Check for updates

Protein–Protein Interactions in Virus–Host Systems

Anderson F. Brito and John W. Pinney*

Department of Life Sciences, Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London, United Kingdom

To study virus–host protein interactions, knowledge about viral and host protein architectures and repertoires, their particular evolutionary mechanisms, and information on relevant sources of biological data is essential. The purpose of this review article is to provide a thorough overview about these aspects. Protein domains are basic units defining protein interactions, and the uniqueness of viral domain repertoires, their mode of evolution, and their roles during viral infection make viruses interesting models of study. Mutations at protein interfaces can reduce or increase their binding affinities by changing protein electrostatics and structural properties. During the course of a viral infection, both pathogen and cellular proteins are constantly competing for binding partners. Endogenous interfaces mediating intraspecific interactions—viral–viral or host–host interactions—are constantly targeted and inhibited by exogenous interfaces mediating viral–host interactions. From a biomedical perspective, blocking such interactions is the main mechanism underlying antiviral therapies. Some proteins are able to bind multiple partners, and their modes of interaction define how fast these "hub proteins" evolve. "Party hubs" have multiple interfaces; they establish simultaneous/stable (domain–domain) interactions, and tend to evolve slowly. On the other hand, "date hubs" have few interfaces; they establish transient/weak (domain–motif) interactions by means of short linear peptides (15 or fewer residues), and can evolve faster. Viral infections are mediated by several protein–protein interactions (PPIs), which can be represented as networks (protein interaction networks, PINs), with proteins being depicted as nodes, and their interactions as edges. It has been suggested that viral proteins tend to establish interactions with more central and highly connected host proteins. In an evolutionary arms race, viral and host proteins are constantly changing their interface residues, either to evade or to optimize their binding capabilities. Apart from gaining and losing interactions via rewiring mechanisms, virus–host PINs also evolve via gene duplication (paralogy); conservation (orthology); horizontal gene transfer (HGT) (xenology); and molecular mimicry (convergence). The last sections of this review focus on PPI experimental approaches and their limitations, and provide an overview of sources of biomolecular data for studying virus–host protein interactions.

Keywords: PPI, virus-host interactions, protein interaction networks, molecular evolution, viral evolution, databases, structural biology, integrative biology

INTRODUCTION

Compared to the relatively well-conserved processes found in cellular organisms, viruses demonstrate huge variations in terms of genomic composition, patterns of evolution, and protein function. While studying protein–protein interactions (PPIs) in virus–host systems, these variations on the pathogen side must be considered. A large proportion of the PPIs are mediated by domain–domain interactions (DDIs), and viruses belonging to different Baltimore groups have specific domain repertoires, providing different strategies and mechanisms of molecular recognition to accomplish their replication cycle (Zheng et al., 2014). In DDIs, molecular recognition is performed via amino acid residues located at interfaces of interaction. Under homeostatic conditions, host proteins interact with each other via (endogenous) interfaces that are also sometimes explored by viruses (exogenous interfaces), leading to competition for such molecular resources between viruses and hosts (Franzosa and Xia, 2011). Protein recognition events can occur as stable or transient interactions, and some proteins can establish interactions with multiple partners, either simultaneously (party hubs), or at different times (date hubs; Han et al., 2004). Such patterns of interaction can be studied in the context of the overall protein interaction network (PIN), in which each node shows particular properties (e.g., connectivity, centrality, etc.; Gursoy et al., 2008).

In terms of evolution, virus and host PPIs often evolve under a regime of arms race. In this phenomenon, one of the partners undergoes mutations that can in turn promote the fixation of mutations in its counterpart, causing both proteins to change over time in a way that retains their mutual recognition capabilities (Daugherty and Malik, 2012). Other common mechanisms of evolution in virus–host systems involve the acquisition of new proteins and interactions via gene duplication, horizontal gene transfer (HGT), and convergence (Alcami, 2003; Koonin et al., 2006; Garamszegi et al., 2013).

Protein interaction data are usually obtained using strategies such as, yeast two-hybrid (Y2H) and affinity-purification mass spectrometry (AP-MS), which present specific advantages and disadvantages (Gavin et al., 2006; Gingras et al., 2007). Different kinds of proteomic data are gathered in multiple independent databases, which provide researchers with information on protein classification, domains, interactions, GO terms, etc.

The purpose of this review is to give an overview on the main concepts required for studying protein interactions in virus–host systems. We also assess the availability of genomic, interaction, and structural data within several databases for all viral families described so far.

PROTEIN DOMAINS IN THE CONTEXT OF VIRUS–HOST INTERACTIONS

Domains are elementary protein structures that evolve independently from each other (Vogel et al., 2004). They have specific biological functions, and most proteins are composed of multiple domains (Apic et al., 2001). Since domains are the basic units by which proteins establish molecular interactions, PPIs can be better understood when they are seen from the level of DDIs (Lee et al., 2006; Yellaboina et al., 2011). Among different viruses, the domain repertoire varies according to the molecular structure of their genomes, which are the basis for their classification into seven viral groups (Baltimore, 1971; **Table 1**).

Viral domains observed among DNA viruses form groups I and II; RNA viruses from groups III, IV, and V; and retro transcribing viruses from groups VI and VII are strictly conserved within these groups, and each viral group uses a unique set of domains to carry out infections (Zheng et al., 2014).
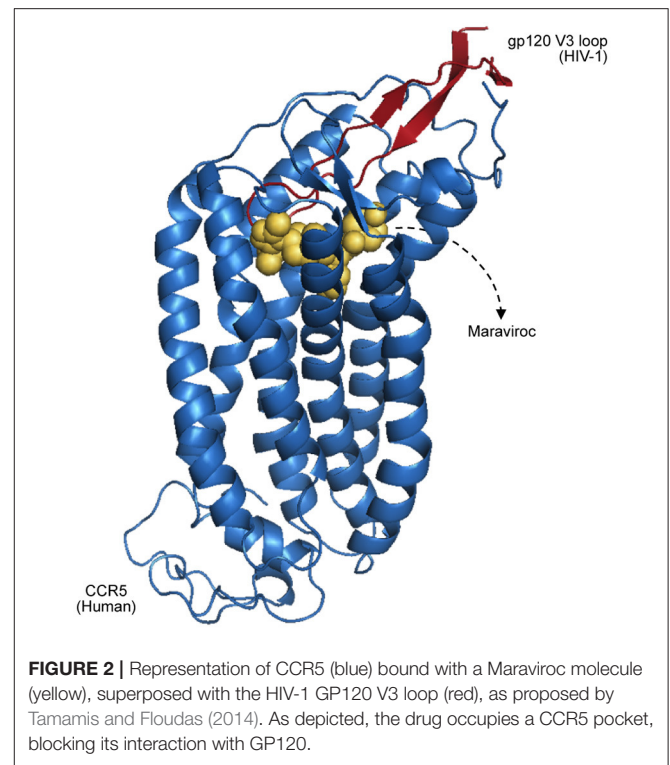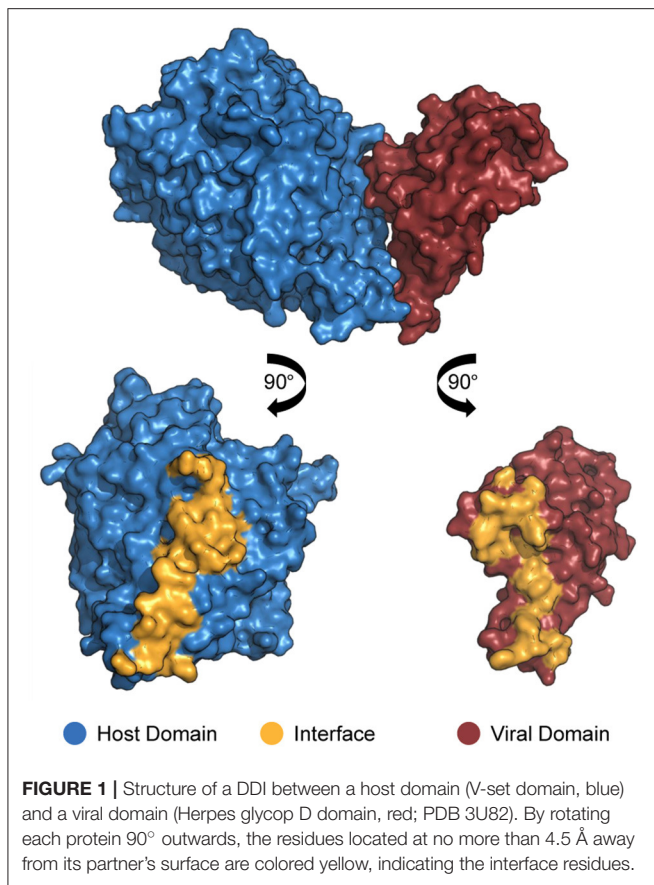
Nearly two-thirds of the viral families are composed by viruses with small genomes of no more than 20,000 nt (see Supplementary Table 1). A direct consequence of this high level of genome compaction is that most viruses encode <30 proteins/domains (see Supplementary Table 1), which in turn are able to interact with multiple host targets and perform a large set of functions (Franzosa and Xia, 2011; Zheng et al., 2014). Another peculiarity of viral domains is their tendency to evolve by convergence, mimicking host interfaces and allowing their proteins to target and compete for host factors usually involved in crucial cellular processes (Franzosa and Xia, 2011; Daugherty and Malik, 2012). At the level of DDIs, RNA viruses tend to be better characterized in terms of the domains encoded and the functions performed. On the contrary, for some viruses with genomes larger than 20 kb (mostly group I DNA viruses, see Supplementary Table 1), information about domains is still limited, and many of their proteins have no assigned domain or known function (Zheng et al., 2014). Unfortunately, such scarcity of information currently limits the conclusions that can be drawn from such markedly incomplete protein data of large DNA viruses.

INTERFACES OF PROTEIN INTERACTIONS IN VIRUS–HOST SYSTEMS

In order for proteins to interact with each other, their respective binding sites must be in direct physical contact, either in a stable or transient mode (Byrum et al., 2012). Such binding sites are called "interfaces": three-dimensional structures formed by sets of amino acid residues directly responsible for the recognition of binding partners (**Figure 1**; Franzosa and Xia, 2011). Deleterious or beneficial mutations occur especially on interfaces, affecting binding affinity due to impairment or improvement of protein

TABLE 1 | Virus classification system (Baltimore groups).

Group	Acronym	Members
I	dsDNA	Double-stranded DNA viruses
II	ssDNA	Single-stranded DNA viruses
III	dsRNA	Double-stranded RNA viruses
IV	ssRNA(+)	Positive-sense single-stranded RNA viruses
V	ssRNA(-)	Negative-sense single-stranded RNA viruses
VI	ssRNA-RT	RNA reverse-transcribing viruses
VII	dsDNA-RT	DNA reverse-transcribing viruses

FIGURE 1 | Structure of a DDI between a host domain (V-set domain, blue) and a viral domain (Herpes glycop D domain, red; PDB 3U82). By rotating each protein 90° outwards, the residues located at no more than 4.5 Å away from its partner's surface are colored yellow, indicating the interface residues.

FIGURE 2 | Representation of CCR5 (blue) bound with a Maraviroc molecule (yellow), superposed with the HIV-1 GP120 V3 loop (red), as proposed by Tamamis and Floudas (2014). As depicted, the drug occupies a CCR5 pocket, blocking its interaction with GP120.

electrostatic and structural properties (Daugherty and Malik, 2012).

In the context of virus–host PPIs, protein-binding sites can be classified as endogenous or exogenous interfaces. Endogenous interfaces are responsible for mediating interactions between proteins belonging to viral or host proteomes, i.e., host–host or virus–virus PPIs. On the other hand, exogenous interfaces mediate interactions between proteins belonging to distinct proteomes, as seen in virus–host PPIs (Franzosa and Xia, 2011).

In virus–host systems, extensive competition for interfaces is common between endogenous and exogenous partners, and viral proteins frequently interfere with host–host protein interactions (Franzosa et al., 2012). Such competition is so frequent that most of the proteins that have at least one known host–host interaction are also involved in virus–host interactions (Franzosa and Xia, 2011).

Having a broader understanding of virus–host PPIs and their interfaces is crucial for the development of new antiviral therapies, like the design of small molecules capable of binding and blocking essential interactions of viral processes (Bailer and Haas, 2009; Gardner et al., 2015). A classic example of virus–host interactions being blocked at the interface level is the action of Maraviroc as an inhibitor of HIV-1 entry to host cells (PDB 4MBS). This drug binds the cellular co-receptor CCR5, preventing it from interacting with GP120 (**Figure 2**),

an essential step of HIV-1 infection (Macarthur and Novak, 2008).

MODES OF PROTEIN INTERACTION

PPIs commonly rely on large interfaces, whilst transient ones involve short linear peptides, as sequence motifs composed of 15 residues or less (Segura-Cabrera et al., 2013). Proteins that show a wide range of binding partners are called "hubs." Hubs having only few interfaces are likely to interact transiently with different partners at different times (date hubs; Han et al., 2004), usually via domain–motif interactions (Franzosa and Xia, 2011). Conversely, proteins showing multiple interfaces tend to establish simultaneous interactions with multiple partners. Such proteins (party hubs) are likely to arrange themselves in complexes, via stable DDIs (Han et al., 2004).

Due to their mode of interaction and number of interface residues, party hubs tend to evolve slowly, as changes in their residues are likely to impair some interactions with specific partners (Fraser et al., 2002). The opposite scenario is observed among proteins that establish transient interactions, which usually evolve faster (Teichmann, 2002). Interestingly, most viral proteins interfering with cell signaling and regulatory pathways perform transient interactions with host proteins (Perkins et al., 2010), leading to severe changes in cellular metabolism (Segura-Cabrera et al., 2013). Unfortunately, compared to stable (domain–domain) interactions, transient (domain–motif) interactions are under-represented in PPI databases, mainly due to limitations associated with the methods used so far to obtain protein–protein interaction data (Russell et al., 2004).

THE EVOLUTION OF PROTEIN INTERFACES AND THE VIRUS–HOST ARMS RACE

In virus–host systems, interacting proteins are constantly losing and regaining their binding sites in order to evade or optimize interspecific PPIs. This process of constant change is known as an "evolutionary arms race" (Franzosa and Xia, 2011; Daugherty and Malik, 2012).

Under an arms race regime, proteins can evolve by offensive or defensive strategies. Host proteins evolve offensively when they are constantly changing as part of an effort to retain or restore their recognition capabilities to bind and neutralize viral factors, which in turn is under recurrent adaptation to evade the host's antagonist actions (Daugherty and Malik, 2012). For example, host immune system proteins in constant interaction with pathogen proteins frequently evolve by an offensive strategy. This scenario is usually found in mammals, whose antiviral proteins are under constant adaptation to recognize their antigens, showing a rapid mode of evolution (Lindblad-Toh et al., 2011). Conversely, defensive strategies are observed when host proteins targeted by viral antagonists undergo mutations to prevent pathogen proteins from binding their interfaces. As a response, this context can favor the fixation of novel mutations on viral interfaces, probably compensating for host evasion (Daugherty and Malik, 2012).

In this intricate virus–host arms race, endogenous and exogenous interfaces show different patterns of evolution. Host interfaces mediating host–host PPIs tend to be less variable than interfaces directly targeted by viral proteins (Franzosa and Xia, 2011). These proteins contain specific residues where small changes can drastically modify protein function and/or structure, and consequently their intraspecific binding affinity (Daugherty and Malik, 2012). This mode of evolution is especially observed in co-evolving host–host interfaces, where mutations can be potentially deleterious, and strong purifying selection acts to maintain the integrity of their binding sites (Franzosa and Xia, 2011; Daugherty and Malik, 2012). However, there are exceptions to this pattern. Taking into account that some endogenous binding sites overlap with exogenous interfaces (Franzosa and Xia, 2011), shared residues of endogenous interfaces can evolve faster due to competition with a viral antagonist (Elde and Malik, 2009).

PROTEIN INTERACTION NETWORKS UNDERLYING VIRAL PROCESSES

In order for pathogens to take over the cellular machinery and replicate themselves, molecular interactions must be established with their hosts. Such interactions are commonly represented using networks, where nodes represent proteins and edges connecting them denote direct physical interactions (Dyer et al., 2008; Bailer and Haas, 2009).

When a PIN is reconstructed, several properties of each protein can be calculated from its network topology, such as, connectivity (degree) and centrality (Gursoy et al., 2008).

Some of these properties have been suggested to be biologically informative, although these findings are not strongly supported by data available so far (Mason and Verwoerd, 2007; Ratmann et al., 2009). As an example, it is well known that highly connected proteins (hubs) tend to interact with low-degree (non-hub) proteins instead of establishing interactions with other hubs (Maslov and Sneppen, 2002). Hub proteins are not always functionally essential, however, their high level of connectivity (degree) could evidence their involvement in multiple biological processes, in such a way that removing them from PINs can probably lead to negative pleiotropic effects (Ratmann et al., 2009).

In terms of virus–host interactions, it has been suggested that viral proteins tend to target more central and highly connected host proteins (Dyer et al., 2011; Zheng et al., 2014). Nonetheless, due to the overrepresentation of extensively studied proteins in PPI databases, assumptions drawn based on network properties could be mere sampling bias, and their meaning must be interpreted with caution (Ratmann et al., 2009; Dickerson et al., 2010). Another downside of host–pathogen PIN analysis is the incompleteness of their interactomes. The scarcity of interaction data in these systems has being a major analysis bottleneck, with most studies being either focused on extensively studied host–pathogen systems, or relying on transferring information by homology from such model systems to neglected ones (Ammari et al., 2016).

HOW DO PROTEIN INTERACTION NETWORKS EVOLVE?

Each viral family encodes a set of protein domains that are classified into several domain families based on their evolutionary relationships (Chothia et al., 2003). Large dsDNA viruses show the most variable protein domain repertoires, while most RNA and retrotranscribing viruses, due to their genome sizes, perform all their processes using few domains, which are reused throughout their entire infection cycles (Zheng et al., 2014). Owing to these proteomic peculiarities, viruses diversify and/or maintain their interaction capabilities via different mechanisms of molecular evolution, including conservation (orthology); HGT (xenology); gene duplication (paralogy); and molecular mimicry (convergence; Alcami, 2003; Koonin et al., 2006; Garamszegi et al., 2013).

In the mode of evolution by orthology, if a domain pair known to interact is found in two closely related systems (organisms), these domains are likely to be real interacting partners in both systems (interologs). Although the mere presence of orthologs in two systems does not necessarily imply a direct interaction among them (Riley et al., 2005), their co-occurrence could be indicative of potentially conserved interactions, making these proteins good targets for further studies (Dyer et al., 2011). Examples of interologs are observed among herpesviruses, DNA viruses that show a large set of genes and interactions shared among almost all members of the family Herpesviridae (Bailer and Haas, 2009). An example of such conservation in observed in the interaction between the cellular receptor Nectin-1 and the
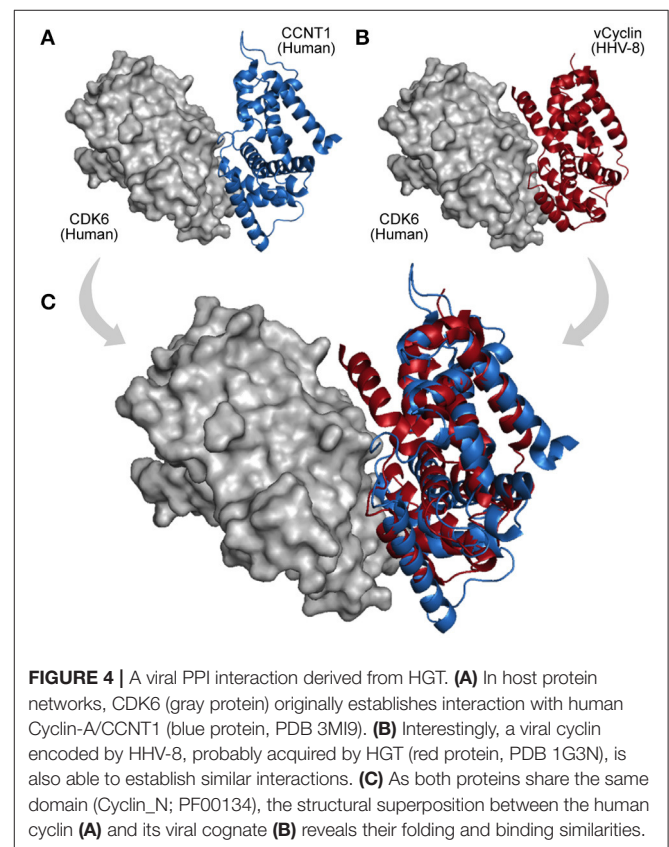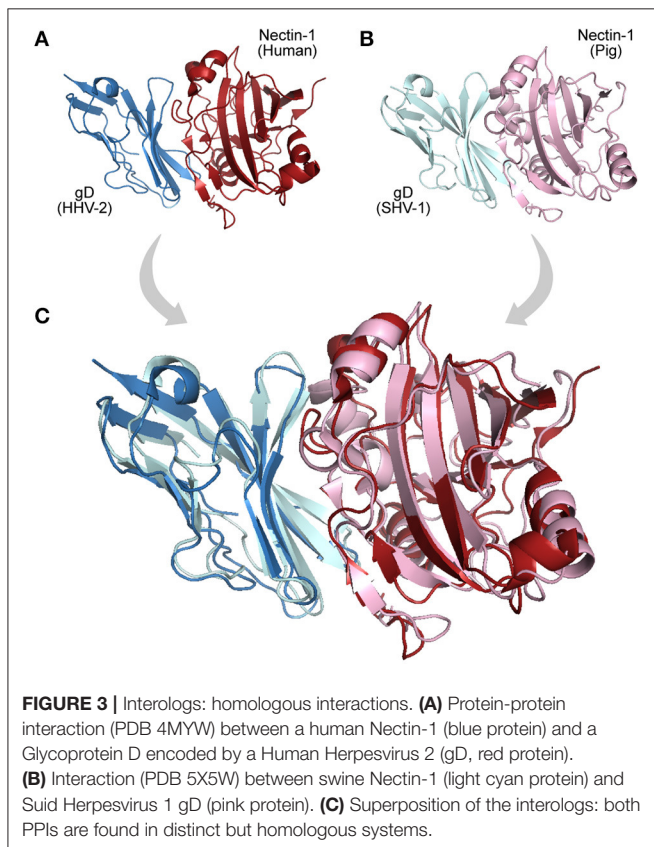
envelope glycoprotein D encoded by alphaherpesviruses. Nectins are commonly found at adherens junction (GO:0005912), and are used by viruses as cell entry mediators (GO:0046718). **Figure 3** illustrates such an interaction in two host–virus pairs: "human × HHV-2," and "pig × SHV-1."

HGT is a process of genome recombination by means of which some viruses acquire one or more genes from non-parental organisms, a mechanism of evolution especially observed among large DNA viruses, which usually acquire new genes from other viruses, bacteria, or from their hosts (Shackelton and Holmes, 2004). Once a viral genome has incorporated a new gene, the protein product can be optimized and integrated into its virus–host network (Daugherty and Malik, 2012). Large dsDNA viruses, such as, poxviruses and herpesviruses, have been shown to be remarkably prone to acquire and domesticate exogenous genes within several functional categories (Raftery et al., 2000; Hughes and Friedman, 2005). **Figure 4B** shows an interaction between a human CDK6 and a Cyclin encoded by the *Human Herpesvirus 8*. This interaction is part of an immune system process (GO:0006955), and takes place in the extracellular region (GO:0005576). The viral cyclin (vCyclin) was probably acquired by HGT, and is capable of modulating cellular growth (GO:0005125) in similar ways to cellular cyclins D (**Figure 4**; Godden-Kent et al., 1997).

Gene duplication (paralogy) is another usual mechanism of protein network evolution, and duplicated genes are common in some viral genomes. After a duplication event, each paralog can undergo independent mutations, giving rise to new biological functions (Barabasi and Oltvai, 2004; Ratmann et al., 2009). Similarly to HGT events, gene duplications are evolutionary processes mainly found among dsDNA viruses, as observed in herpesviruses, adenoviruses and poxviruses (Shackelton and Holmes, 2004). An example of evolution by gene duplication is observed for the herpesvirus Glycoprotein D previously shown in **Figure 1**. Some alphaherpesviruses express a second copy of that protein, glycoprotein G (gG), a paralog that does not act as a viral entry mediators, but in fact shows a modified function, binding a broad range of chemokines to prevent their interaction with specific cellular receptors (Bryant et al., 2003). Interestingly, the presence of paralogs in PINs is not exclusive to viruses.

Finally, acquisition of a new interaction partner via convergent evolution is also a recurrent mechanism in virus–host networks. As they evolve at faster mutation rates, viruses can rapidly acquire new binding partners by mimicking and targeting interfaces of host proteins (Elde and Malik, 2009; Standfuss, 2015). A particular example is observed among *Dengue* viruses, *Vaccinia* viruses, and HIV-1, which independently acquired similar mechanisms of protein interaction and RNA recognition, which are essential to promote genome replication and mRNA translation (Garcia-Montalvo et al., 2004; Alvarez et al., 2006; Katsafanas and Moss, 2007). In this way, viruses can evolve not only by homology (HGT, duplication, and conservation) but also by analogy, allowing them to share interacting partners and adopt common strategies of infection (Dyer et al., 2008; Bailer

FIGURE 3 | Interologs: homologous interactions. **(A)** Protein-protein interaction (PDB 4MYW) between a human Nectin-1 (blue protein) and a Glycoprotein D encoded by a Human Herpesvirus 2 (gD, red protein). **(B)** Interaction (PDB 5X5W) between swine Nectin-1 (light cyan protein) and Suid Herpesvirus 1 gD (pink protein). **(C)** Superposition of the interologs: both PPIs are found in distinct but homologous systems.

FIGURE 4 | A viral PPI interaction derived from HGT. **(A)** In host protein networks, CDK6 (gray protein) originally establishes interaction with human Cyclin-A/CCNT1 (blue protein, PDB 3MI9). **(B)** Interestingly, a viral cyclin encoded by HHV-8, probably acquired by HGT (red protein, PDB 1G3N), is also able to establish similar interactions. **(C)** As both proteins share the same domain (Cyclin_N; PF00134), the structural superposition between the human cyclin **(A)** and its viral cognate **(B)** reveals their folding and binding similarities.

and Haas, 2009; Segura-Cabrera et al., 2013). **Figure 5** shows an example of convergent evolution. The human Ephrin-B2 is a cell surface transmembrane ligand of Ephrin receptors (**Figure 5A**; Qin et al., 2010), and the Glycoprotein G encoded by the Paramixovirus *Hendra henipavirus* is an envelope component (GO:0019031) that mimics this interaction with Ephrin receptors using an interface similar to the one explored by Ephrin type-A receptors 4 (**Figure 5B**).

PROTEIN INTERACTION DATA: EXPERIMENTAL APPROACHES AND LIMITATIONS

Among the experimental techniques applied to identify virus–host protein interactions, Y2H and AP-MS are the most extensively used, together contributing to more than 90% of the information available in public databases (Guirimand et al., 2014), with the remaining data being obtained by GST-pull-down, luminescence, protease assay, surface plasmon resonance (SPR), and other techniques.

Y2H is efficient at detecting weak/transient domain–motif interactions (Gavin et al., 2006), however, one drawback is that it does not provide precise information about the domains involved in the interactions (Riley et al., 2005; Lee et al., 2006; Segura-Cabrera et al., 2013). Another disadvantage of Y2H screens is that PPI detection takes place within the nucleus. As some proteins

FIGURE 5 | Interaction convergence. **(A)** In physiological conditions the cell surface Ephrin (gray protein) binds its Ephrin type-A receptor 4 (blue protein, PDB 3GXU). **(B)** However, during infections of the Paramixovirus *Hendra henipavirus*, the Ephrin interface is also used by the viral Glycoprotein G, which by convergence evolved its binding capacity (red protein, PDB 2VSK). **(C)** As shown in the superposition, Ephrin type-A receptor 4 and the viral Glycoprotein G can compete for the same interface on the Ephrin surface.

are not naturally found in this cellular compartment, they are usually not identified as interactors, resulting in a bias that increases the proportion of false negatives (Von Mering et al., 2002). Additionally, a large number of entries in some databases describe binary interactions that are not physiologically feasible, i.e., even though two proteins are biochemically able to interact, if they do not share the same temporal and spatial compartment in a given biological process, their physical contact would not happen in natural conditions (Russell and Aloy, 2008).

AP-MS works in a different way. In these assays, proteins of interest (baits) are tagged with a recombinant fusion tag, which is then used to purify baits and their respective interacting partners (preys). Once purified, each component in the protein complex can be determined by mass spectrometry (Von Mering et al., 2002; Gingras et al., 2007). As a consequence, this technique is more efficient at identifying stable interactions among proteins of the same functional category (Gavin et al., 2006; Chiang et al., 2007). On the other hand, there are some disadvantages associated with affinity purification assays. An intrinsic issue is the use of tags by itself. By tagging the target proteins, their folding can be affected, preventing their normal functioning and raising the level of false positives (Bauer and Kuster, 2003; Meyer and Selbach, 2015). Another downside of this method is that reactions occur in ectopic environments, in other words, outside the normal physiological context of the PPIs. Additionally, for detecting interactions, affinity purification assays require proteins to be overexpressed, which can give rise to artifactual results (Bauer and Kuster, 2003).

To tackle these problems, methods based on Förster Resonance Energy Transfer (FRET) have increasingly gaining popularity. FRET-Based techniques allow the expression of each respective target protein in their native environment, as a single molecule fused with a fluorescent donor/acceptor. Such methods rely on a physical phenomenon of energy transferring between two fluorophores (in this context, interacting proteins): one protein receives light at a specific wavelength and transfers the energy to its interactor, which in turn emits light of other wavelength/color, which is captured by sensors (Xing et al., 2016). Although still an expensive alternative, FRET-based methods could become a suitable solution for detecting dynamic PPI in viral infections (Pfleger and Eidne, 2006; Xing et al., 2016).

So far, the main problem associated with PPI data is the poor overlap among datasets, as large proportions of the interactions are not shared among different experimental screens (Ratmann et al., 2009). This implies that although current PPI assays are high-throughput, their levels of completeness are low, leading to incomplete coverage of the interaction space (Von Mering et al., 2002). This is especially observed in Y2H assays, for which sensitivity (true positive rate) is low, ranging from 20 to 40%, implying that most pairwise interactions are not identified, falling in the false negative space (Bailer and Haas, 2009; Braun et al., 2009).

Another problem is that many databases are constructed by integrating data derived from several low throughput studies, which leads to some proteins being overrepresented in PINs (Bailer and Haas, 2009). It occurs mainly due to the bias associated with highly studied genes, such as, *p53* and other
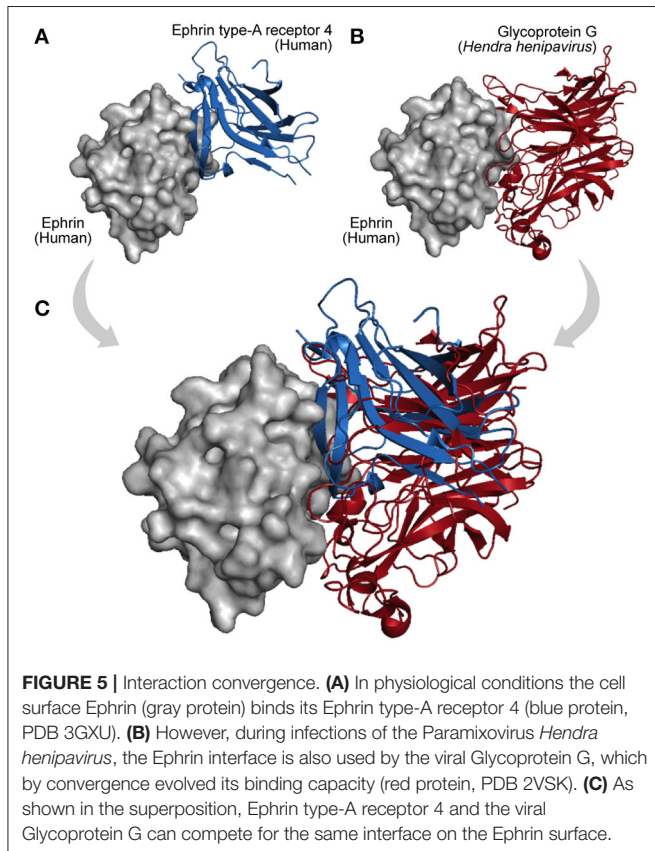
genes often related to human diseases (Sinen and Koyutürk, 2010). Proteins encoded by these genes usually emerge as highly connected nodes in the PIN, which may lead to erroneous conclusions, especially when it comes to applying the available data as input for predictive models (Riley et al., 2005; Bailer and Haas, 2009).

However, although high confidence data are still scarce, the virus–host PPI data available so far allow us to conduct large-scale comparative studies to understand the fundamental cellular functions targeted by viruses from different viral families, as well as the evolution of virus–host PINs (Navratil et al., 2009).

SOURCES OF BIOMOLECULAR DATA FOR VIRUS–HOST INTERACTION STUDIES

Integrative approaches are essential for the deep understanding of the evolutionary aspects of virus–host PINs. The data types applied for this purpose comes from different biological dimensions: protein sequence, domain composition, PPIs, DDIs,

gene ontology terms, and taxonomic data, and several databases providing such biological information are available (see **Table 2**).

SCOP (Andreeva et al., 2014), Superfamily (Gough and Chothia, 2002), and Pfam (Finn et al., 2014a) provide valuable information on protein and domain classifications, such as, domain organizations, functional annotations, and taxonomic distributions of proteins encoded by completely sequenced genomes. To explore genetic diversity, databases such as (UniProt Consortium, 2014) and Genbank/NCBI (Benson et al., 2014) offer a variety of genome and protein sequences from viruses and their hosts. Protein structure data, such as, protein interfaces and chemical properties, can be found on PQS/PISA (Henrick and Thornton, 1998) and especially on PDB (Berman et al., 2000). Supplementary Table 2 includes a list of 1,100 virus–host PPIs with solved PDB structures, which are classified by viral/host taxonomy, GO terms, protein domains, among other biological information.

Viral and host PPI data can be retrieved from multiple databases, but VirHostNet (Guirimand et al., 2014) and

TABLE 2 | Sources of biomolecular data for studies on virus-host interactions.

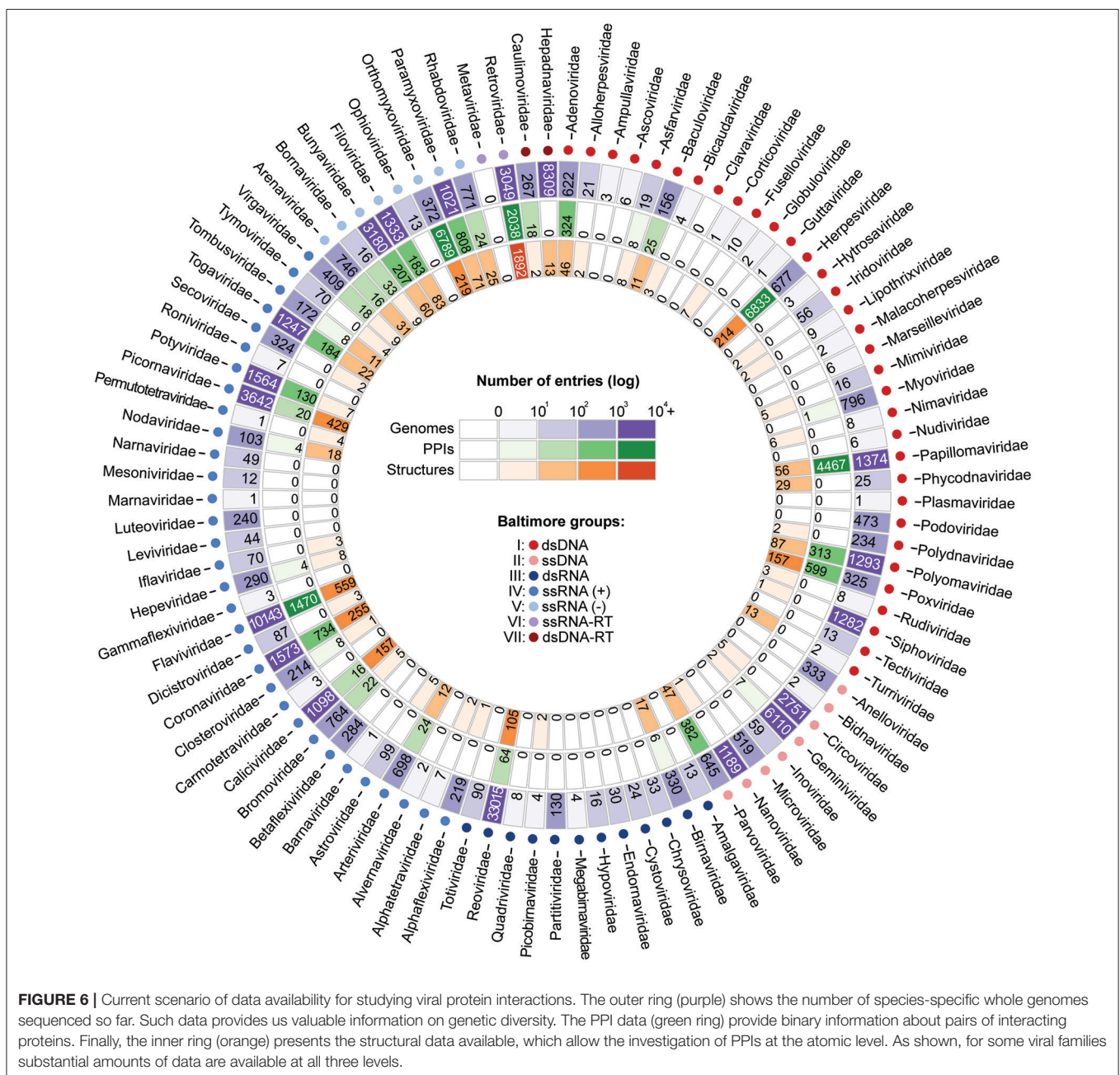| Data types[†] & Databases	Protein classification	Protein sequence	Protein structure	Protein Interaction	Domain interaction	Genomic sequence	GO terms	PSICQUIC member?	Data source*	References
3DID			○	●	●		○	No	[I]	Mosca et al., 2014
BIND	○			●				Yes	[E]	Bader et al., 2003
BIOGRID				●				Yes	[E] [T]	Chatr-Aryamontri et al., 2015
DIP	○	○		●			○	Yes	[E]	Xenarios et al., 2002
GENERIF				●				No	[E] [T]	Mitchell et al., 2003
HIV-1 Human Protein Interaction		○		●		●		No	[T]	Ptak et al., 2008
HPIDb				●				Yes	[I] [P]	Kumar and Nanduri, 2010
HPRD	○	○	○	●			○	Yes	[E]	Prasad et al., 2009
InnateDB				●			○	Yes	[E] [T]	Lynn et al., 2008
IntAct	●	○		●			○	Yes	[E] [T]	Aranda et al., 2010
InterPro	●	○	○	○			○	–	[E] [I]	Mitchell et al., 2015
iPfam	●		○	●	●			No	[I]	Finn et al., 2014b
MatrixDB	○	○	○	●			○	Yes	[E] [P]	Launay et al., 2015
MINT	○	○	○	●			○	Yes	[E] [T]	Licata et al., 2012
MPPI				●				No	[T]	Pagel et al., 2005
Negatome				●	●			No	[P] [T]	Blohm et al., 2014
NetworKIN				●				No	[P]	Linding et al., 2008
Pfam	●	●	○	●	●		○	No	[I] [P]	Finn et al., 2014a
PDB	○	●	●	●	●		○	No	[E]	Berman et al., 2000
PQS/PISA			●	●	●			No	[A] [I]	Henrick and Thornton, 1998
Reactome	●			●			○	Yes	[I] [T]	Joshi-Tope et al., 2005
SCOP	●	●	○					–	[A] [I]	Andreeva et al., 2014
Superfamily	●	○				○	○	–	[A] [P]	Gough and Chothia, 2002
UniProt	○	●	○	○		●	○	Yes	[I]	UniProt Consortium, 2014
Viral Genomes NCBI		○				●		–	[I]	Brister et al., 2015
VirHostNet				●				Yes	[I] [T]	Guirimand et al., 2014
VirusMentha				●				No	[I] [T]	Calderone et al., 2015

*A, Annotation; E, Experimental data; I, Integration of multiple databases; P, Prediction; T, Text mining. [†] ●, primary data type; ○, data retrieved from links to secondary sources.

VirusMentha (Calderone et al., 2015) are databases entirely dedicated to virus–host interactions. Most of their entries come from external and semi-independent sources, such as, MINT (Licata et al., 2012), IntAct (Aranda et al., 2010), DIP (Xenarios et al., 2002), and BIOGRID (Chatr-Aryamontri et al., 2015), which predominantly store experimental and literature-derived data. Hence, these datasets show slightly different sets of binary interactions, which cover different parts of the interaction space (Ratmann et al., 2009). To circumvent these problems, an integrative platform was proposed as a solution to integrate all entries, building a single, comprehensive, and non-redundant database. The Proteomics Standard Initiative developed the Common Query Interface (PSICQUIC), a

platform to retrieve molecular interaction data from multiple databases storing binary interactions in PSI-MI format (Aranda et al., 2011).

Finally, most of the entries of the aforementioned databases are usually associated with external gene ontology (GO) information (Ashburner et al., 2000), such as, biological processes, molecular functions, and cellular compartments where interactions are likely to occur. Altogether, functional, evolutionary and interaction data make it possible to integrate all knowledge accumulated so far, and construct predictive models for virus–host interactions.

For those interested in studying the evolution of viral PPIs by means of integrative approaches, three types of

FIGURE 6 | Current scenario of data availability for studying viral protein interactions. The outer ring (purple) shows the number of species-specific whole genomes sequenced so far. Such data provides us valuable information on genetic diversity. The PPI data (green ring) provide binary information about pairs of interacting proteins. Finally, the inner ring (orange) presents the structural data available, which allow the investigation of PPIs at the atomic level. As shown, for some viral families substantial amounts of data are available at all three levels.

information are of particular interest: sequence diversity, binary protein interactions, and protein structures. By collecting viral biomolecular data from NCBI (Viral Genomes), VirHostNet (PPIs), and PDB (Structures) it is possible to assess the level of data availability for several viral families. As shown in **Figure 6**, most families still have only a limited amount of data, far from faithfully representing the true complexity of a virus protein interaction network (Dyer et al., 2011). However, the amount of data available for at least 15 families has shown to be favorable for evolutionary studies on viral PPIs. Among the viral families with consistent data availability there are five dsDNA families (Adenoviridae, Herpesviridae, Papillomaviridae, Polyomaviridae, Poxviridae); one ssDNA (Parvoviridae); one dsRNA (Reoviridae); three ssRNA(+) (Coronaviridae, Flaviviridae, Picornaviridae); four ssRNA(-) (Bunyaviridae, Filoviridae, Orthomyxoviridae, Paramyxoviridae), and one ssRNA-RT (Retroviridae; **Figure 6**). As previously mentioned, these distinct viral groups evolve under particular genetic mechanisms, and are good models for understanding the evolution of PPIS at the molecular and network levels.

CONCLUSION

Viruses are pathogens with rather compact genomes that nevertheless provide them with versatile molecular tools able to cause extensive changes in cellular processes. Such versatility can be credited to domain repertoires encoded by viruses, as well as to their mechanisms of molecular evolution. The aspects addressed in this review provide starting points not just to virologists willing to explore integrative approaches to understand viral evolution, but also to computational biologists wanting to understand more about the peculiarities of viral biology in order to develop predictive models for virus–host PPIs.

Advances in biomolecular research over the last decades now allow us to tackle important questions regarding virus–host interactions by integrating data from multiple levels of biological complexity. As shown in **Figure 6**, for at least 15 viral families, large amounts of information on sequence diversity, protein interactions, and structures are available, allowing us to better understand how viruses evolve their mechanisms of interaction alongside their hosts. Additional studies on such evolutionary aspects could help us to develop new strategies for PPI inhibition and provide us with extra knowledge to explain cases of host switch, as well as the expansion of pathogen virulence and host range in emerging viral diseases.

AUTHOR CONTRIBUTIONS

AB and JP wrote the manuscript. AB collected the data and created the figures. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01557/full#supplementary-material

REFERENCES

Alcami, A. (2003). Viral mimicry of cytokines, chemokines and their receptors. *Nat. Rev. Immunol.* 3, 36–50. doi: 10.1038/nri980

Alvarez, E., Castello, A., Menendez-Arias, L., and Carrasco, L. (2006). HIV protease cleaves poly(A)-binding protein. *Biochem. J.* 396, 219–226. doi: 10.1042/BJ20060108

Ammari, M. G., Gresham, C. R., McCarthy, F. M., and Nanduri, B. (2016). HPIDB 2.0: a curated database for host-pathogen interactions. *Database (Oxford)* 2016:baw103. doi: 10.1093/database/baw103

Andreeva, A., Howorth, D., Chothia, C., Kulesha, E., and Murzin, A. G. (2014). SCOP$_2$ prototype: a new approach to protein structure mining. *Nucleic Acids Res.* 42, D310–D314. doi: 10.1093/nar/gkt1242

Apic, G., Gough, J., and Teichmann, S. A. (2001). An insight into domain combinations. *Bioinformatics* 17(Suppl. 1), S83–S89. doi: 10.1093/bioinformatics/17.suppl_1.S83

Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., Bridge, A., Derow, C., et al. (2010). The intact molecular interaction database in 2010. *Nucleic Acids Res.* 38, D525–D531. doi: 10.1093/nar/gkp878

Aranda, B., Blankenburg, H., Kerrien, S., Brinkman, F. S., Ceol, A., Chautard, E., et al. (2011). PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods* 8, 528–529. doi: 10.1038/nmeth.1637

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bader, G. D., Betel, D., and Hogue, C. W. (2003). BIND: the biomolecular interaction network database. *Nucleic Acids Res.* 31, 248–250. doi: 10.1093/nar/gkg056

Bailer, S. M., and Haas, J. (2009). Connecting viral with cellular interactomes. *Curr. Opin. Microbiol.* 12, 453–459. doi: 10.1016/j.mib.2009.06.004

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* 35, 235–241.

Barabasi, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Bauer, A., and Kuster, B. (2003). Affinity purification-mass spectrometry. Powerful tools for the characterization of protein complexes. *Eur. J. Biochem.* 270, 570–578. doi: 10.1046/j.1432-1033.2003.03428.x

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2014). GenBank. *Nucleic Acids Res.* 42, D32–D37. doi: 10.1093/nar/gkt1030

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235

Blohm, P., Frishman, G., Smialowski, P., Goebels, F., Wachinger, B., Ruepp, A., et al. (2014). Negatome 2.0: a database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. *Nucleic Acids Res.* 42, D396–D400. doi: 10.1093/nar/gkt1079

Braun, P., Tasan, M., Dreze, M., Barrios-Rodiles, M., Lemmens, I., Yu, H., et al. (2009). An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* 6, 91–97. doi: 10.1038/nmeth.1281

Brister, J. R., Ako-Adjei, D., Bao, Y., and Blinkova, O. (2015). NCBI viral genomes resource. *Nucleic Acids Res.* 43, D571–D577. doi: 10.1093/nar/gku1207

Bryant, N. A., Davis-Poynter, N., Vanderplasschen, A., and Alcami, A. (2003). Glycoprotein G isoforms from some alphaherpesviruses function as broad-spectrum chemokine binding proteins. *EMBO J.* 22, 833–846. doi: 10.1093/emboj/cdg092

Byrum, S., Smart, S. K., Larson, S., and Tackett, A. J. (2012). Analysis of stable and transient protein-protein interactions. *Methods Mol. Biol.* 833, 143–152. doi: 10.1007/978-1-61779-477-3_10

Calderone, A., Licata, L., and Cesareni, G. (2015). VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Res.* 43, D588–D592. doi: 10.1093/nar/gku830

Chatr-Aryamontri, A., Breitkreutz, B. J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., et al. (2015). The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* 43, D470–D478. doi: 10.1093/nar/gku1204

Chiang, T., Scholtens, D., Sarkar, D., Gentleman, R., and Huber, W. (2007). Coverage and error models of protein-protein interaction data by directed graph analysis. *Genome Biol.* 8:R186. doi: 10.1186/gb-2007-8-9-r186

Chothia, C., Gough, J., Vogel, C., and Teichmann, S. A. (2003). Evolution of the protein repertoire. *Science* 300, 1701–1703. doi: 10.1126/science.1085371

Daugherty, M. D., and Malik, H. S. (2012). Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.* 46, 677–700. doi: 10.1146/annurev-genet-110711-155522

Dickerson, J. E., Pinney, J. W., and Robertson, D. L. (2010). The biological context of HIV-1 host interactions reveals subtle insights into a system hijack. *BMC Syst. Biol.* 4:80. doi: 10.1186/1752-0509-4-80

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. doi: 10.1371/journal.ppat.0040032

Dyer, M. D., Murali, T. M., and Sobral, B. W. (2011). Supervised learning and prediction of physical interactions between human and HIV proteins. *Infect. Genet. Evol.* 11, 917–923. doi: 10.1016/j.meegid.2011.02.022

Elde, N. C., and Malik, H. S. (2009). The evolutionary conundrum of pathogen mimicry. *Nat. Rev. Microbiol.* 7, 787–797. doi: 10.1038/nrmicro2222

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014a). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Finn, R. D., Miller, B. L., Clements, J., and Bateman, A. (2014b). iPfam: a database of protein family and domain interactions found in the protein data bank. *Nucleic Acids Res.* 42, D364–D373. doi: 10.1093/nar/gkt1210

Franzosa, E. A., Garamszegi, S., and Xia, Y. (2012). Toward a three-dimensional view of protein networks between species. *Front. Microbiol.* 3:428. doi: 10.3389/fmicb.2012.00428

Franzosa, E. A., and Xia, Y. (2011). Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10538–10543. doi: 10.1073/pnas.1101440108

Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C., and Feldman, M. W. (2002). Evolutionary rate in the protein interaction network. *Science* 296, 750–752. doi: 10.1126/science.1068696

Garamszegi, S., Franzosa, E. A., and Xia, Y. (2013). Signatures of pleiotropy, economy and convergent evolution in a domain-resolved map of human-virus protein-protein interaction networks. *PLoS Pathog.* 9:e1003778. doi: 10.1371/journal.ppat.1003778

Garcia-Montalvo, B. M., Medina, F., and Del Angel, R. M. (2004). La protein binds to NS5 and NS3 and to the 5′ and 3′ ends of Dengue 4 virus RNA. *Virus Res.* 102, 141–150. doi: 10.1016/j.virusres.2004.01.024

Gardner, M. R., Kattenhorn, L. M., Kondur, H. R., Von Schaewen, M., Dorfman, T., Chiang, J. J., et al. (2015). AAV-expressed eCD4-Ig provides durable protection from multiple SHIV challenges. *Nature* 519, 87–91. doi: 10.1038/nature14264

Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440, 631–636. doi: 10.1038/nature04532

Gingras, A. C., Gstaiger, M., Raught, B., and Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* 8, 645–654. doi: 10.1038/nrm2208

Godden-Kent, D., Talbot, S. J., Boshoff, C., Chang, Y., Moore, P., Weiss, R. A., et al. (1997). The cyclin encoded by Kaposi's sarcoma-associated herpesvirus stimulates cdk6 to phosphorylate the retinoblastoma protein and histone H1. *J. Virol.* 71, 4193–4198.

Gough, J., and Chothia, C. (2002). SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. *Nucleic Acids Res.* 30, 268–272. doi: 10.1093/nar/30.1.268

Guirimand, T., Delmotte, S., and Navratil, V. (2014). VirHostNet 2.0: surfing on the web of virus/host molecular interactions data. *Nucleic Acids Res.* 43, D583–D587. doi: 10.1093/nar/gku1121

Gursoy, A., Keskin, O., and Nussinov, R. (2008). Topological properties of protein interaction networks from a structural perspective. *Biochem. Soc. Trans.* 36, 1398–1403. doi: 10.1042/BST0361398

Han, J. D., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., et al. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430, 88–93. doi: 10.1038/nature02555

Henrick, K., and Thornton, J. M. (1998). PQS: a protein quaternary structure file server. *Trends Biochem. Sci.* 23, 358–361. doi: 10.1016/S0968-0004(98)01253-5

Hughes, A. L., and Friedman, R. (2005). Poxvirus genome evolution by gene gain and loss. *Mol. Phylogenet. Evol.* 35, 186–195. doi: 10.1016/j.ympev.2004.12.008

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'eustachio, P., Schmidt, E., De Bono, B., et al. (2005). Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* 33, D428–D432. doi: 10.1093/nar/gki072

Katsafanas, G. C., and Moss, B. (2007). Colocalization of transcription and translation within cytoplasmic poxvirus factories coordinates viral expression and subjugates host functions. *Cell Host Microbe* 2, 221–228. doi: 10.1016/j.chom.2007.08.005

Koonin, E. V., Senkevich, T. G., and Dolja, V. V. (2006). The ancient virus world and evolution of cells. *Biol. Direct* 1:29. doi: 10.1186/1745-6150-1-29

Kumar, R., and Nanduri, B. (2010). HPIDB–a unified resource for host-pathogen interactions. *BMC Bioinformatics* 11(Suppl. 6):S16. doi: 10.1186/1471-2105-11-S6-S16

Launay, G., Salza, R., Multedo, D., Thierry-Mieg, N., and Ricard-Blum, S. (2015). MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. *Nucleic Acids Res.* 43, D321–D327. doi: 10.1093/nar/gku1091

Lee, H., Deng, M., Sun, F., and Chen, T. (2006). An integrated approach to the prediction of domain-domain interactions. *BMC Bioinformatics* 7:269. doi: 10.1186/1471-2105-7-269

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., et al. (2012). MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 40, D857–D861. doi: 10.1093/nar/gkr930

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476–482. doi: 10.1038/nature10530

Linding, R., Jensen, L. J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., et al. (2008). NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* 36, D695–D699. doi: 10.1093/nar/gkm902

Lynn, D. J., Winsor, G. L., Chan, C., Richard, N., Laird, M. R., Barsky, A., et al. (2008). InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol. Syst. Biol.* 4, 218. doi: 10.1038/msb.2008.55

Macarthur, R. D., and Novak, R. M. (2008). Maraviroc: the first of a new class of antiretroviral agents. *Clin. Infect. Dis.* 47, 236–241. doi: 10.1086/589289

Maslov, S., and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science* 296, 910–913. doi: 10.1126/science.1065103

Mason, O., and Verwoerd, M. (2007). Graph theory and networks in Biology. *IET Syst. Biol.* 1, 89–119. doi: 10.1049/iet-syb:20060038

Meyer, K., and Selbach, M. (2015). Quantitative affinity purification mass spectrometry: a versatile technology to study protein-protein interactions. *Front. Genet.* 6:237. doi: 10.3389/fgene.2015.00237

Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., et al. (2015). The interpro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 43, D213–D221. doi: 10.1093/nar/gku1243

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.

Mosca, R., Ceol, A., Stein, A., Olivella, R., and Aloy, P. (2014). 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 42, D374–D379. doi: 10.1093/nar/gkt887

Navratil, V., De Chassey, B., Meyniel, L., Delmotte, S., Gautier, C., Andre, P., et al. (2009). VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Res.* 37, D661–D668. doi: 10.1093/nar/gkn794

Pagel, P., Kovac, S., Oesterheld, M., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., et al. (2005). The MIPS mammalian protein-protein interaction database. *Bioinformatics* 21, 832–834. doi: 10.1093/bioinformatics/bti115

Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G., and Orengo, C. (2010). Transient protein-protein interactions: structural, functional, and network properties. *Structure* 18, 1233–1243. doi: 10.1016/j.str.2010.08.007

Pfleger, K. D. G., and Eidne, K. A. (2006). Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat. Methods* 3, 165–174. doi: 10.1038/nmeth841

Prasad, T. S., Kandasamy, K., and Pandey, A. (2009). Human protein reference database and human proteinpedia as discovery tools for systems biology. *Methods Mol. Biol.* 577, 67–79. doi: 10.1007/978-1-60761-232-2_6

Ptak, R. G., Fu, W., Sanders-Beer, B. E., Dickerson, J. E., Pinney, J. W., Robertson, D. L., et al. (2008). Cataloguing the HIV type 1 human protein interaction network. *AIDS Res. Hum. Retroviruses* 24, 1497–1502. doi: 10.1089/aid.2008.0113

Qin, H., Noberini, R., Huan, X., Shi, J., Pasquale, E. B., and Song, J. (2010). Structural characterization of the EphA4-Ephrin-B2 complex reveals new features enabling Eph-ephrin binding promiscuity. *J. Biol. Chem.* 285, 644–654. doi: 10.1074/jbc.M109.064824

Raftery, M., Muller, A., and Schonrich, G. (2000). Herpesvirus homologues of cellular genes. *Virus Genes* 21, 65–75. doi: 10.1023/A:1008184330127

Ratmann, O., Wiuf, C., and Pinney, J. W. (2009). From evidence to inference: probing the evolution of protein interaction networks. *HFSP J.* 3, 290–306. doi: 10.2976/1.3167215

Riley, R., Lee, C., Sabatti, C., and Eisenberg, D. (2005). Inferring protein domain interactions from databases of interacting proteins. *Genome Biol.* 6:R89. doi: 10.1186/gb-2005-6-10-r89

Russell, R. B., Alber, F., Aloy, P., Davis, F. P., Korkin, D., Pichaud, M., et al. (2004). A structural perspective on protein-protein interactions. *Curr. Opin. Struct. Biol.* 14, 313–324. doi: 10.1016/j.sbi.2004.04.006

Russell, R. B., and Aloy, P. (2008). Targeting and tinkering with interaction networks. *Nat. Chem. Biol.* 4, 666–673. doi: 10.1038/nchembio.119

Segura-Cabrera, A., Garcia-Perez, C. A., Guo, X., and Rodriguez-Perez, M. A. (2013). A viral-human interactome based on structural motif-domain interactions captures the human infectome. *PLoS ONE* 8:e71526. doi: 10.1371/journal.pone.0071526

Shackelton, L. A., and Holmes, E. C. (2004). The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol.* 12, 458–465. doi: 10.1016/j.tim.2004.08.005

Sinen, E., and Koyutürk, M. (2010). "Role of centrality in network-based prioritization of disease genes," in *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics,* eds C. Pizzuti, M. D. Ritchie, and M. Giacobini (Berlin: Springer).

Standfuss, J. (2015). Structural biology. Viral chemokine mimicry. *Science* 347, 1071–1072. doi: 10.1126/science.aaa7998

Tamamis, P., and Floudas, C. A. (2014). Molecular recognition of CCR5 by an HIV-1 gp120 V3 loop. *PLoS ONE* 9:e95767. doi: 10.1371/journal.pone.0095767

Teichmann, S. A. (2002). The constraints protein-protein interactions place on sequence divergence. *J. Mol. Biol.* 324, 399–407. doi: 10.1016/S0022-2836(02)01144-0

UniProt Consortium. (2014). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989

Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., and Teichmann, S. A. (2004). Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* 14, 208–216. doi: 10.1016/j.sbi.2004.03.011

Von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., et al. (2002). Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399–403. doi: 10.1038/nature750

Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S. M., and Eisenberg, D. (2002). DIP, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 30, 303–305. doi: 10.1093/nar/30.1.303

Xing, S., Wallmeroth, N., Berendzen, K. W., and Grefen, C. (2016). Techniques for the analysis of protein-protein interactions *in vivo*. *Plant Physiol.* 171, 727–758. doi: 10.1104/pp.16.00470

Yellaboina, S., Tasneem, A., Zaykin, D. V., Raghavachari, B., and Jothi, R. (2011). DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.* 39, D730–D735. doi: 10.1093/nar/gkq1229

Zheng, L. L., Li, C., Ping, J., Zhou, Y., Li, Y., and Hao, P. (2014). The domain landscape of virus-host interactomes. *Biomed. Res. Int.* 2014:867235. doi: 10.1155/2014/867235