target abundantly expressed and slow evolving proteins

We investigated hVIPs and non-hVIPs for protein abundance, presence of PEST motifs (sites for ubiquitination) and their overall evolutionary rates. We found that hVIPs are abundantly expressed when compared with non-hVIPs ($P < 10^{-16}$) (Fig. 4a). hVIPs do not differ from non-hVIPs in the number of PEST motifs they harbour ($P = 0.06$) (Supplementary Fig. S3), suggesting that both the groups of proteins have similar propensities for their regulation. hVIPs show smaller d*N*/d*S* ratios when compared with non-hVIPs suggesting that they are evolving at much slower rates than non-hVIPs ($P = 10^{-13}$) (Fig. 4b). It has been shown that immune system related proteins usually show higher d*N*/d*S* values (Fumagalli *et al.*, 2011; Nielsen *et al.*, 2005). Our investigations too reveal that immune-system-related proteins show distinctly higher evolutionary rates when compared with other proteins among hVIPs ($P < 10^{-8}$) and non-hVIPs ($P < 10^{-14}$). However, between the immune related proteins of hVIPs and that of non-hVIPs, the former group shows lower d*N*/d*S* values when compared with the latter group ($P < 10^{-3}$) (Supplementary Fig. S4). These results collectively suggest that viral targets are abundantly expressed and are relatively conserved when compared with non-hVIPs.

3.4 Linear motifs enriched in hVIPs

We calculated d*N*/d*S* (divergence from orthologous mouse sequence) values separately for disordered and ordered regions in hVIPs and

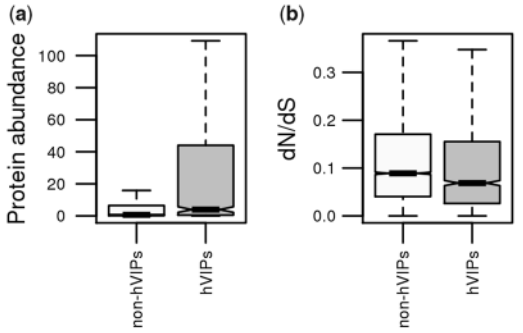


Fig. 4. hVIPs are abundantly expressed and slow evolving. (**a**) Averaged protein abundance of hVIPs and non-hVIPs in the human tissues; hVIPs are more abundantly expressed than non-hVIPs ($P \ll 10^{-16}$). (**b**) d*N*/d*S* (divergence from mouse protein sequence) values for hVIPs and non-hVIPs; hVIPs evolve at slower rates than non-hVIPs ($P < 10^{-13}$)

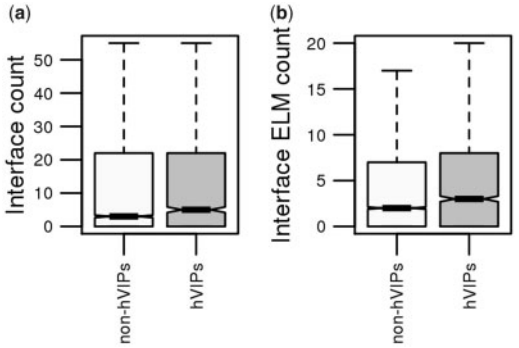


Fig. 5. Binding interface and ELM distribution in disordered regions of hVIPs and non-hVIPs. (**a**) Number of binding interfaces in disordered regions of hVIPs and non-hVIPs predicted by ANCHOR; hVIPs have higher number of binding interfaces than non-hVIPs ($P < 10^{-4}$). (**b**) Number of ELMs in binding interfaces of hVIPs and non-hVIPs; hVIPs have higher number of interface ELMs than non-hVIPs ($P < 10^{-3}$)

non-hVIPs using the program PAML4.8 (Yang, 2007). As reported in the literature (Brown *et al.*, 2002, 2010), disordered residues have higher d*N*/d*S* values when compared with the d*N*/d*S* values of ordered regions in hVIPs and non-hVIPs ($P < 10^{-48}$). When we compared d*N*/d*S* values of disordered regions in hV IPs with those in non-hVIPs, we found that the disordered regions in hVIPs have slightly lower d*N*/d*S* values ($P < 10^{-3}$) (Supplementary Fig. S5).

Disorder in proteins can be classified into three types: (i) constrained disorder (highly conserved), (ii) flexible disorder and (iii) non-conserved disorder (Bellay *et al.*, 2011). Constrained disordered regions usually harbour motifs referred to as SLiMs/ELMs (Brown *et al.*, 2010; Diella *et al.*, 2008; Nguyen Ba *et al.*, 2012), which confer them differential regulation and varied functions (Babu *et al.*, 2011; Diella *et al.*, 2008; Dinkel *et al.*, 2012). With the help of ANCHOR (Dosztányi *et al.*, 2009), we predicted binding interfaces and ELMs in proteins. We found that hVIPs harbour higher number of binding interfaces and ELMs when compared with non-hVIPs ($P \ll 10^{-05}$) (Fig. 5). These binding surfaces and ELMs enable viral proteins to interact with many partners. This explains why hVIPs show higher centrality values in the interaction network when compared with non-hVIPs.

Further we calculated enrichment score for ELMs in hVIPs. Of the 118 ELMs found in 1735 hVIPs, 62 were enriched in hVIPs ($P \ll 0.05$; Fisher test FDR corrected) (Supplementary Table S2). The enriched motifs are involved in regulation of various signalling pathways through kinase actions (such as GSK3_1, CK1, MAPK and LIG_PDZ/SH3/SH2) and those corresponding to post-translational modification sites [the motifs GlcNHglycan, N_GLC_1 and SUMO (Small ubiquitin like molecule)]. It is to be noted that ELMs act as molecular switches (Van Roey *et al.*, 2012) in regulation of many processes and used for transient binding with variety of proteins.

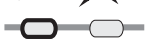
3.5 DMIs and DDIs involved in Hu-Vir PPIs

ELMs interact with their LMBDs leading to distinct ELM-LMDB pairs. We prepared a list of potential ELM-LMDB pairs as a union set of (i) 116 high confidence ELM-LMBD pairs predicted from human protein interaction pairs using iELM webserver (Weatheritt *et al.*, 2012) and (ii) the 200 curated ELM-LMBD pairs present in ELM database (Dinkel *et al.*, 2012). Union set ELM-LMBD pairs or DMIs consists of 240 non-redundant pairs. The Hu-Vir interacting proteins were scanned for the presence of ELM-LMBD pairs, which resulted in 985 LMBDs (hVIPs) and ELMs (VPs) interactions in 295 Hu-Vir pairs (Table 1) between 54 unique DMIs are found. Of the 54 unique pairs of DMIs, 20 are experimentally known (Davey *et al.*, 2011) (Supplementary Table S3). We also found 47 ELM (hVIP)-LMBD (VPs) interactions in 12 Hu-Vir PPIs (Table 1 and Supplementary Table S4) between 11 unique pairs of motif–domain interactions (MDIs). MDIs are mostly formed between SH2 and SH3 domains in Sarcoma viral proteins or Cyclin_1 in Herpes viral proteins and corresponding interacting motifs in human proteins.

In addition to the DMIs, DDIs can also mediate viral Hu-PPIs. We, therefore, identified interacting domain pairs (or DDIs) in the interacting human and viral proteins. The interacting domain pairs were taken from iPfam (Finn *et al.*, 2005) and DOMINE (Yellaboina *et al.*, 2011) as described in Section 2. We could identify 225 DDI interactions in 172 unique Hu-Vir PPIs between 90 pairs of DDIs (Table 1 and Supplementary Table S5). These DDIs are formed between 51 unique Pfam domains in 126 unique human proteins and 49 unique Pfam domains in 53 unique viral proteins. About 70% of the DDIs (156 of 225) resolved in our study are from ordered viral proteins interacting with both ordered and disordered human proteins. As can be seen, viral Hu-PPIs are predominantly mediated by DMIs than DDIs, and this result is consistent with the recently published work (Garamszegi *et al.*, 2013). Our investigations also revealed subcellular localization-specific utilization of DDIs and DMIs. hVIPs that utilize DDIs for viral interactions are mostly found in extracellular space, plasma membrane, cytoskeleton and peroxisome, whereas those utilizing DMIs are found in cytoplasmic, mitochondrial and nuclear localizations (Supplementary Fig. S6a–h).

A bipartite network (Fig. 6) of DDIs and DMIs in Hu-Vir PPI revealed that ELMs interact with their unique domain counterparts, whereas domains interact with multiple ELMs and domains ($P = 0.01$) (Fig. 7a). There are more DMIs per Hu-Vir PPI than DDIs ($P < 10^{-16}$) (Fig. 7b), and multiple DMIs are used by individual Hu-Vir PPIs when compared with DDIs between protein pairs ($P = 10^{-09}$) (Fig. 7c). It is interesting to note that some of the linear motifs often used by viruses to interact with host proteins are utilized in signalling pathways, which include kinase functions (Pkinase, SH2, SH3_1 domains), cell cycle regulation (TRAF6, TRAF2, APPCC_Dbox1, WW domains) or protein degradation (MATH domain) or cleavage (Peptidase_S8) pathways (Diella *et al.*, 2008). Viruses, therefore, by mimicking human motifs interfere in the regulation of the concerned pathways. A DMI, when compared with DDI, is utilized by multiple viruses ($P = 10^{-09}$) (Fig. 7d) suggesting that such DMIs form common mode of molecular interactions by multiple viruses, whereas DDIs are virus specific. Further

Table 1. Summary of DMIs and DDIs in Hu-Vir PPIs

Domain/motif profile in hVIP	Domain/motif profile in VP	Data for interaction search	No. of interactions in Hu-Vir PPI	Total no. of Hu-Vir PPI (%)	Unique no. of DMIs/DDIs
		iELM/ELMdb	985 	295 (~9%)	54[a]
		iELM/ELMdb	47 	12 (0.35%)	11
		DOMINE/iPfam	225 	172 (5%)	90

ELM in human protein:  ; ELM in viral protein: ; domain in hVIP:  and domain in viral protein: . Domain profile was obtained using InterProScan. ELM profile for protein obtained using ANCHOR.

[a]Twenty have experimental evidences (Davey *et al.*, 2011) (detailed list in Supplementary Table S3).

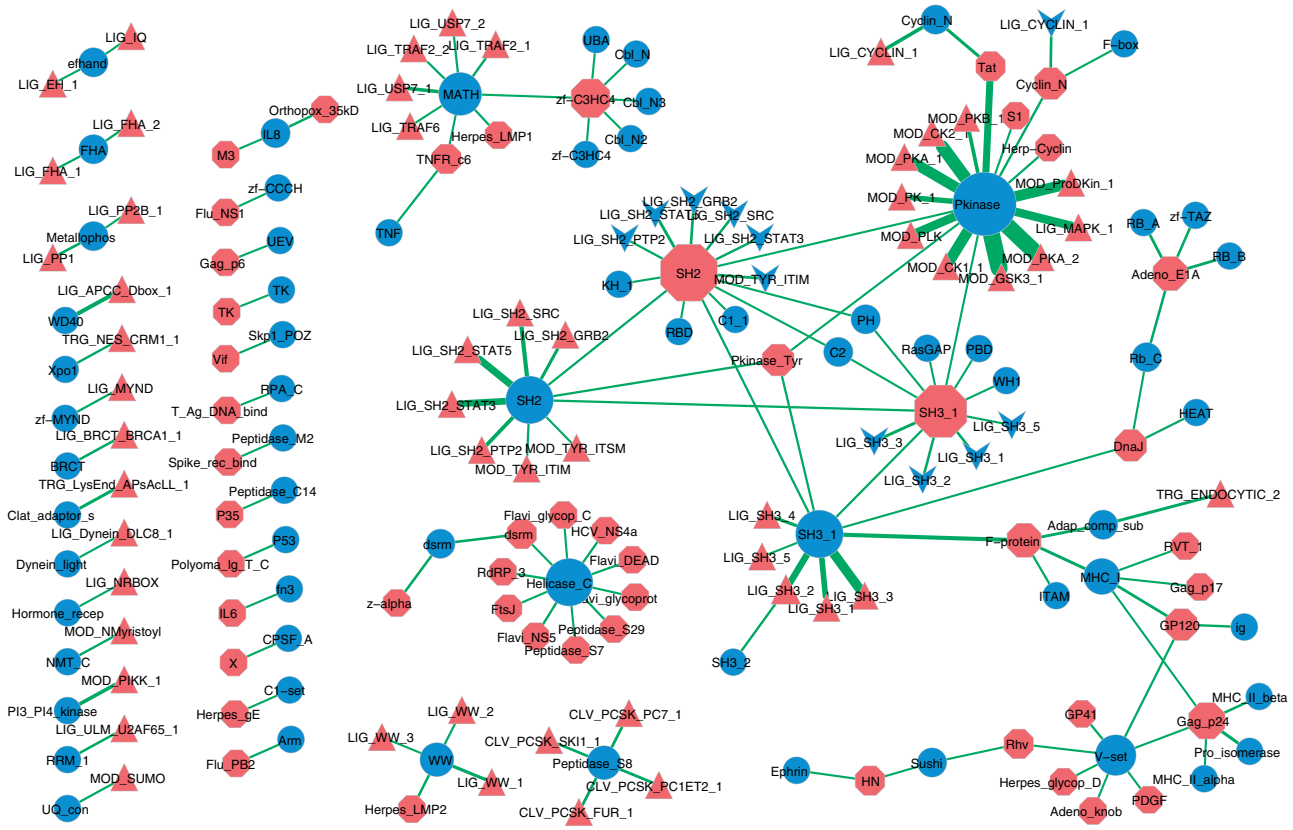


Fig. 6. Domain and motif utilization in Hu-Vi PPIs represented by means of human–virus domain–motif and domain–domain bipartite interaction network. Domains corresponding to human protein are represented by blue circles and motifs by blue vee. Domains corresponding to viral proteins are represented by coral octagon and motifs by coral triangles. Edge thickness corresponds to number of Hu-Vir PPI using particular DMI or DDI and node size corresponds to degree in the current network. Network comprises 54 domain (in hVIP)–motif (in VP) associations from 302 Hu-Vir PPI, 11 motif (hVIP)–domain (VP) associations from 47 Hu-Vir PPI and 90 DDI from 172 Hu-Vir PPIs. Network was visualized using Cytoscape (Shannon *et al.*, 2003)

we calculated KEGG pathway enrichment of human proteins involved in DDIs and DMIs among different families of viruses, and comparison of the enriched pathways ($P < 0.05$) reveals that common set of pathways are targeted via DMIs by different families of viruses (Supplementary Fig. S8a and b). Pathways uniquely enriched in hVIPs that utilize DDIs are related to immune recognition such as antigen processing, IgA production, NOD-like receptor signalling and mRNA surveillance (e.g. Fig. 6 for MHC I, MHC II and Ig domains utilizing DDI). In the hVIPs that utilize DDI and DMI with their viral partners many immune signalling pathways and cell cycle regulation pathways are enriched. It is interesting to note that few signalling pathways (mTOR, insulin, adipocytokine signalling pathways) related to regulation of metabolism are also enriched in hVIPs utilizing DMI with viral proteins (Supplementary Fig. S8a and b). In our previous work, we have shown certain viral proteins acting as articulation points that bridge human proteins in metabolic pathways to the giant component in Hu-PPI network (Halehalli and Nagarajaram, 2014).

4 Discussion

Viral replication inside host is marked by various stages of life cycle. During various stages of viral life cycle, viral proteins interact with human proteins localized to distinct cellular locations and distinct functions: extra cellular space or plasma membrane for virus entry and egress, cytoplasm during unpackaging or virion assembly, endoplasmic reticulum for protein synthesis and nucleus for viral transcription/replication/mRNA processing. Our study has revealed that Hu-Vir interactions mainly involve disordered proteins (60% Hu-

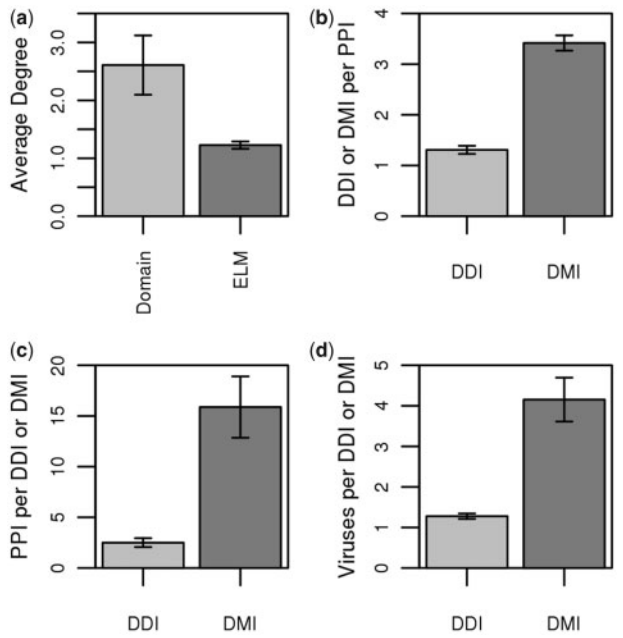


Fig. 7. Average properties of DDIs and DMIs. (**a**) Average degree of domains and motifs in bipartite DDI-DMI network. Domains are involved in higher number of interactions than linear motifs ($P = 0.01$). (**b**) Average number of DDI or DMI per Hu-Vir PPI. There is higher number of DMIs per known Hu-Vir PPI than DDIs ($P \ll 10^{-16}$). (**c**) Average number of Hu-Vir PPIs using a given DDI/DMI. DMIs are utilized by higher number of Hu-Vir PPIs than DDIs ($P < 10^{-08}$). (**d**) Average number of viruses using a particular DDI/DMI. Multiple viruses use same DMI, whereas DDIs are virus specific ($P < 10^{-08}$). Error bars represent ± standard error in distribution

Vir PPI) (Fig. 1a and b), which provide inherent conformational flexibility to interacting partners perhaps as a means to compensate structural variations caused by mutations that frequently happen in viral proteins. Protein structural disorder in hVIPs and viral proteins shows subcellular location-specific variation (Fig. 2a–h and supplementary fig. S1a and b). In most of the Hu-Vir PPIs, human and viral proteins complement each other in terms of their structural disorder—when one is disordered, the other is usually ordered. Disorder-associated conformational flexibility thus seems to be the common feature of Hu-Vi PPIs.

Viruses target human proteins that occupy central positions and hence make targeted attacks on Hu-PPI by which they gain control over information flow in Hu-PPI network. hVIPs are expressed across multiple tissues suggesting that they are housekeeping genes which are required for cellular processes. hVIPs are associated with splice variation, involved in multiple pathways and localized to multiple subcellular locations thus presenting diversity and variation to interacting viral proteins. Viruses are known to jump across species to cause diseases (Morens *et al.*, 2004; Woolhouse *et al.*, 2005) as a consequence of their co-evolution with host proteins. Immune system proteins evolve at higher rates when compared with non-immune proteins (Fumagalli *et al.*, 2011; Nielsen *et al.*, 2005). Our study also has revealed that immune hVIPs evolve at faster rate, whereas non-immune hVIPs show slowest evolutionary rate. These observations suggest two facets of viral infections: (i) evolutionary arms race between viral and antiviral proteins consistent with (Daugherty and Malik, 2012) and (ii) viruses' capability to infect different hosts (Sawyer and Elde, 2012). Viral pathogen of one host species cause a severe pathological symptoms in an another closely related host species. For example, HIV-1 (Huet *et al.*, 1990) and HIV-2 (Hirsch *et al.*, 1989) are thought to have jumped from Chimpanzee and Sooty mangabey, respectively. Influenza H1N1 causative of swine flu has jumped from pigs to humans and KFDV (Kyasanoor forest disease virus) that causes the disease in human has jumped from monkey.

hVIPs show higher number of interfaces in disordered regions compared with non-hVIPs, and they can form multi-interface hubs (Kim *et al.*, 2008). Large number of interfaces in a protein enables it to interact with multiple proteins. hVIPs are also enriched with ELMs, which function in regulation of activities of proteins in cellular processes such as endocytosis, signalling, transcription, cell cycle regulation and protein degradation pathways.

When we resolved Hu-Vir PPI using known DMI and DDI information, we found that viruses mainly utilize ELMs with hVIPs. In a recent study, Garamszegi *et al.* (2013) reported that interspecies PPIs between human and viral proteins are mostly characterized by DMIs compared with intraspecies interactions They also reported that 30% Hu-Vir PPIs can be resolved using DDI and DMIs. In our study, with high coverage of almost all known Hu-Vir PPIs and higher number of DDIs and DMIs than previously reported in literature (Garamszegi *et al.*, 2013), we found only 11.6% of the interactions utilize either DDI or DMI.

ELMs in viral proteins mimic host motifs. Mimicking SLiMs is evolutionarily favourable for viruses as they undergo mutations rapidly driving convergent evolution (Davey *et al.*, 2011). Mimicked ELMs in viral proteins mostly target kinase functions, cell cycle regulation, protein degradation and cleavage pathways. In bipartite DDI and DMI network, we found that DDIs are widespread, and viral domains are involved in multiple interactions with different domains, whereas motifs are involved in single interaction with corresponding LMBD and are utilized by multiple Hu-Vir PPIs. It was also found that DDIs are mostly virus specific, whereas DMIs are used by multiple viruses.

KEGG pathways enriched in hVIPs involved in DMI with viral proteins of different viral families indicate that they are common among families and those of DDIs are more specific to families of viruses. This further confirms convergent evolution of linear motifs in viruses belonging to different families as a strategy to target common pathways. KEGG pathways uniquely enriched in DDI and DMI are distinct. In DDI, they are related to immune recognition of viruses, whereas in DMI, they are related to metabolism, cell cycle regulation and immune signalling. It is well established that certain viral proteins modulate pattern recognition receptor (PRR) signalling pathways (Bowie and Unterholzner, 2008), and our results show downstream signalling proteins of PRR pathways form DMI with viral proteins indicating viruses use linear motifs for manipulating these pathways. Proteins containing WW domain, which have role in cell cycle regulation (Salah and Aqeilan, 2011), are targeted by viruses by mimicking WW domain binding motifs consistent with previous studies (Freed, 2002; Winberg *et al.*, 2000). Non-homologous end joining pathway is enriched in hVIPs involved in DMIs of *Retroviridae* and *Herpesviridae* whose activation is required for retroviruses (Li *et al.*, 2001) and inhibition during lytic cycle of herpes viruses (Rennekamp and Lieberman, 2010). Interestingly, we found a few pathways involved in regulation of metabolism such as carbohydrate absorption, insulin signalling, adipocytokine signalling, mTOR signalling and inositol phosphate metabolism, which are also enriched in hVIPs targeted via ELMs in viral proteins. Viral modulation of metabolic pathways (Birch *et al.*, 2012; Heaton and Randall, 2011) is being studied in recent years. mTOR signalling pathway is known to be manipulated by pathogens (Martin *et al.*, 2012), our study revealed that AKT, GSK and PIK3K are targeted by many viruses by mimicking pkinase domain binding motifs (Fig. 6 and Supplementary Table 3). HCV has been shown to stimulate phosphoinositol 4 phosphate production via activation of

PI4K3 kinase activation (Berger *et al.*, 2011) consistent with our finding inositol phosphate metabolism being targeted by means of DMIs. Our previous study on Hu-Vir-bridged networks has revealed certain viral proteins acting as articulation points and as a consequence bridge metabolic pathways to the signalling pathways. Aforementioned observations and examples in a nutshell support that host-like linear motifs in viral protein sequences are used for modulation of host functions by viruses.

Although the known protein disorder in the viral genomes vary broadly in the range from 1 to 70%, majority of them are in the range of 10–30% (Pushker *et al.*, 2013; Xue *et al.*, 2012). Our study has also shown that majority of viral proteins are ODPs and MDPs structured (% of disordered residues <30%). Disordered viral proteins (Meyniel-Schicklin *et al.*, 2012) show high-degree centrality. Though viral disordered proteins are less in number, we see that they interact with higher number of hVIPs owing to their disorder and higher number of ELMs. Some viral proteins have shown to contain higher density of ELMs (Garamszegi *et al.*, 2013). Few studies focusing presence of ELMs in disordered proteins of HCV (Fan *et al.*, 2014) and across viral genomes (Hagai *et al.*, 2014) have been reported. There are large numbers of uncharacterized ELMs in disordered regions of viral proteins, which may have functional implications (Hagai *et al.*, 2014; Tompa *et al.*, 2014). In our study, we see viruses utilizing ELMs as a common strategy to manipulate certain pathways. Presence of large numbers of putative ELMs indicate that viruses may use molecular mimicry of ELMs as a strategy to rewire host's regulatory and signalling pathways. Current approaches based on limited structural data are inadequate to uncover all the ELMs involved in Hu-Vir PPIs.

5 Conclusion

Our study has shed light on general molecular characteristics of targeted viral attack on host cells. Viruses have evolved in such a way that they employ targeted attack strategy by interacting with important proteins of Hu-PPI network, which are mostly IDPs. In most of the cases, Hu-Vir PPIs commonly involve at least one IDP providing conformational flexibility to the protein interaction. Viruses were found to utilize intrinsic disorder in human proteins and of viral proteins differentially in different subcellular locations. Interactions with central proteins enable viruses to take control of host network and redirect it towards synthesis of viral genome and proteins. hVIPs were found to be enriched with binding interfaces in disordered region and SLiMs/ELMs in them. Viruses mimic host motifs and utilize them for targeted attack on certain host cellular pathways. Presence of motifs which act as molecular switches in regulation of multiple pathways, high connectivity and global expression, involvement in multiple pathways, slow evolutionary rate, protein abundance and high LE in human proteins makes them Achilles' heel for targeted attack by viral proteins. Viruses irrespective of their families mimic linear motifs in their protein sequences as a common strategy to subvert specified functions undergoing convergent evolution. Information control, functional diversity, protein abundance, evolutionary plasticity of viruses to mimic host ELMs and conformational flexibility are the governing principles of Hu-Vir PPIs. This knowledge can be further exploited for screening and developing broad spectrum of antiviral therapies.

Acknowledgement

The authors thank Dr. Rashna Bhandari for discussion on kinases in signalling pathways and anonymous reviewers for their constructive comments.

Funding

This work was supported by Centre for DNA Fingerprinting and Diagnostics (CDFD) core grant [to H.A.N.] and R.R.H. received CSIR Senior Research Fellowship.

Conflict of interest: none declared.

References

Albert,R. *et al.* (2000) Error and attack tolerance of complex networks. *Nature*, **406**, 378–382.

Apweiler,R. *et al.* (2004) Protein sequence databases. *Curr. Opin. Chem. Biol.*, **8**, 76–80.

Aranda,B. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.

Babu,M.M. *et al.* (2011) Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.*, **21**, 432–440.

Barabási,A. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.

Bellay,J. *et al.* (2011) Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.*, **12**, R14.

Berger,K.L. *et al.* (2011) Hepatitis C virus stimulates the phosphatidylinositol 4-kinase III alpha-dependent phosphatidylinositol 4-phosphate production that is essential for its replication. *J. Virol.*, **85**, 8870–8883.

Birch,E.W. *et al.* (2012) Determining host metabolic limitations on viral replication via integrated modeling and experimental perturbation. *PLoS Comput. Biol.*, **8**, e1002746.

Bowie,A.G. and Unterholzner,L. (2008) Viral evasion and subversion of pattern-recognition receptor signalling. *Nat. Rev. Immunol.*, **8**, 911–922.

Brown,C.J. *et al.* (2002) Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol.*, **55**, 104–110.

Brown,C.J. *et al.* (2010) Comparing models of evolution for ordered and disordered proteins. *Mol. Biol. Evol.*, **27**, 609–621.

Calderwood,M.A. *et al.* (2007) Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl Acad. Sci. USA*, **104**, 7606–7611.

Chatr-aryamontri,A. *et al.* (2009) VirusMINT: a viral protein interaction database. *Nucleic Acids Res.*, **37**, D669–D673.

Daugherty,M.D. and Malik,H.S. (2012) Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.*, **46**, 677–700.

Davey,N.E. *et al.* (2011) How viruses hijack cell regulation. *Trends Biochem. Sci.*, **36**, 159–169.

de Chassey,B. *et al.* (2008) Hepatitis C virus infection protein network. *Mol. Syst. Biol.*, **4**, 230.

Diella,F. *et al.* (2008) Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.*, **13**, 6580–6603.

Dinkel,H. *et al.* (2012) ELM—the database of eukaryotic linear motifs. *Nucleic Acids Res.*, **40**, D242–D251.

Dosztányi,Z. *et al.* (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. *J. Proteome Res.*, **5**, 2985–2995.

Dosztányi,Z. *et al.* (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics*, **25**, 2745–2746.

Driscoll,T. *et al.* (2009) PIG—the pathogen interaction gateway. *Nucleic Acids Res.*, **37**, D647–D650.

Dyer,M.D. *et al.* (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog.*, **4**, e32.

Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.

Fan,X. *et al.* (2014) The intrinsic disorder status of the human hepatitis C virus proteome. *Mol. Biosyst.*, **10**, 1345–1363.

Finn,R.D. *et al.* (2005) iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.

Franzosa,E.A. and Xia,Y. (2011) Structural principles within the human-virus protein-protein interaction network. *Proc. Natl Acad. Sci. USA*, **108**, 10538–10543.

Franzosa,E.A. *et al.* (2012) Toward a three-dimensional view of protein networks between species. *Front. Microbiol.*, **3**, 428.

Freed,E.O. (2002) Viral late domains. *J. Virol.*, **76**, 4679–4687.

Fumagalli,M. *et al.* (2011) Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.*, **7**, e1002355.

Garamszegi,S. *et al.* (2013) Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human–virus protein–protein interaction networks. *PLoS Pathog.*, **9**, e1003778.

Gsponer,J. *et al.* (2008) Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science*, **322**, 1365–1368.

Hagai,T. *et al.* (2014) Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Rep.*, **7**, 1729–1739.

Halehalli,R.R. and Nagarajaram,H.A. (2014) Viral proteins that bridge unconnected proteins and components in human PPI network. *Mol. Biosyst.* **10**, 2448–2458.

Heaton,N.S. and Randall,G. (2011) Multifaceted roles for lipids in viral infection. *Trends Microbiol.*, **19**, 368–375.

Hirsch,V.M. *et al.* (1989) An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature*, **339**, 389–392.

Horton,P. *et al.* (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*, **35**, W585–W587.

Huet,T. *et al.* (1990) Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature*, **345**, 356–359.

Jäger,S. *et al.* (2012) Global landscape of HIV-human protein complexes. *Nature*, **481**, 365–370.

Kanehisa,M. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

Keshava Prasad,T.S. *et al.* (2009) Human protein reference database—2009 update. *Nucleic Acids Res.*, **37**, D767–D772.

Khadka,S. *et al.* (2011) A physical interaction network of dengue virus and human proteins. *Mol. Cell. Proteomics*, **10**, M111.012187.

Kim,P.M. *et al.* (2008) The role of disorder in interaction networks: a structural analysis. *Mol. Syst. Biol.*, **4**, 179.

Kinsella,R.J. *et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)*, **2011**, bar030.

König,R. *et al.* (2008) Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*, **135**, 49–60.

Krishnan,M.N. *et al.* (2008) RNA interference screen for human genes associated with West Nile virus infection. *Nature*, **455**, 242–245.

Li,L. *et al.* (2001) Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J.*, **20**, 3272–3281.

Martin,S. *et al.* (2012) The battle over mTOR: an emerging theatre in host-pathogen immunity. *PLoS Pathog.*, **8**, e1002894.

Meyniel-Schicklin,L. *et al.* (2012) Viruses and interactomes in translation. *Mol. Cell. Proteomics*, **11**, M111.014738.

Midic,U. *et al.* (2009) Unfoldomics of human genetic diseases: illustrative examples of ordered and intrinsically disordered members of the human diseasome. *Protein Pept. Lett.*, **16**, 1533–1547.

Morens,D.M. *et al.* (2004) The challenge of emerging and re-emerging infectious diseases. *Nature*, **430**, 242–249.

Navratil,V. *et al.* (2011) When the human viral infectome and diseasome networks collide: towards a systems biology platform for the aetiology of human diseases. *BMC Syst. Biol.*, **5**, 13.

Nguyen Ba,A.N. *et al.* (2012) Proteome-wide discovery of evolutionary conserved sequences in disordered regions. *Sci. Signal.*, **5**, rs1.

Nielsen,R. *et al.* (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.*, **3**, e170.

Pichlmair,A. *et al.* (2012) Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature*, **487**, 486–490.

Pushker,R. *et al.* (2013) Marked variability in the extent of protein disorder within and between viral families. *PLoS One*, **8**, e60724.

Quevillon,E. *et al.* (2005) InterProScan: protein domains identifier. *Nucleic Acids Res.*, **33**, W116–W120.

Rennekamp,A.J. and Lieberman,P.M. (2010) Initiation of lytic DNA replication in Epstein-Barr virus: search for a common family mechanism. *Future Virol.*, **5**, 65–83.

Salah,Z. and Aqeilan,R.I. (2011) WW domain interactions regulate the Hippo tumor suppressor pathway. *Cell Death Dis.*, **2**, e172.

Sawyer,S.L. and Elde,N.C. (2012) A cross-species view on viruses. *Curr. Opin. Virol.*, **2**, 561–568.

Shannon,P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Simonis,N. *et al.* (2012) Host-pathogen interactome mapping for HTLV-1 and -2 retroviruses. *Retrovirology*, **9**, 26.

Snel,B. *et al.* (2002) The identification of functional modules from the genomic association of genes. *Proc. Natl Acad. Sci. USA*, **99**, 5890–5895.

Tompa,P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci.*, **27**, 527–533.

Tompa,P. (2009) *Structure and Function of Intrinsically Disordered Proteins*, 1st ed. Chapman and Hall/CRC Press, London.

Tompa,P. *et al.* (2014) A million peptide motifs for the molecular biologist. *Mol. Cell*, **55**, 161–169.

Uetz,P. *et al.* (2006) Herpesviral protein networks and their interaction with the human proteome. *Science*, **311**, 239–242.

Uversky,V.N. and Dunker,A.K. (2010) Understanding protein non-folding. *Biochim. Biophys. Acta*, **1804**, 1231–1264.

Uversky,V.N. *et al.* (2008) intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.*, **37**, 215–246.

Van Roey,K. *et al.* (2012) Motif switches: decision-making in cell regulation. *Curr. Opin. Struct. Biol.*, **22**, 378–385.

Wang,M. *et al.* (2012) PaxDb, a database of protein abundance averages across all three domains of life. *Mol. Cell. Proteomics*, **11**, 492–500.

Ward,J.J. *et al.* (2004a) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, **337**, 635–645.

Ward,J.J. *et al.* (2004b) The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**, 2138–2139.

Weatheritt,R.J. *et al.* (2012) iELM—a web server to explore short linear motif-mediated interactions. *Nucleic Acids Res.*, **40**, W364–W369.

Winberg,G. *et al.* (2000) Latent membrane protein 2A of Epstein-Barr virus binds WW domain e3 protein-ubiquitin ligases that ubiquitinate B-cell tyrosine kinases. *Mol. Cell. Biol.*, **20**, 8526–8535.

Woolhouse,M.E.J. *et al.* (2005) Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.*, **20**, 238–244.

Wu,C. *et al.* (2009) BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.*, **10**, R130.

Wuchty,S. *et al.* (2010) Viral organization of human proteins. *PLoS One*, **5**, e11796.

Xue,B. *et al.* (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.*, **30**, 137–149.

Yang,Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586–1591.

Yellaboina,S. *et al.* (2011) DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.*, **39**, D730–D735.

Zhang,L. *et al.* (2009) Analysis of vaccinia virus-host protein-protein interactions: validations of yeast two-hybrid screenings. *J. Proteome Res.*, **8**, 4311–4318.

Zhou,H. *et al.* (2008) Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host Microbe*, **4**, 495–504.