

VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus–host interaction networks

Vincent Navratil^{1,2,3,4,*}, Benoît de Chassey^{1,3,4}, Laurène Meyniel^{1,3,4},
Stéphane Delmotte^{1,4,5}, Christian Gautier^{1,4,5}, Patrice André^{1,3,4,6},
Vincent Lotteau^{1,3,4} and Chantal Rabourdin-Combe^{1,3,4}

¹Université de Lyon, ²INRA, UMR754, Ecole Nationale Vétérinaire de Lyon, ³INSERM, U851, 21 avenue Tony Garnier, Lyon, F-69007, ⁴Université Lyon 1, IFR128, ⁵CNRS, UMR5558, Laboratoire de biologie et de biométrie évolutive and ⁶Hospices Civils de Lyon, Hôpital de la Croix-Rousse, Laboratoire de virologie, France

Received August 8, 2008; Revised October 9, 2008; Accepted October 10, 2008

ABSTRACT

Infectious diseases caused by viral agents kill millions of people every year. The improvement of prevention and treatment of viral infections and their associated diseases remains one of the main public health challenges. Towards this goal, deciphering virus–host molecular interactions opens new perspectives to understand the biology of infection and for the design of new antiviral strategies. Indeed, modelling of an infection network between viral and cellular proteins will provide a conceptual and analytic framework to efficiently formulate new biological hypothesis at the proteome scale and to rationalize drug discovery. Therefore, we present the first release of VirHostNet (Virus–Host Network), a public knowledge base specialized in the management and analysis of integrated virus–virus, virus–host and host–host interaction networks coupled to their functional annotations. VirHostNet integrates an extensive and original literature-curated dataset of virus–virus and virus–host interactions (2671 non-redundant interactions) representing more than 180 distinct viral species and one of the largest human interactome (10672 proteins and 68252 non-redundant interactions) reconstructed from publicly available data. The VirHostNet Web interface provides appropriate tools that allow efficient query and visualization of this infected cellular network. Public access to the VirHostNet knowledge-based system is available at <http://pbildb1.univ-lyon1.fr/virhostnet>.

INTRODUCTION

Eukaryotic cells express a large panel of proteins that coordinately participate to the cellular machinery through a highly connected and regulated network of protein–protein interactions (1). Physical architecture of model organisms and human cellular protein networks exhibits a strong robustness against random failures, and strikingly a high sensitivity to targeted attacks on highly connected and central proteins, also called ‘hubs’ (2,3). Cellular protein network is not static and its robustness may change dynamically according to various factors like tissue and cell-line origins, signals received by cellular environment or more specifically during viral infections (4). Replication and pathogenesis of viruses depend on a complex interplay between viral and host cellular proteins both acting through a complex network of protein–protein interactions. In order to evade the cell innate immune response and/or to favour their own replication and transmission, viruses have developed strategies to hijack central functions of the cell (5–7). Viruses also use intra-viral, i.e. virus–virus, protein–protein interactions for virion assembly or viral egress from the cell. Accumulation of functional perturbations associated with such virus–virus and virus–host protein–protein interactions may lead to severe and complex diseases, like the development of cancers (8,9). From a systems biology perspective, a deeper understanding of infectious diseases may rely on an exhaustive characterization of all potential interactions occurring between proteins encoded by viruses and those expressed in infected cells (10). Thus, integration of all protein–protein interactions into an infected cellular network, or ‘infectome’, is a great challenge that may provide a powerful framework for virtual modelling and analysis of viral infection.

*To whom correspondence should be addressed. Tel: +33 4 37 28 23 57; Fax: +33 4 37 28 23 41; Email: navratil@univ-lyon1.fr

The first draft of the human cellular network, also referred to the human interactome, has been explored at the proteome-wide level by the mean of high-throughput experiments such as yeast-two hybrid screens (11,12) or tap-tag procedure (13). The overall quality and completeness of this human cellular network has been significantly improved thanks to systematic approaches based on text-mining and literature-curated interactions extracted from low-throughput experiments. Many generalized and specialized databases are involved in the integration of these protein–protein interactions, such as BIND (14), MINT (15), INTACT (16), HPRD (17), DIP (18), BIOGRID (19), REACTOME (20), GENERIF (21) and NETWORKIN (22). However, the low redundancy of interactions found between these databases has raised the need to unify such data resources for human and model organisms (23). Concerning virus–virus and virus–host protein–protein interactions, few high-throughput experiments have been achieved, except some yeast-two hybrid screens completed for Herpes viruses (EBV, KSH, VZV, HSV-1) (24–26) and SARS (27). Although some generalist databases like BIND, MINT, INTACT and HIV-GENERIF provide access to virus–virus and virus–host protein–protein interactions, no systematic approach has been reported to exhaustively mine and curate all interactions that have accumulated in scientific publications.

In this context, we have developed VirHostNet (Virus–Host Network), a public knowledge-based system specialized in the management, analysis and integration of virus–virus, virus–host and host–host interactions as well

as their functional annotations in the cell. Based on an extensive scientific literature expertise, VirHostNet provides a high-confidence resource of manually curated interactions defined for a wide range of viral species. The content of this high-confidence dataset has been illustrated by the analysis of cellular functions and pathways enriched in proteins targeted by one or many viruses. An integrated cellular network has also been reconstructed from public data and combined with viral data to provide the first draft of the infected cellular network. In addition, an original Web interface has been developed, which provides multi-criteria query and visualization tools for infection network navigation. The utility of the visualisator has been exemplified by network representation of the mTOR pathway and its interplay with viruses.

INFECTION NETWORK INTEGRATION

A bioinformatics pipeline was developed to fully integrate virus–virus, virus–host and host–host protein–protein interactions gathered from a wide range of public databases, with those mined from scientific literature and curated by VirHostNet experts (Figure 1). In addition to the management of this large protein interaction resource, VirHostNet integrates contextual information concerning interacting proteins, like structural and functional annotations of proteins: Gene Ontology term (28), KEGG pathway (29), INTERPRO domain (30). All these data were integrated into a knowledge-based system implemented by

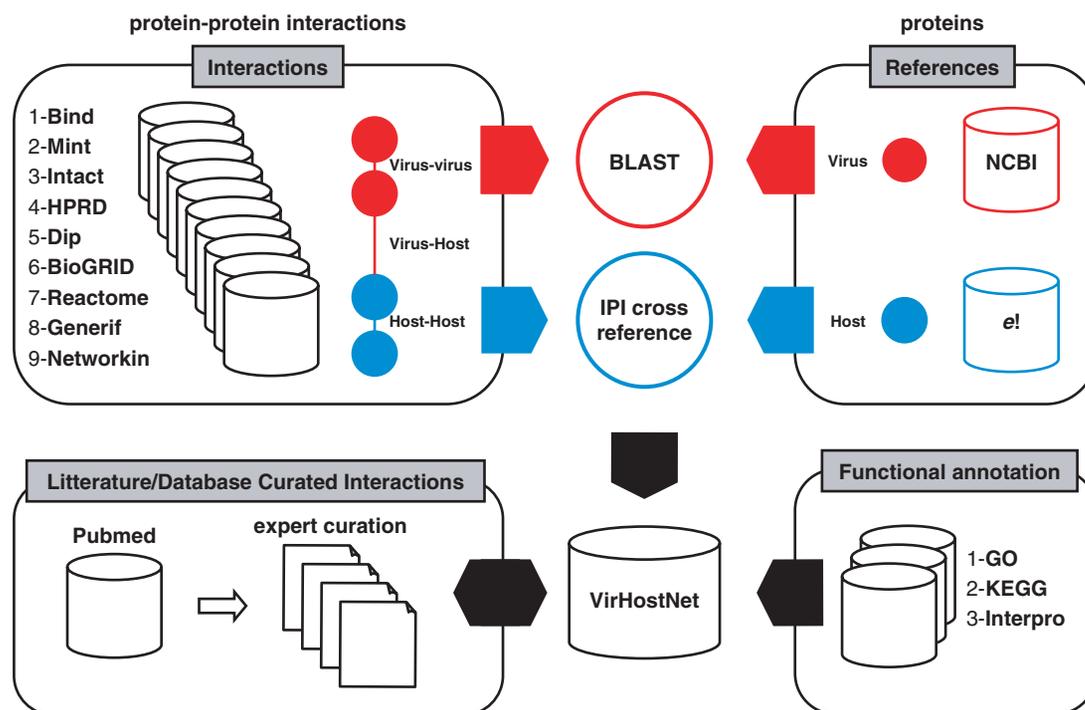


Figure 1. Flow chart of the VirHostNet knowledge-based system integration process. Virus–virus, virus–host and host–host protein–protein interactions were integrated from 10 external public databases. BLAST and IPI cross-reference databases were used to link protein interactor identification numbers to NCBI (viruses) and ENSEMBL protein accession numbers (e!) (host) databases respectively. Virus–Virus and Virus–Host protein–protein interactions were also extracted from literature (Literature Curated Interactions) and curated from databases (Database Curated Interactions). Functional annotations of cellular proteins were extracted from Gene Ontology, KEGG and INTERPRO databases.

using PostgreSQL DataBase Management System (release 8.2.6).

Public Database integration

The low level of redundancy observed among available databases involved in molecular interactions management has emphasized the need to integrate these heterogeneous data sources (23). Virus–virus, virus–host and host–host protein–protein interactions and meta-data related to experimental procedures or publications were extracted from 10 databases (BIND, MINT, INTACT, HPRD, DIP, BIOGRID, REACTOME, GENERIF, HIV-GENERIF, NETWORKIN) (Figure 1). Due to the heterogeneity of protein sequence identification found across these databases (i.e. gene identification number, gene name, protein accession number, protein name), NCBI and ENSEMBL protein sequence databases were chosen to unify virus and host proteins respectively (see Supplementary Table 1A). Towards this end, the IPI database system (31) was chosen to cross-reference all the human protein sequences to ENSEMBL protein accession numbers. In addition, viral protein sequences defined at EMBL and UNIPROT were mapped on NCBI protein sequences by using BLAST Alignment software (32). Protein cross-referencing led to the definition of non-redundant protein–protein interactions that were in many cases defined in different databases, publications or supported by distinct experimental procedures (see Supplementary Table 1B). Thus, all information associated with non-redundant interactions, like database origin, experimental procedure description in PSI-MI 2.5 standard format (33) or PUBMED identification (PMID) number, were retrieved in VirHostNet to provide the most documented interactions. This compilation of interaction meta-data will facilitate data quality filtering based on the number of databases, methods or PMIDs used (34,35).

Literature- and Database-curated interactions

An automatic text-mining pipeline was developed and plugged into the VirHostNet system in order to prioritize scientific papers for protein–protein interaction curation. As a first step, all abstracts containing keywords related to both viruses and experimental procedures used for interactions identification (mainly yeast-two hybrid, co-immunoprecipitation, pull-down and tandem affinity purification) were extracted for an in-depth expertise. During curation, protein–protein interactions were carefully annotated according to: (i) the protein accession numbers of each of the protein interactor, the human and/or viral proteins being respectively referenced to ENSEMBL and NCBI accession numbers; (ii) the molecular interaction methods based on the PSI-MI 2.5 ontology vocabulary; and (iii) the PMIDs. Based on 1174 selected PMIDs, literature curation led to the annotation of 2186 redundant interactions in 723 papers (Supplementary Table 2). This effort significantly complemented data from public databases with 1297 new non-redundant protein–protein interactions. In order to provide a higher level of data accuracy, virus–virus and virus–host protein–protein interactions from public databases were also carefully inspected.

From 2294 PMIDs for which at least one protein–protein interaction was defined, database curation led to the validation of 2261 redundant interactions found in 789 papers, corresponding to 1374 confirmed non-redundant protein–protein interactions (Supplementary Table 2). Strikingly, our experts confirmed 20% of BIND and GENERIF (HIV) against 90–95% for MINT and INTACT data. One reason is that all protein–protein interactions defined by functional associations and/or genetic interactions between proteins were discarded from BIND and HIV-GENERIF.

Infection network content

To our knowledge, VirHostNet provides the largest and the most confident infected cellular network. This network is composed of 2671 virus–virus and virus–host non-redundant protein–protein interactions concerning 180 distinct viral species. The curated protein–protein interactions were mainly defined by low-throughput and high-throughput yeast two-hybrid screens (40%), co-immunoprecipitation (24%) and pull-down (21%) (Figure 2A). Even if only 65% of interactions rely on a single experimental procedure, a total of 944 protein–protein interactions (35%) were defined by at least two independent methods, in good agreement with other high-confidence databases (36) (Figure 2B). All these interactions were defined in 36 distinct viral families, underlying the broad taxonomical diversity provided by

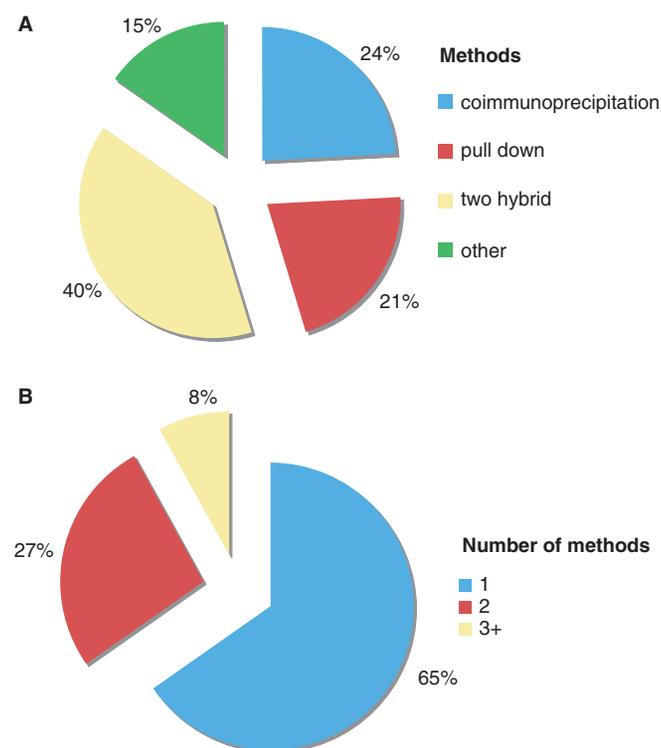


Figure 2. Virus–virus and virus–host protein–protein interaction methods summary. (A) Distribution of protein–protein interactions according to experimental procedures (yeast-two hybrid, co-immunoprecipitation, pull-down and others). (B) Distribution of protein–protein interactions identified by one, two or more distinct methods.

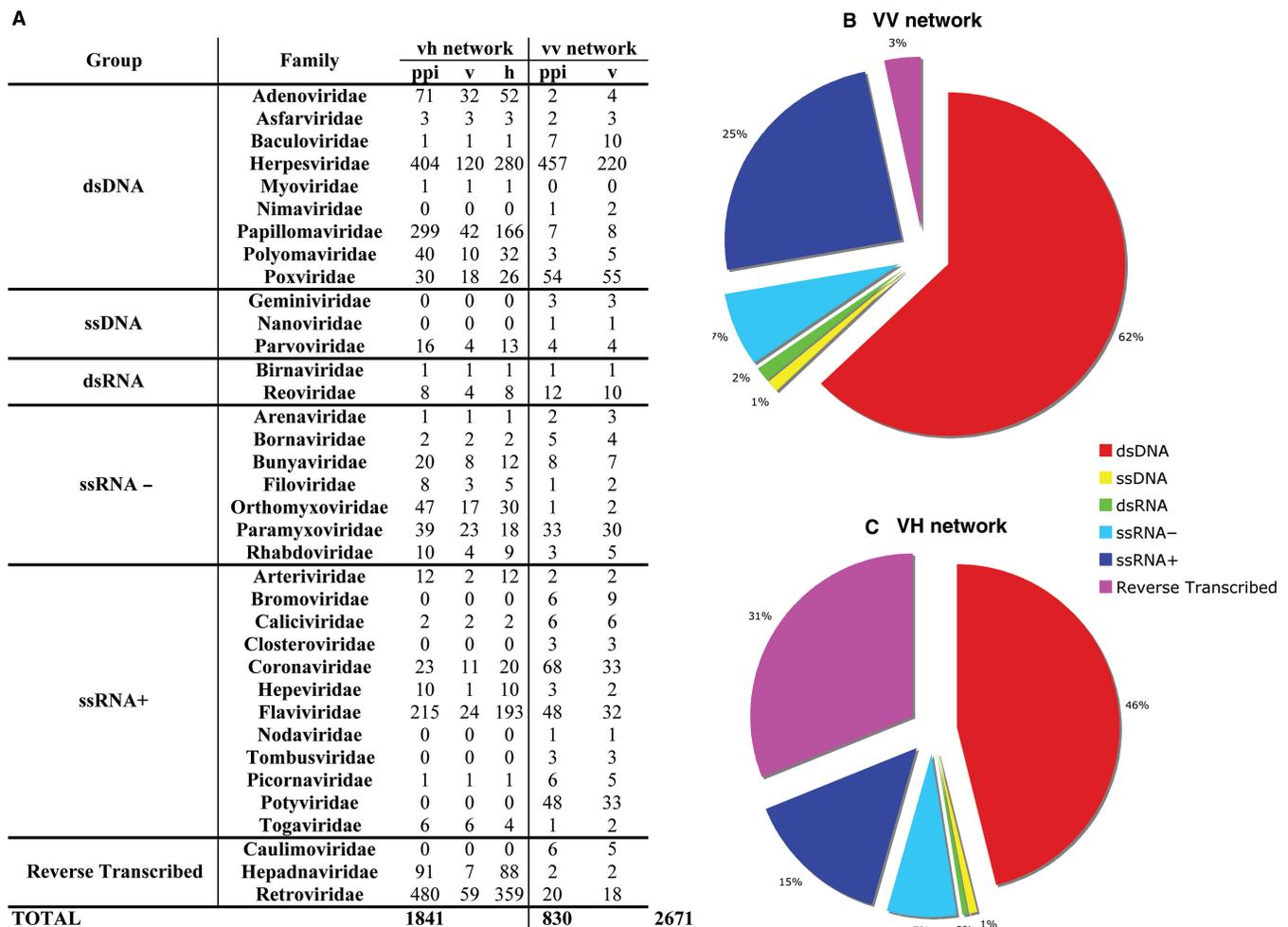


Figure 3. VirHostNet protein-protein interaction content. (A) In each Baltimore group (dsDNA, ssDNA, dsRNA, ssRNA-, ssRNA+, retro transcribed), for each viral family, the number of virus-host protein-protein interactions (ppi) and the associated number of viral (v) and host cellular protein (h) interactors, the number of virus-virus protein-protein interactions (ppi) and the associated number of viral (v) interactors are given. (B) Distribution of virus-virus interactions according to viral Baltimore groups. (C) Distribution of virus-host interactions according to viral Baltimore groups.

VirHostNet (Figure 3A). In addition, the distribution of interactions observed among viral Baltimore groups should allow large-scale comparative study of virus-virus (Figure 3B) and virus-host networks (Figure 3C).

INFECTION NETWORK ANALYSIS

In the infection network, the virus-host interactions occurred between 407 viral proteins and 1012 human proteins, suggesting the strong tendency of viruses to interact with a large number of cellular proteins. In order to characterize cellular functions targeted by the viral machinery, we performed functional enrichment analysis of host proteins interacting with viruses, by using Gene Ontology and KEGG databases and the same methodology described by Zheng and Wang (37) (Supplementary Tables 3 and 4, respectively). The results showed that viruses interact significantly with a large panel of cellular functions (e.g. cell cycle, apoptosis, cell communication, protein transport)

and with canonical signalling pathways (e.g. Jak-Stat, Toll-like Receptor, MAPK, TGF- β , mTOR). The majority of these functions and pathways have already been described to participate in either viral infectious cycle, cellular anti-viral mechanisms or viral associated diseases (38). Interestingly, analysis of KEGG pathways revealed cellular mechanisms poorly documented in the case of viral infections. One example is focal adhesion, a pathway involved in cell contact with the extracellular matrix and in many other cellular processes including invasion, motility, proliferation and apoptosis (39). Indeed, on 202 protein members of the focal adhesion pathway, more than 25% (59) were found significantly targeted (exact Fisher test, Benjamini-Hochberg multiple correction test P -value < 0.05) by at least one viral protein in 36 distinct viral taxons. This may suggest the central role of focal adhesion during viral infections and its potential impact on viral induced cancer development that might be associated for instance to the loss of cellular adhesion. Although cellular functions of proteins are far from

being completely known and/or annotated in public databases, based on the 'guilty by association' concept the human protein-protein network may serve as a template to complete our understanding on cellular functions perturbed during viral infection. In order to include virus-virus and virus-host interactions in their cellular context, a human-human protein interaction network containing roughly 70 000 non-redundant protein-protein interactions and 10 000 proteins was built from public databases (details on interaction methods distribution are given in Supplementary Figure 1). Thus, based on roughly 40 000 unique proteins annotated in ENSEMBL, 25% (10 000/40 000) are connected within the human protein network. Analysis of the infection network revealed that surprisingly 88% (881/1012) of targeted human proteins interact with at least one cellular protein. Thus, targeted proteins tend to physically interact in the cell and may probably participate in cross-linked functions and pathways. Based on protein neighbourhood or sub-networks, the human protein-protein interaction network may help to elucidate new protein regulators or modular functions associated to viral or cellular anti-viral strategies.

VIRHOSTNET WEB INTERFACE

A user friendly and powerful Web interface based on PHP, JAVA and AJAX technologies was developed. This interface is intended to facilitate: (i) protein and contextual based queries (ii) protein-protein interaction quality filtering and display; (iii) protein-protein interaction network query (viral and host neighbours, virus-virus, virus-host, host-host sub-networks); and (iv) protein network graphical visualization. Description and examples of the database features are available in the Wiki page of the VirHostNet Web site (<http://pbildb1.univ-lyon1.fr/virhostnet/wiki>).

VirHostNet query interface

Once logged-in, VirHostNet users can directly query the knowledge base by using a wide range of information concerning viral (e.g. NCBI protein name or accession number) or human proteins (e.g. ENSEMBL gene or protein accession number, NCBI gene name, REFSEQ protein accession number and UNIPROT primary and secondary accession numbers) (Figure 4A). AJAX technology was incorporated to control protein name and accession number availability in VirHostNet. Another important feature of the interface is batch query. It allows in-depth analysis of interaction profiles with cellular and/or viral proteins from a list of proteins defined for instance in high-throughput studies (microarray, yeast-two hybrid). A list of genes or proteins of interest can also be assessed by the mean of contextual information, such as taxonomical information, Gene Ontology terms, KEGG pathways and INTERPRO protein domains. These properties offer a unique access to protein-protein interaction networks: (i) associated to a specific virus taxon or (ii) underlying canonical sub-cellular localization, cellular functions and pathways.

To access protein-protein interactions from a list of proteins (Figure 4B), users have: (i) to select all or a subset of proteins of interest; (ii) to define the kind of interactions to retrieve (virus-virus, virus-host and host-host) and their database origin and (iii) to select the mode of navigation to perform, either protein neighbours or protein subgraph (Figure 4C). Neighbours are viral or cellular proteins interacting directly with a protein of interest. A subgraph (or subnetwork) is a graph made of all interactions between a set of proteins. The resulting host-host, virus-host and/or virus-virus protein-protein interactions are then given into a tabulated format in three independent tabbed panels (Figure 4D). For each protein-protein interaction, users have a privileged access to interaction meta-data (Figure 4E) and a colour code highlights interactions that have been checked by VirHostNet experts.

VirHostNet network visualisator

Beside table representation of protein-protein interactions, a more dynamic and interactive network visualization tool was specifically developed for graph representation of infection networks. This new network visualisator was fully implemented in Jung 2.0 (<http://jung.sourceforge.net>) as a Java Web applet. It efficiently takes into account viruses and host nodes dichotomy for both graph rendering (colour of nodes) and navigation (host and viral neighbours). The visualisator provides also sliders to dynamically filter graphs based on the number of PMIDs and experimental procedures used to identify interactions. Additional features are also provided to draw protein node size according to the number of viral or host interacting partners (i.e. their degree into the virus-virus, virus-host, host-host networks) or to highlight targeted proteins. As a case study, we built a protein sub-network view of the mTOR KEGG signalling pathway that has been found significantly enriched in targeted proteins (Supplementary Table 4). Indeed, the modulation of PI3K-Akt-mTOR signal transduction pathway by viruses has been shown to play a crucial role in inhibition of apoptosis, cell survival, cell transformation, viral replication and viral assembly (40,41). To identify and compare how viruses interplay with this network, virus-host protein interactions annotated by VirHostNet were added. This mTOR-infection network is composed of 42 cellular interacting proteins (blue nodes), 10 viral proteins (coloured nodes according to viral taxonomy), 84 host-host (blue edges) and 14 virus-host (red edges) physical protein-protein interactions (Figure 5). Protein network visualization showed that cellular proteins of this pathway are highly inter-connected in contrast to the classical representation given by the KEGG pathway, underlying the extreme complexity and regulation of this pathway. Moreover, graph visualization allows identifying viral proteins targeting multiple cellular proteins (e.g. NS5A protein interacting with AKT1, PDPK1 and PIK3CB) and reciprocally cellular proteins interacting with multiple viral proteins (e.g. HIF1A interacting with LANA of Human Herpes Virus 8 type P and X protein of Hepatitis B Virus). Hence, the VirHostNet interface allows users to visualize

A

- Home
- Query
- Gene
- Protein
- Function
- Pathway
- Domain
- Publication
- Browse
- Info
- Account

B

Number of genes : 49

acc	name
<input checked="" type="checkbox"/> ENSG00000017427	IGF1 Insulin-like growth factor (Somatomedin C) [Source:Uniprot]
<input checked="" type="checkbox"/> ENSG000000051382	PIK3CB Phosphatidylinositol 3-kinase class III beta subunit (PtdIn3K) [Source:Uniprot]
<input checked="" type="checkbox"/> ENSG000000071242	RPS6KA2 Ribosomal protein S6K-alpha 2 (p90-RSK 2) (Ribosomal protein S6K) [Source:Uniprot]
<input checked="" type="checkbox"/> ENSG000000072133	RPS6KA6 Ribosomal protein S6K-alpha 6 (p90-RSK 6) (Ribosomal protein S6K) [Source:Uniprot]
<input checked="" type="checkbox"/> ENSG000000083290	ULK2 Serine/threonine-protein kinase 2 [Source:Uniprot]
<input checked="" type="checkbox"/> ENSG000000100030	MAPK1 Mitogen-activated protein kinase 1 (Mitogen-activated protein kinase) [Source:Uniprot]

C

1 ---- select check all uncheck all save list

2 ---- preference[+]

3 ---- operation

D

x	taxonomy_name_1	acc_1	official_1	taxonomy_name_2	acc_2	official_2	info
<input checked="" type="checkbox"/>	Homo sapiens	ENSG000000117461	PIK3R3	Homo sapiens	ENSG000000121879	PIK3CA	?
<input checked="" type="checkbox"/>	Homo sapiens	ENSG000000117676	RPS6KA1	Homo sapiens	ENSG000000117676	RPS6KA1	?
<input checked="" type="checkbox"/>	Homo sapiens	ENSG000000118046	STK11	Homo sapiens	ENSG000000118046	STK11	?

E

-- molecular interaction informations --

nr_ppi_id	ppi_id	pmid	method_id	method_name	database	edit
15736	50497	17353931	MI:0006	anti bait coimmunoprecipitation	virhostnet	<input type="button" value="edit"/>
15736	50498	14730349	MI:0105	structure based prediction	bind	<input type="button" value="edit"/>

Figure 4. The VirHostNet Query interface. (A) Menu of the VirHostNet interface. Users are allowed to search for protein–protein interactions according to gene, protein, biological function (GO), pathway (KEGG), protein domain (INTERPRO) or PMID information. For example, in the ‘Pathway’ query interface of the ‘Query’ menu, users can enter ‘mTOR’, then select the right pathway defined in KEGG database and ask for proteins annotated in this pathway (see result in Figure 4B). (B) List of the 49 proteins belonging to the mTOR pathway. The ‘?’ link at the end of each protein line allows users to access functional annotations of the protein in external databases. (C) Protein–protein interactions research panel. Users can select one or more proteins from the list of proteins (by default, all proteins are selected). Then, from the preferences panel, they have to choose the kind of interaction to retrieve (virus–virus, virus–host or host–host) or the database origin. Finally, they have to choose an operation (‘neighbours’ or ‘sub-graph’). The ‘neighbours’ button allows users to retrieve all direct protein–protein interactions from a single protein or a list of selected proteins. The ‘sub-graph’ button allows users to retrieve only protein–protein interactions occurring between selected proteins (see result in Figure 4D). (D) List of the 89 protein–protein interactions occurring between proteins of the mTOR pathway. For the interacting proteins, NCBI taxonomy name, ENSEMBL gene acc (host)—NCBI protein acc (viruses) and official gene name (host)—product name (viruses) are given. The ‘?’ link at the end of each interaction line allows users to access interaction annotations (Figure 4E). In the protein–protein interactions research panel, in addition to ‘neighbours’ and ‘sub-graph’ buttons, the ‘visualize’ button allows from a list of interactions the graphical visualization of the resulting network (see result in Figure 5). (E) More detailed information concerning protein–protein interactions are available. PMID, PSI-MI method accession number and database origin of the selected interaction are given.

protein interaction networks associated to any kind of GO term, KEGG pathway, list of proteins or keywords and to analyse how they interplay with viruses.

CONCLUSION AND PERSPECTIVES

VirHostNet provides now a public access to the largest known resource of integrated virus–virus, virus–host and host–host protein interaction networks. Literature- and database-curated interactions have led to the definition of an original and high-confidence protein–protein interactions dataset. We have briefly illustrated the need of this high-confidence dataset for the characterization of cellular functions targeted by viruses. This resource may also be crucial for network-based analysis of molecular mechanisms involved during viral infections, such as cellular network properties disturbed after the connection of viral proteins. VirHostNet will also provide a backbone for

automatic screening of specific protein domains or peptides motifs associated to virus–host interactions and hence may help to delineate at the proteome-wide scale footprints in both viral and host proteins sequences. VirHostNet will allow systematic prediction of virus–host protein–protein interactions based on sequence homology criteria between closely related viral proteins. The knowledge-based system is also intended to integrate virus–host protein–protein interactions data derived by our team from high-throughput yeast-two hybrids experiments (Orthomyxovirus, Paramyxovirus, Flavivirus...). Thus, the availability of virus–virus and virus–host networks for a broad range of viruses will encourage comparative analysis and will be very helpful for the identification of molecular interactions associated to viral pathogenesis or virulence. As virus–host and virus–virus protein–protein interactions curation is one of the central features of the VirHostNet knowledge base, one of our missions is to keep these data up to date

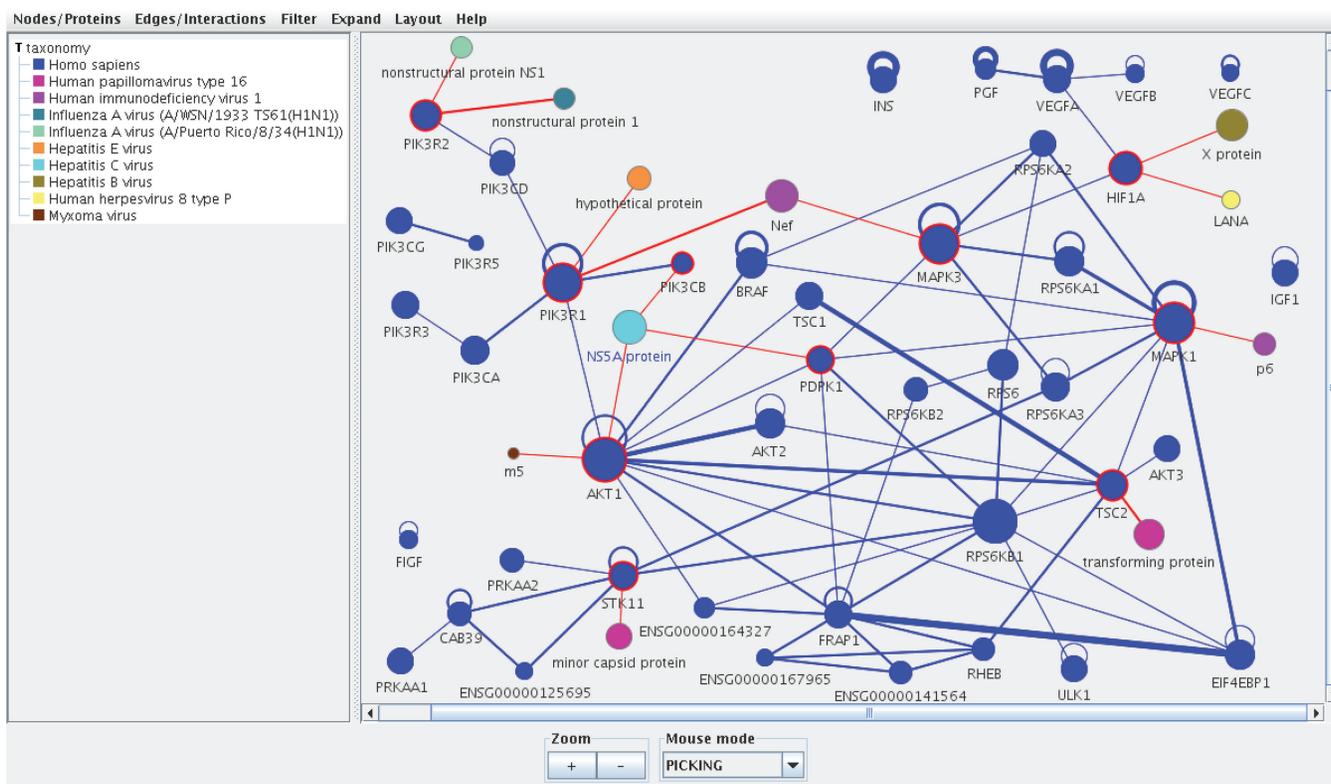


Figure 5. VirHostNet Visualisator. Graph representation of the mTOR-infection network is drawn (right of the applet). Only interacting proteins of the mTOR pathway are represented (cellular proteins = blue nodes; host–host protein–protein interactions = blue edges). Viral interacting proteins are also added to define the infection network (viral proteins = coloured nodes according to viral taxonomy; virus–host protein–protein interactions = red edges). Cellular proteins of the network targeted by at least one viral protein (red stroke) are highlighted. The width of the edges is roughly proportional to the number of PMIDs used to describe the interaction. The width of the nodes is roughly proportional to the cellular degree, i.e. the number of cellular partners in the whole network. The taxonomical classification of viral proteins interacting with the mTOR network is given (left panel of the applet). The menu of the applet (top of the applet) allows users to define nodes and edges preferences (width, colour, labelling), interaction filtering (according to the number of methods, number of PMIDs or both), network navigation (expand viral or host neighbours) and graph layout. To obtain this figure, from the list of the 89 selected protein–protein interactions obtained in Figure 4E, users have to click on ‘visualize’ button. Then from the applet, in picking mode, they have to select all proteins and ask for viral proteins neighbours (in expand menu).

continuously from data published in scientific journals. The update of public databases will occur at least once or twice a year in order to keep the data as current as possible. In the next future, integration of other host species, such as mammals or insects, is envisaged. This will facilitate comparison of interaction profiles among different hosts and thus may help to elucidate the molecular basis underlying the ability of some viruses to overcome the inter-species barrier. Efforts will be made to facilitate data exchange with other generalist databases (MINT, INTACT) and to add Web2.0 capabilities to the Web interface (save, comparison and analysis of user customized networks). Altogether, VirHostNet provides an entry gate for proteome wide analysis of the virus–host system and will greatly help scientists willing to take advantage of functional genomic and systems biology to decipher viral infection, evolution and pathogenesis mechanisms and/or to rationalize anti-viral drug design.

DATABASE ACCESS AND FEEDBACK

Public access to the VirHostNet knowledge base is available at <http://pbildb1.univ-lyon1.fr/virhostnet>. Access can

be made either anonymously (by default) or by creating a personal account (register in the account menu). On simple request, this personal account allows users to participate to the literature-curation effort. Literature-curated and Database-curated protein–protein interactions flat files are available in a tabulated format on request. Contact V.N. (navratil@univ-lyon-1.fr) for more information.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors wish to thank Philippe Mangeot, Isabel Pombo-Grégoire, Agnès Pommier, Stéphane Schicklin, Juliette de Chasse, Sandy Navratil and Justine Navratil for critical reading of the manuscript. We also acknowledge Pierre-Olivier Vidalain and all the I-MAP team for helpful discussions and PRABI DOUA for technical assistance and maintenance of the VirHostNet server.

FUNDING

This work was funded by INRA, Université Lyon 1, INSERM, Rhône Alpes region and the French Ministry of Industry. V.N. is supported by a grant from INRA.

Conflict of interest statement. None declared.

REFERENCES

- Hartwell,L.H., Hopfield,J.J., Leibler,S. and Murray,A.W. (1999) From molecular to modular cell biology. *Nature*, **402**, C47–C52.
- Albert,R., Jeong,H. and Barabasi,A.L. (2000) Error and attack tolerance of complex networks. *Nature*, **406**, 378–382.
- Jeong,H., Mason,S.P., Barabasi,A.L. and Oltvai,Z.N. (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41–42.
- Kitano,H. (2007) Biological robustness in complex host-pathogen systems. *Prog. Drug Res.*, **64**, 239, 241–263.
- Katze,M.G., He,Y. and Gale,M. Jr. (2002) Viruses and interferon: a fight for supremacy. *Nat. Rev. Immunol.*, **2**, 675–687.
- Goff,S.P. (2007) Host factors exploited by retroviruses. *Nat. Rev. Microbiol.*, **5**, 253–263.
- Moore,C.B. and Ting,J.P. (2008) Regulation of mitochondrial antiviral signaling pathways. *Immunity*, **28**, 735–739.
- Hebner,C.M. and Laimins,L.A. (2006) Human papillomaviruses: basic mechanisms of pathogenesis and oncogenicity. *Rev. Med. Virol.*, **16**, 83–97.
- Ahmed,N. and Heslop,H.E. (2006) Viral lymphomagenesis. *Curr. Opin Hematol.*, **13**, 254–259.
- Tan,S.L., Ganji,G., Paeper,B., Proll,S. and Katze,M.G. (2007) Systems biology and the host response to viral infection. *Nat. Biotechnol.*, **25**, 1383–1389.
- Stelzl,U., Worm,U., Lalowski,M., Haenig,C., Brembeck,F.H., Goehler,H., Stroedicke,M., Zenkner,M., Schoenherr,A. *et al.* (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957–968.
- Rual,J.F., Venkatesan,K., Hao,T., Hirozane-Kishikawa,T., Dricot,A., Li,N., Berriz,G.F., Gibbons,F.D., Dreze,M. *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173–1178.
- Ewing,R.M., Chu,P., Elisma,F., Li,H., Taylor,P., Climie,S., McBroom-Cerajewski,L., Robinson,M.D., O'Connor,L., Li,M. *et al.* (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.*, **3**, 89.
- Alfarano,C., Andrade,C.E., Anthony,K., Bahroos,N., Bajec,M., Bantoft,K., Betel,D., Bobechko,B., Boutilier,K., Burgess,E. *et al.* (2005) The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.*, **33**, D418–D424.
- Chatr-aryamontri,A., Ceol,A., Palazzi,L.M., Nardelli,G., Schneider,M.V., Castagnoli,L. and Cesareni,G. (2007) MINT: the Molecular INTERaction database. *Nucleic Acids Res.*, **35**, D572–D574.
- Kerrien,S., Alam-Faruque,Y., Aranda,B., Bancarz,I., Bridge,A., Derow,C., Dimmer,E., Feuermann,M., Friedrichsen,A., Huntley,R.D. *et al.* (2007) IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.
- Mishra,G.R., Suresh,M., Kumaran,K., Kannabiran,N., Suresh,S., Bala,P., Shivakumar,K., Anuradha,N., Reddy,R., Raghavan,T.M. *et al.* (2006) Human protein reference database—2006 update. *Nucleic Acids Res.*, **34**, D411–D414.
- Salwinski,L., Miller,C.S., Smith,A.J., Pettit,F.K., Bowie,J.U. and Eisenberg,D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
- Breitkreutz,B.J., Stark,C., Reguly,T., Boucher,L., Breitkreutz,A., Livstone,M., Oughtred,R., Lackner,D.H., Bahler,J., Wood,V. *et al.* (2008) The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.*, **36**, D637–D640.
- Joshi-Tope,G., Gillespie,M., Vastrik,I., D'Eustachio,P., Schmidt,E., de Bono,B., Jassal,B., Gopinath,G.R., Wu,G.R., Matthews,L. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
- Mitchell,J.A., Aronson,A.R., Mork,J.G., Folk,L.C., Humphrey,S.M. and Ward,J.M. (2003) Gene indexing: characterization and analysis of NLM's GeneRIFs. *AMIA Annu. Symp. Proc.*, **2003**, 460–464.
- Linding,R., Jensen,L.J., Pasculescu,A., Olhovskiy,M., Colwill,K., Bork,P., Yaffe,M.B. and Pawson,T. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.*, **36**, D695–D699.
- Chaurasia,G., Iqbal,Y., Hanig,C., Herzel,H., Wanker,E.E. and Futschik,M.E. (2007) UniHI: an entry gate to the human protein interactome. *Nucleic Acids Res.*, **35**, D590–D594.
- Uetz,P., Dong,Y.A., Zeretzke,C., Atzler,C., Baiker,A., Berger,B., Rajagopala,S.V., Roupelieva,M., Rose,D., Fossum,E. *et al.* (2006) Herpesviral protein networks and their interaction with the human proteome. *Science*, **311**, 239–242.
- Lee,J.H., Vittone,V., Diefenbach,E., Cunningham,A.L. and Diefenbach,R.J. (2008) Identification of structural protein-protein interactions of herpes simplex virus type 1. *Virology*, **378**, 347–354.
- Calderwood,M.A., Venkatesan,K., Xing,L., Chase,M.R., Vazquez,A., Holthaus,A.M., Ewence,A.E., Li,N., Hirozane-Kishikawa,T., Hill,D.E. *et al.* (2007) Epstein-Barr virus and virus human protein interaction maps. *Proc. Natl Acad. Sci. USA*, **104**, 7606–7611.
- von Brunn,A., Teepe,C., Simpson,J.C., Pepperkok,R., Friedel,C.C., Zimmer,R., Roberts,R., Baric,R. and Haas,J. (2007) Analysis of intraviral protein-protein interactions of the SARS coronavirus ORF6. *PLoS ONE*, **2**, e459.
- Camon,E., Barrell,D., Lee,V., Dimmer,E. and Apweiler,R. (2004) The Gene Ontology Annotation (GOA) Database—an integrated resource of GO annotations to the UniProt Knowledgebase. *In Silico Biol.*, **4**, 5–6.
- Kanehisa,M. (2002) The KEGG database. *Novartis Found Symp.*, **247**, 91–101. discussion 101–103, 119–128, 244–152.
- Mulder,N.J. and Apweiler,R. (2008) The InterPro database and tools for protein domain analysis. *Curr. Protoc. Bioinformatics*, **Chap 2**, Unit 2.7.
- Kersey,P.J., Duarte,J., Williams,A., Karavidopoulou,Y., Birney,E. and Apweiler,R. (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics*, **4**, 1985–1988.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
- Kerrien,S., Orchard,S., Montecchi-Palazzi,L., Aranda,B., Quinn,A.F., Vinod,N., Bader,G.D., Xenarios,I., Wojcik,J., Sherman,D. *et al.* (2007) Broadening the horizon—level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.
- Chen,J., Chua,H.N., Hsu,W., Lee,M.L., Ng,S.K., Saito,R., Sung,W.K. and Wong,L. (2006) Increasing confidence of protein-protein interactomes. *Genome Inform.*, **17**, 284–297.
- Chua,H.N. and Wong,L. (2008) Increasing the reliability of protein interactomes. *Drug Discov. Today*, **13**, 652–658.
- Zanzoni,A., Montecchi-Palazzi,L., Quondam,M., Ausiello,G., Helmer-Citterich,M. and Cesareni,G. (2002) MINT: a Molecular INteraction database. *FEBS Lett.*, **513**, 135–140.
- Zheng,Q. and Wang,X.J. (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic Acids Res.*, **36**, 358–363.
- Dyer,M.D., Murali,T.M. and Sobral,B.W. (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog.*, **4**, e32.
- Han,E.K. and McGonigal,T. (2007) Role of focal adhesion kinase in human cancer: a potential target for drug discovery. *Anticancer Agents Med. Chem.*, **7**, 681–684.
- Cooray,S. (2004) The pivotal role of phosphatidylinositol 3-kinase-Akt signal transduction in virus survival. *J. Gen. Virol.*, **85**, 1065–1076.
- Buchkovich,N.J., Yu,Y., Zampieri,C.A. and Alwine,J.C. (2008) The TORrid affairs of viruses: effects of mammalian DNA viruses on the PI3K-Akt-mTOR signalling pathway. *Nat. Rev. Microbiol.*, **6**, 266–275.