Structural principles within the human-virus protein-protein interaction network

Eric A. Franzosa[a] and Yu Xia[a,b,c,1]

[a]Bioinformatics Program; [b]Department of Chemistry; and [c]Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, MA 02215

Edited* by Michael Levitt, Stanford University School of Medicine, Stanford, CA, and approved May 23, 2011 (received for review January 27, 2011)

General properties of the antagonistic biomolecular interactions between viruses and their hosts (*exogenous interactions*) remain poorly understood, and may differ significantly from known principles governing the cooperative interactions within the host (*endogenous interactions*). Systems biology approaches have been applied to study the combined interaction networks of virus and human proteins, but such efforts have so far revealed only low-resolution patterns of host-virus interaction. Here, we layer curated and predicted 3D structural models of human-virus and human-human protein complexes on top of traditional interaction networks to reconstruct the *human-virus structural interaction network*. This approach reveals atomic resolution, mechanistic patterns of host-virus interaction, and facilitates systematic comparison with the host's endogenous interactions. We find that exogenous interfaces tend to overlap with and mimic endogenous interfaces, thereby competing with endogenous binding partners. The endogenous interfaces mimicked by viral proteins tend to participate in multiple endogenous interactions which are transient and regulatory in nature. While interface overlap in the endogenous network results largely from gene duplication followed by divergent evolution, viral proteins frequently achieve interface mimicry without any sequence or structural similarity to an endogenous binding partner. Finally, while endogenous interfaces tend to evolve more slowly than the rest of the protein surface, exogenous interfaces—including many sites of endogenous-exogenous overlap—tend to evolve faster, consistent with an evolutionary "arms race" between host and pathogen. These significant biophysical, functional, and evolutionary differences between host-pathogen and within-host protein-protein interactions highlight the distinct consequences of antagonism versus cooperation in biological networks.

structural bioinformatics | structural systems biology | virus-host interaction

In host-virus protein-protein interactions (PPIs), components from the viral system invade and modulate the biological networks of the host in a highly antagonistic manner. As a result of this competitive relationship, the organizational, functional, and evolutionary principles of the host-virus PPI network are expected to differ from the better-understood principles governing the cooperative PPI network naturally occurring within the host. However, the evolved strategies by which viral pathogens evade the surveillance of the host immune system and hijack host cellular machinery for their own replication are not completely understood.

Traditional pathogen research studies host-virus PPIs in a one-at-a-time fashion. Recently, systems biology approaches have been applied to immunology (1) and pathogen research (2). Significant progress has been made in genome-wide mapping of host-pathogen PPIs for selected pathogens (3–8). This work has been successful in revealing systematic trends in host-pathogen interaction networks, e.g., that viruses tend to target host protein interaction hubs (2). Despite these recent experimental and computational advances in the analysis of host-pathogen interaction networks, there have been no attempts thus far to integrate interaction network and 3D structural data to extract general principles of host-pathogen PPI. Such an approach is essential, as structural information complements network information in a key way: while network information provides a holistic but low-resolution ("big picture") view of cellular events, 3D structural information provides a mechanistic and high-resolution view of interactions between specific biomolecules (9–13). The power of a structural approach in systems biology has been demonstrated in recent work employing protein structure information in the analysis of PPI networks (14–16). For example, structural models of PPIs in yeast were used to demonstrate a physical distinction (17) between the empirically observed "party"- and "date"-type protein hubs (18), and to elucidate patterns of intrinsic protein disorder among single- and multiinterface hubs (19).

In this work, we apply techniques and principles from structural systems biology to extract unique insights from the networks of human-virus PPIs. We find that viral proteins tend to bind to and mimic existing within-host PPI interfaces otherwise occupied by multiple, transiently bound regulators, and accelerate the evolution of those interfaces. Moreover, our work definitively demonstrates that the host-virus PPI network is governed by structural, functional, and evolutionary principles that are distinct from those governing the within-host PPI network. Compared to within-host PPI interfaces, host-virus PPI interfaces tend to be more transient, targeted by more host proteins, more regulatory in function, faster evolving, and rely more on convergent evolution to achieve interface mimicry. These results highlight the distinct consequences of antagonism vs. cooperation in PPI networks, with significant implications for the study of biological and social networks in general.

Results

Building a Human-Virus Structural Interaction Network. In a structural interaction network (SIN), every PPI is associated with a high-confidence 3D structural model. To reconstruct the human-virus SIN, we assembled structural models of interactions between pairs of human proteins (*endogenous* interactions) and structural models of virus proteins targeting human proteins (*exogenous* interactions). Models were based on biological assemblies from the Protein Data Bank (20), with human and virus proteins assigned to structures based directly on annotation or based on sequence homology (see *Methods*). The resulting human-virus SIN contains 3,039 endogenous interactions among 2,435 human proteins (Fig. 1), as well as 53 exogenous interactions between 50 virus proteins from 36 viral species and their 50 human target proteins (Fig. 1; Table S1). Additional details of the SIN are given in Datasets S1, S2, and S3; Figs. S1 and S2.

Author contributions: E.A.F. and Y.X. designed research; E.A.F. and Y.X. performed research; E.A.F. and Y.X. analyzed data; and E.A.F. and Y.X. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: yuxia@bu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1101440108/-/DCSupplemental.

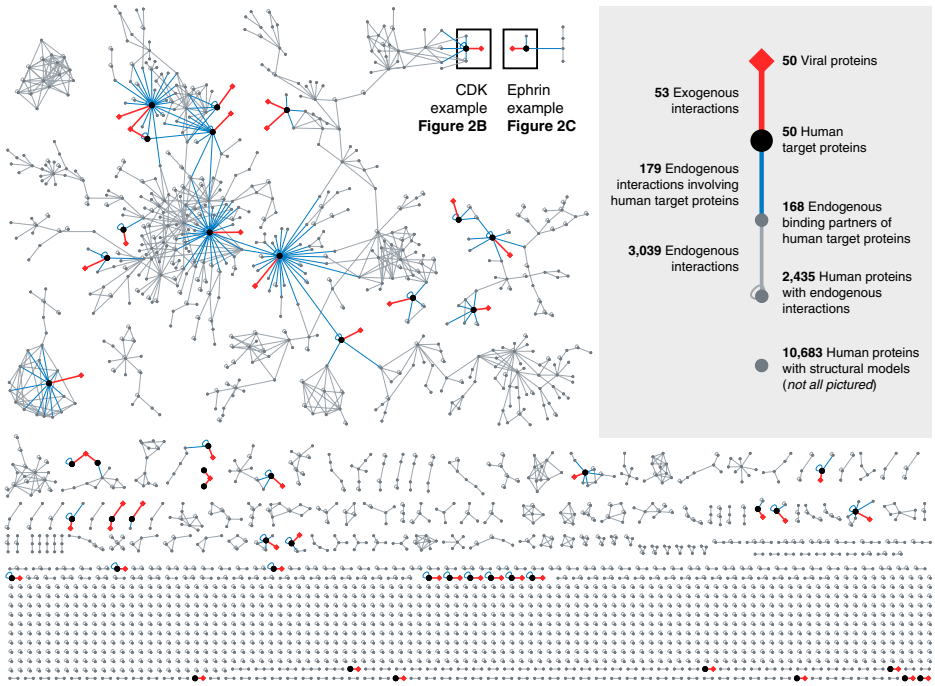

Fig. 1. The human-virus SIN. The network contains 3,039 endogenous (human-human) interactions among 2,435 human proteins alongside 53 exogenous (human-virus) interactions between 50 virus proteins from 36 viral species and their 50 human target proteins. See inset for symbol guide.

Extensive Overlap Between Exogenous and Endogenous Interfaces. Of the 53 exogenous interactions in the human-virus SIN, 41 are mediated by viral proteins targeting human proteins that are also involved in structured endogenous interactions. For these exogenous interactions, we investigated how often the involved viral proteins bind to their target proteins such that they overlap with one of the target proteins' endogenous interfaces (otherwise occupied by a human binding partner). We refer to this phenomenon as *interface mimicry*. Mimicry of host components is a general strategy employed by pathogens in their efforts to evade immune system detection and hijack host cellular machinery for their own purposes (21–25). Given that PPIs are responsible for mediating many cellular events, and given that these interactions depend on specific physical interfaces between proteins, interface mimicry is expected to be a common mechanism by which viruses modulate the biology of their hosts.

Mimicry of an endogenous interface by an exogenous interface can be quantified by the degree of similarity between the sets of target protein residues involved in the two interfaces, which we measure using the Jaccard similarity index: i.e., the number of residues involved in both interfaces, divided by the number of residues involved in either interface (subsequently referred to as *Jaccard interface similarity*). A target protein residue is "involved" in an interface with a human or viral protein binding partner if its solvent accessible surface area (SASA) in the unbound state decreases when the binding partner is present. Of the 53 total exogenous interfaces, 25 overlap with an endogenous interface such that the sets of involved target protein residues have Jaccard interface similarity greater than or equal to 0.25 (Fig. 2*A*; Table S1). The endogenous interfaces overlapped in these cases often recognize more than one human binding partner, such that a total of 92 endogenous interactions have Jaccard interface similarity greater than or equal to 0.25 with one of these 25 exogenous interactions. We subsequently refer to these groups of exogenous and endogenous interfaces as "extensively overlapping" with one another. Examples of extensive overlaps are highlighted in Fig. 2 *B* and *C*.

In total, 702 of 1,362 target protein residues involved in an exogenous interface (51.5%) are also involved in a known endogenous interface. To establish the statistical significance of this finding, we calculated the probability of observing a total overlap at least this extensive by making comparisons with randomly defined exogenous interfaces (*P*-value). We generated 1,000 sets of random interfaces with the same sizes (i.e., number of residues) as the observed exogenous interfaces by sampling from all viral target protein residues while maintaining an interface residue-like unbound SASA distribution. This procedure ensures that we are not biasing the randomization toward buried residues, which are less likely to be involved in endogenous interfaces. Even with this control, none of the 1,000 randomizations resulted in an overlap at least as extensive as the observed overlap between exogenous and endogenous interface residues. Hence, we conclude that the observed overlap is highly statistically significant (resampling-based one-tailed $P < 0.001$). The average expected fraction of exogenous interface residues overlapping with endogenous interfaces is 32.1% from 1,000 random trials, substantially smaller than the observed value of 51.5%. As the mapping of endogenous interfaces is not guaranteed to be exhaustive, it is likely that the observed degree of exogenous-endogenous overlap is an underestimate of the real value. Finally, exogenous interfaces (average size 949 Å^2) tend to be smaller than endogenous interfaces (average size 1,783 Å^2) across the SIN (permutation test, two-tailed $P < 0.01$), suggesting that the viral genome is under intense selection to reduce its size compared to the host genome.

Different Mechanisms for Evolving Interface Similarity in Virus and Host. Mimicry of a protein's interface can result from mimicry of the protein's structure, which can in turn result from mimicry of the protein's sequence. To investigate the extent to which the observed interface mimicry can be explained by structural and sequence mimicry, we measured the similarity at the structural level [Dali (26) *z*-score] and the sequence level [BLAST (27) *E*-value] between viral proteins and the human binding partners

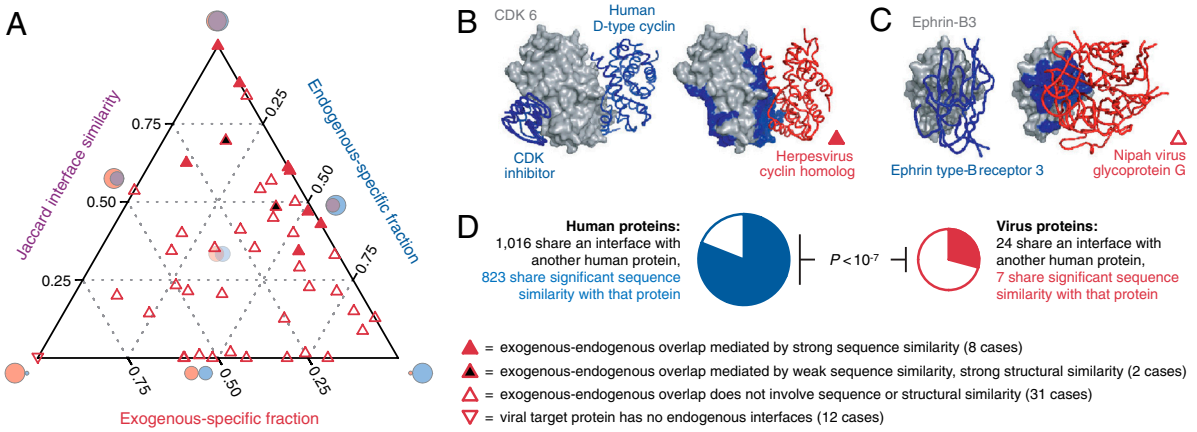

Fig. 2. Endogenous-exogenous interface overlap in the human-virus SIN. (*A*) The most significant case of endogenous-exogenous interface overlap is plotted for each exogenous interaction. If an exogenous interface extensively overlaps with multiple endogenous interfaces, we plot the endogenous interface whose human binding partner shares the greatest structural similarity with the viral protein. If no such structural similarity exists, we plot the endogenous interface with the greatest Jaccard interface similarity to the exogenous interface. Filled points indicate confirmed common ancestry between viral protein and mimicked human binding partner. (*B*) An example of extensive interface overlap with significant sequence similarity. In the left rendering, based on PDB structures 2iw8 (41) and 1bi8 (42), human *cell division protein kinase 6* (CDK6) is shown in complex with two of its endogenous binding partners, a *D-type cyclin* and a *CDK inhibitor*. In the right rendering, based on PDB structure 1g3n (43), the human binding partners have been removed, leaving their endogenous interfaces highlighted, and saimiriine herpesvirus 2 *cyclin homolog* is shown binding to the CDK. (*C*) An example of extensive interface overlap without significant sequence similarity. In the left rendering, based on PDB structure 3gxu (44), human *ephrin-B3* is shown in complex with *ephrin type-B receptor 3*. In the right rendering, based on PDB structure 2vsk (45), the receptor has been removed, leaving its endogenous interface highlighted, and Nipah virus *glycoprotein G* is shown binding to *ephrin-B3*. (*D*) Interface overlap in the endogenous network is significantly more likely to be mediated by sequence homology than in the exogenous network (Fisher's exact test, two-tailed $P < 10^{-7}$).

of their target proteins whose interfaces they overlap (Fig. 2*A*; Table S1).

Seven viral proteins mediating eight exogenous interactions have significant sequence similarity to at least one endogenous binding partner of their target protein whose interface they overlap extensively (BLAST *E*-value $<10^{-5}$; see Fig. 2*B* for a specific example). As a consequence, the structures of these virus-human protein pairs also tend to be significantly similar [Dali *z*-score >2; cutoff based on suggestion from (28)]. These exogenous interactions represent clear cases of horizontal gene transfer between the virus and the host (or a relative of the host), followed by divergent evolution. Two cases of extensive exogenous-endogenous interface overlap involve significant structural similarity but weak sequence similarity between a viral protein and the mimicked endogenous binding partner of its target protein; additional evidence [e.g., shared SCOP (29) superfamily] suggests that these can also be explained by horizontal gene transfer followed by divergent evolution.

The remaining 15 viral proteins mediating 15 further exogenous interfaces extensively overlapping with an endogenous interface show no obvious signs of structural or sequence similarity to the corresponding endogenous binding partners (see Fig. 2*C* for a specific example). The abundance of interface mimicry in the absence of sequence and structural similarity among viral proteins is surprising in comparison with properties of the endogenous interaction network. Within endogenous interaction networks, gene duplication is thought to play an important role in network expansion (30). Indeed, 1,016 of the human proteins in the SIN have an interface that extensively overlaps with the interface of at least one other human protein (Jaccard interface similarity ≥0.25), and 823 of these proteins (81.0%) share significant sequence similarity (BLAST *E*-value $<10^{-5}$) with the proteins whose interfaces they overlap. In contrast, only seven of the 24 viral proteins (29.2%) extensively overlapping with a human protein's interface share significant sequence similarity with that human protein, which is significantly less than the endogenous network (Fisher's exact test, two-tailed $P < 10^{-7}$; Fig. 2*D*).

This result suggests that, while both human and virus proteins may use sequence similarity (divergent evolution) as a means to establish interface similarity, the phenomenon is much more common among endogenous interactions. The evolution of the host genome is dominated by the regular expansion and reorganization of a rich repertoire of protein functions through domain shuffling and gene duplication (divergent evolution); reinventing a function by convergent evolution is less common in the host as a result of its slow mutation rate. Conversely, viral genomes tend to encode only a small number of proteins critical to survival, and as a result prefer different evolutionary mechanisms for establishing interface similarity. In addition to rapid divergent evolution following acquisition of coding sequences from the host, viruses appear to regularly establish interface similarity by convergent evolution as a result of their fast mutation rates, which is comparatively rare within the host.

Viruses Tend to Target Interfaces Transiently Bound by Multiple Regulators. Next, we characterized the dynamical properties of the endogenous interfaces mimicked by viral proteins. In particular, we asked if these interfaces are involved in permanent or transient interactions with endogenous binding partners. We infer that an endogenous interaction is more likely to be permanent if the proteins involved have been reported as coexpressed at the mRNA level in ref. 31; homodimeric interactions are excluded from this analysis, as they have coexpression values of 1 by definition. Of the 84 nonhomodimeric endogenous interactions extensively overlapped by an exogenous interaction, none are reported as coexpressed (0%), compared with 271 out of 1,608 (16.9%) total nonhomodimeric endogenous interactions coexpressed in the SIN background. This result suggests that mimicked endogenous interactions tend to be transient (Fisher's exact test, two-tailed $P < 10^{-6}$; Fig. 3*A*).

For an endogenous interface residue in the SIN, we define a property called *occupancy*, which is the number of endogenous interactions in which the residue participates. Interfaces with high average occupancy are "date"-like: they are used transiently by different endogenous binding partners at different times. The average occupancy of endogenous interface residues in the SIN is 1.47; 76.0% of endogenous interface residues are exclusive to one endogenous interface (have occupancy of 1; Fig. 3*B*). In contrast,

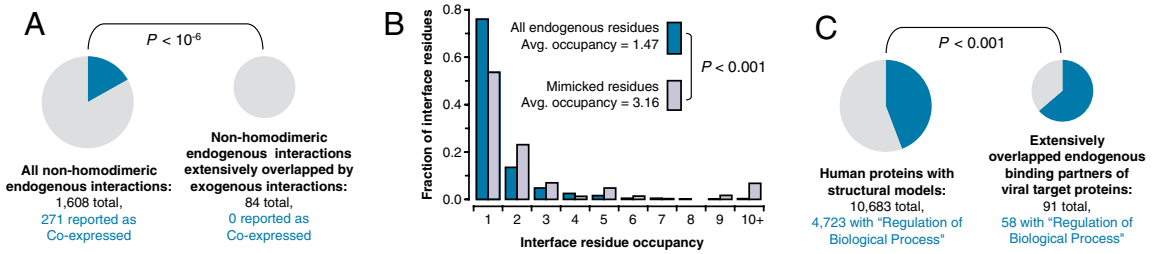

Fig. 3. Functional properties of endogenous interfaces targeted by viral proteins. (*A*) Compared to all endogenous interactions in the SIN, those extensively overlapped by exogenous interactions are significantly less likely to be coexpressed (Fisher's exact test, two-tailed $P < 10^{-6}$). (*B*) Distribution of interface residue occupancy (i.e., number of endogenous interactions in which the residue participates) for endogenous and mimicked interface residues. Compared to all endogenous interface residues in the SIN, residues mimicked by a viral protein participate in significantly more endogenous interactions (resampling-based one-tailed $P < 0.001$). (*C*) Human proteins binding to mimicked endogenous interfaces are significantly enriched for the GO slim term "Regulation of Biological Process" relative to generic human proteins with structural models (Fisher's exact test, two-tailed $P < 0.001$).

the average occupancy of mimicked interface residues—those targeted by both viral proteins and human binding partners—is 3.16, a significantly larger number than the population average (resampling-based one-tailed $P < 0.001$; Fig. 3*B*). Thus, the endogenous interfaces targeted by viruses tend to be date-like: on average, they are utilized by more human binding partners than generic endogenous interfaces. This result is consistent with the transient nature of the mimicked interactions identified above.

Lastly, we investigated functional enrichment among the mimicked endogenous binding partners using Gene Ontology (GO) slim (32). Among the 91 human proteins whose endogenous interfaces extensively overlap with an exogenous interface, the most common GO slim annotation is "Regulation of Biological Process," assigned to 58 proteins (63.7%). In comparison, this GO slim term is assigned to 4,723 out of 10,683 total human proteins with structural models (44.2%), and hence its enrichment among mimicked binding partners is highly statistically significant (Fisher's exact test, two-tailed $P < 0.001$; Fig. 3*C*). Thus, viruses tend to interrupt regulatory interactions.

Opposite Conservation Patterns in Exogenous vs. Endogenous Interfaces. Finally, we investigated the evolutionary properties of exogenous and endogenous interfaces in the human-virus SIN (Fig. 4; Tables S2 and S3). When we compare residues involved in endogenous interfaces to random sets of generic residues of equivalent unbound SASA distribution ("surface residues"), we find that the interface residues are on average significantly more evolutionarily conserved (human-mouse comparison, 7.8% divergence vs. 10.0% divergence; rejection-resampling-based one-tailed $P < 0.001$), in agreement with previous findings (33). In contrast, if we compare target protein residues involved in exogenous interfaces to generic surface residues, we find that they tend to be significantly less conserved (19.0% divergence vs. 10.0% divergence; $P < 0.001$). This opposite evolutionary behavior remains significant even if we restrict the analysis to surface residues of viral target proteins (15.6% divergence; $P < 0.001$).

This evolutionary distinction between exogenous and generic endogenous interfaces is intuitive: unlike generic endogenous interfaces, an exogenous interface actively targeted by viral proteins contributes to viral pathogenicity, has a negative impact on the fitness of the host, and therefore has no reason to be selectively constrained. Moreover, exogenous interface residues are not simply unconstrained, but rather tend to evolve faster than the rest of the target protein surface. Although direct evidence for positive selection ($dN/dS > 1$; Table S2) exists for only one exogenous interface in the SIN, the observed network-wide evolutionary rate acceleration within exogenous interfaces is consistent with an evolutionary "arms race" between host and virus: i.e., the virus evolves to bind to a host protein, and the host protein evolves to disrupt viral binding. Many examples of coevolutionary arms races between pathogens and their hosts are known (34), and they involve several human protein families actively targeted by positive selection (35).

Exogenous interfaces frequently overlap with endogenous interfaces, yet they exhibit opposite conservation patterns. This apparent contradiction points to something unusual about the endogenous interfaces of viral target proteins (Fig. 4). Endogenous interfaces from viral target proteins are not significantly more conserved than the surfaces of those target proteins (15.4% divergence vs. 15.6% divergence; $P = 0.340$). This finding is likely due in part to the fact that these endogenous interfaces are overlapped to a large extent by exogenous interfaces, which are fast evolving. To further explore this phenomenon, we divided interface residues from viral target proteins into three categories: those that participate in both endogenous and exogenous interfaces (mimicked residues), those that are exclusive to endogenous interfaces, and those that are exclusive to exogenous interfaces (see Fig. 4 for analysis results). We found that mimicked residues

	Human protein endogenous interface residues	Human protein surface residues †	Target protein endogenous-specific interface residues	Target protein endogenous interface residues	Target protein surface residues ††	Target protein exogenous-specific interface residues	Target protein exogenous interface residues	Target protein mimicked interface residues
# of residues	165,029	n/a	2,230	2,932	n/a	660	1,362	702
% divergence human vs. mouse	7.8	10.0	13.5	15.4	15.6	16.9	19.0	20.8
Surface comparison † Human / †† Target	*** / ***	n/a	*** / ***	*** / ns	n/a	*** / ns	*** / ***	*** / ***

Fig. 4. Evolutionary properties of exogenous and endogenous interfaces. Levels of human-mouse amino acid sequence divergence are quantified for sets of exogenous and endogenous interface residues and compared with "surface residues" (i.e., residues which follow an interface residue-like unbound SASA distribution) from generic human proteins and human proteins targeted by viruses. Residues involved in exogenous interfaces consistently evolve faster than the protein surface. Statistical significance (***$P < 0.001$; $^{ns}P > 0.05$) was determined from 1,000 rounds of rejection-resampling. Surface residues of both types are defined only during rejection-resampling, and so they have no absolute counts.

are evolving the fastest, and significantly faster than target protein surface residues (20.8% divergence vs. 15.6% divergence; $P < 0.001$). On the contrary, endogenous-specific interface residues are the most conserved, and—in line with the trend for generic endogenous interfaces—are significantly more conserved than target protein surface residues (13.5% divergence vs. 15.6% divergence; $P < 0.001$). Hence, while selection generally acts to conserve sites participating in endogenous interfaces, including those from viral target proteins, mutations at sites involved in both endogenous and exogenous interactions are less likely to be eliminated, perhaps due to an associated selective advantage resulting from disruption of viral binding. Finally, exogenous-specific interface residues have intermediate evolutionary rate which is slightly faster than that of target protein surface residues, but the difference is not statistically significant (16.9% divergence vs. 15.6% divergence; $P = 0.216$). This finding demonstrates that exogenous-specific interface residues are not more conserved than the rest of the surface residues, consistent with the general trend for all exogenous interface residues, and opposite to the general trend for all endogenous interface residues.

Following these analyses of sequence conservation, we explored additional enrichments for sequence-based traits among viral target proteins, endogenous interfaces, and exogenous interfaces (Table S3). We found that generic endogenous interfaces tend to be significantly enriched for hydrophobic amino acids and order-promoting amino acids, but significantly depleted for charged amino acids, all relative to generic protein surface residues. Residues involved in the endogenous interactions of viral target proteins are weakly enriched for order-promoting residues, but are neither significantly more hydrophobic nor significantly less charged than target protein surface residues: these results provide biophysical support for the earlier claim that target proteins are enriched for transient interfaces, which, as they spend time in an exposed state, are expected to be more "surface-like." Exogenous interfaces are similar to their target proteins' surfaces in terms of these biophysical properties.

Discussion

We employed a SIN analysis to study the properties of host-virus PPIs at high resolution. A SIN provides many advantages over a traditional PPI network, including precise definitions of the residues involved in the interfaces between interacting proteins. By enforcing that all PPIs in the human-virus SIN be supported by a solved 3D structure, we drastically reduce the influence of noise (false positives), a common problem in traditional conglomerated biological networks. Incompleteness (false negatives), experimental bias, and investigator bias remain as potential limitations for the SIN approach given the inherent difficulties associated with protein structure determination. For example, the SIN will tend to underrepresent interactions involving highly disordered proteins, and may be biased toward well studied viruses. Nevertheless, we emphasize that the SIN encompasses most of the human-human and virus-human PPIs for which 3D structural models are presently available. Most importantly, given that our comparisons and contrasts between exogenous and endogenous interactions are carried out within the SIN, our conclusions should be minimally confounded by any inherent biases from the structural approach.

We extracted multiple lines of evidence from the human-virus SIN, all of which point in the same direction: consistent with our intuition, antagonistic host-virus interactions exhibit a variety of general patterns in their behavior, many of which are unique or even opposite from the patterns found in the cooperative endogenous interactions within a single organism. Structural evidence revealed that mimicry of host binding interfaces by viruses is common, and that it is often achieved without structural similarity to the mimicked human binding partner. This finding is consistent with a pattern of convergent evolution on the part of the viruses, and is significantly different from the endogenous interaction network, in which interface similarity most often results from gene duplication and sequence homology. Structural and functional evidence further revealed that virus proteins tend to target date-like endogenous interfaces used by multiple, transiently bound human proteins enriched for regulatory activities. In contrast to endogenous interfaces within the host, evolutionary evidence revealed that viral target protein residues involved in exogenous interfaces tend to evolve faster than the rest of the protein surface.

The exogenous interactions in the SIN can be partitioned into two classes: (*i*) the globular-globular interaction class, involving multiple-span interfaces between a globular host protein and a globular viral protein; and (*ii*) the globular-peptide interaction class, where viral proteins interact with globular host proteins containing linear motif-binding domains by mimicking the peptide linear motifs recognized by those domains. The majority of the exogenous interactions in the SIN belong to the globular-globular class. Our conclusions are the same when the globular-globular class is analyzed separately (Table S4). Although the number of globular-peptide interactions in the SIN is too small for a separate assessment of statistical significance (Table S4), globular-peptide interactions as a group are expected to follow many of our trends. For example, unlike a globular domain, a peptide motif can easily arise in a viral protein by convergent evolution. Furthermore, like other mimicked interactions in the SIN, globular-peptide interactions tend to be transient, date-like (the same linear motif is recognized in multiple proteins), and of a regulatory nature.

Although the human-virus SIN represents a small sample of the universe of endogenous and exogenous PPIs, we expect these trends to be highly general in light of their clear mechanistic relevance. Taking these results together, a universal trend appears to emerge in which a virus, needing to infiltrate and modulate the biology of its host, evolves new molecular mechanisms to facilitate those goals. In the interest of economy, the virus hijacks the key components of the host's existing regulatory machinery, using interfaces and functions evolved for the host's benefit to its own advantage. This strategy imposes selective pressure on the host favoring counterstrategies that disrupt viral binding—and the cycle repeats. These insights are difficult to glean from a network of binary interactions alone, and they illustrate the power of integrating 3D structure data in network and systems biology.

Methods

We collected protein sequences and annotations for 20,328 reviewed human proteins and 1,013,058 virus protein sequences from the UniProt database (36). We collected atomic coordinate, biological assembly, and sequence data for all protein structures from the Protein Data Bank (20). Structures were mapped to human and virus sequences via BLAST-based sequence alignments.

We assembled structural models of exogenous interactions by three methods. The first method was based on structure annotation. If, in a given structure, one subunit was annotated as derived from human, and another was annotated as derived from a virus species, then this subunit pair was taken to represent an exogenous interaction between the human and virus proteins to which the subunits map most significantly (as determined by BLAST *E*-value). The second method was based on a combination of homology modeling and structure annotation. If, in a given structure, one subunit was annotated as derived from a nonhuman species, but mapped to a human protein with BLAST *E*-value $<10^{-10}$, and another subunit was annotated as derived from a virus species, then this subunit pair was taken to represent an exogenous interaction between the human and virus proteins to which the subunits map most significantly. The third and final method was based only on homology modeling. If, in a given structure, one subunit was annotated as derived from a nonhuman species, but mapped to a human protein with BLAST *E*-value $<10^{-10}$, and another subunit was annotated as derived from a nonvirus species, but mapped to a virus protein with BLAST *E*-value $<10^{-10}$, this subunit pair was taken to represent an exogenous interaction between the human and virus proteins if the two proteins were indepen-

dently known to interact (2). A unique exogenous interaction can have more than one structural model.

Once a structural model of an exogenous interaction was constructed, the exogenous interface was determined by measuring the SASA of the human target protein subunit first in isolation ("unbound"), and then with the viral protein subunit present ("bound"). Target protein residues whose SASA values decreased by at least 1 $Å^2$ upon binding of the viral protein subunit were considered to be part of the exogenous interface. SASA was determined with hydrogen atoms excluded using the program MSMS (37) and a 1.4 Å spherical solvent probe ("water molecule") (38). Residue interface participation and SASA data computed from structural models were aligned to the human target protein's reference sequence; SASA was averaged in regions of overlapping coverage by different structural models.

Exogenous interactions were filtered to remove redundancy (e.g., orthologous viral proteins from different strains with the same human protein target). Two exogenous interactions were said to be redundant if the human proteins involved were the same or highly similar (BLAST *E*-value $<10^{-5}$) and the virus proteins involved were the same or highly similar. Among redundant exogenous interactions, only the interaction with the largest number of structural models was considered. The exogenous interactions were further filtered manually to remove additional redundancy and cases of canonical immunological interaction (e.g., human antibodies targeting viral proteins).

We assembled structural models of human proteins and endogenous interactions following similar methods. To build structural models of human proteins, we mapped subunits of protein structures to (*i*) the most significantly aligned human protein, if the subunit was annotated as derived from human, and (*ii*) any human protein that shares highly significant sequence homology (BLAST *E*-value $<10^{-10}$) with the subunit. To build structural models of endogenous interactions, we employed two methods based on structure annotation and homology modeling. First, interacting pairs of human-derived subunits were automatically included as endogenous interactions between their associated proteins. Second, pairs of human proteins mapped to interacting subunits from the PDB by homology ("interologs") were included as endogenous interactions only if these protein pairs were independently reported as interacting in the IntAct database (39). Six additional interaction annotations from the Human Protein Reference Database (40) were included for viral target proteins based on manual curation. The remaining procedures were identical to those employed for the exogenous interactions.

Interface residues were resampled in several statistical analyses following a rejection-resampling method conditioned on the unbound SASA distribution of all interface residues (Fig. S1*A*); resampled residues were always drawn from structured regions of proteins. Human amino acid sequence conservation was determined by comparison with mouse (*Mus musculus*) protein sequences using ortholog relationships determined from the reciprocal best hit method.

ACKNOWLEDGMENTS. The authors thank John Connor, Rachel Fearns, Simon Kasif, Elke Mühlberger, and Michael Sorenson for helpful discussions; Sara Garamszegi for assisting with the curation and visualization of the dataset; and the anonymous reviewers for their constructive comments. This work was supported by a fellowship from National Science Foundation Integrative Graduate Education and Research Traineeship grant 0654108 (to E.A.F.) and a Research Starter Grant in Informatics from the Pharmaceutical Research and Manufacturers of America Foundation (to Y.X.).

1. Gardy JL, Lynn DJ, Brinkman FS, Hancock RE (2009) Enabling a systems biology approach to immunology: focus on innate immunity. *Trends Immunol* 30:249–262.
2. Dyer MD, Murali TM, Sobral BW (2008) The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathog* 4:e32.
3. Maxwell KL, Frappier L (2007) Viral proteomics. *Microbiol Mol Biol R* 71:398–411.
4. Tan SL, Ganji G, Paeper B, Proll S, Katze MG (2007) Systems biology and the host response to viral infection. *Nat Biotechnol* 25:1383–1389.
5. Calderwood MA, et al. (2007) Epstein-Barr virus and virus human protein interaction maps. *Proc Natl Acad Sci USA* 104:7606–7611.
6. von Schwedler UK, et al. (2003) The protein network of HIV budding. *Cell* 114:701–713.
7. Brass AL, et al. (2008) Identification of host proteins required for HIV infection through a functional genomic screen. *Science* 319:921–926.
8. Uetz P, et al. (2006) Herpesviral protein networks and their interaction with the human proteome. *Science* 311:239–242.
9. Jones S, Thornton JM (1996) Principles of protein-protein interactions. *Proc Natl Acad Sci USA* 93:13–20.
10. Chothia C, Janin J (1975) Principles of protein-protein recognition. *Nature* 256:705–708.
11. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein-protein recognition sites. *J Mol Biol* 285:2177–2198.
12. Teichmann SA, Murzin AG, Chothia C (2001) Determination of protein function, evolution and interactions by structural genomics. *Curr Opin Struct Biol* 11:354–363.
13. Aloy P, et al. (2004) Structure-based assembly of protein complexes in yeast. *Science* 303:2026–2029.
14. Aloy P, Russell RB (2006) Structural systems biology: modelling protein interactions. *Nat Rev Mol Cell Biol* 7:188–197.
15. Beltrao P, Kiel C, Serrano L (2007) Structures in systems biology. *Curr Opin Struct Biol* 17:378–384.
16. Kiel C, Beltrao P, Serrano L (2008) Analyzing protein interaction networks using structural information. *Annu Rev Biochem* 77:415–441.
17. Kim PM, Lu LJ, Xia Y, Gerstein MB (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 314:1938–1941.
18. Han JD, et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430:88–93.
19. Kim PM, Sboner A, Xia Y, Gerstein M (2008) The role of disorder in interaction networks: a structural analysis. *Mol Syst Biol* 4:179.
20. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242.
21. Benedict CA, Norris PS, Ware CF (2002) To kill or be killed: viral evasion of apoptosis. *Nat Immunol* 3:1013–1018.
22. Finlay BB, McFadden G (2006) Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* 124:767–782.
23. Bernet J, Mullick J, Singh AK, Sahu A (2003) Viral mimicry of the complement system. *J Biosci* 28:249–264.
24. Alcami A (2003) Viral mimicry of cytokines, chemokines and their receptors. *Nat Rev Immunol* 3:36–50.
25. Elde NC, Malik HS (2009) The evolutionary conundrum of pathogen mimicry. *Nat Rev Microbiol* 7:787–797.
26. Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233:123–138.
27. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
28. Holm L, Kaariainen S, Rosenstrom P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24:2780–2781.
29. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247:536–540.
30. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5:101–113.
31. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P (2004) Coexpression analysis of human genes across many microarray datasets. *Genome Res* 14:1085–1094.
32. Barrell D, et al. (2009) The GOA database in 2009—an integrated Gene Ontology Annotation resource. *Nucleic Acids Res* 37:D396–403.
33. Choi SS, Vallender EJ, Lahn BT (2006) Systematically assessing the influence of 3-dimensional structural context on the molecular evolution of mammalian proteomes. *Mol Biol Evol* 23:2131–2133.
34. Woolhouse ME, Webster JP, Domingo E, Charlesworth B, Levin BR (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet* 32:569–577.
35. Vallender EJ, Lahn BT (2004) Positive selection on the human genome. *Hum Mol Genet* 13:R245–R254.
36. The UniProt Consortium (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* 38:D142–148.
37. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38:305–320.
38. Franzosa EA, Xia Y (2009) Structural determinants of protein evolution are context-sensitive at the residue level. *Mol Biol Evol* 26:2387–2395.
39. Aranda B, et al. (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res* 38:D525–531.
40. Keshava Prasad TS, et al. (2009) Human Protein Reference Database—2009 update. *Nucleic Acids Res* 37:D767–772.
41. Pratt DJ, et al. (2006) Dissecting the determinants of cyclin-dependent kinase 2 and cyclin-dependent kinase 4 inhibitor selectivity. *J Med Chem* 49:5470–5477.
42. Russo AA, Tong L, Lee JO, Jeffrey PD, Pavletich NP (1998) Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* 395:237–243.
43. Jeffrey PD, Tong L, Pavletich NP (2000) Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. *Genes Dev* 14:3115–3125.
44. Qin H, et al. (2009) Structural characterization of the EphA4-Ephrin-B2 complex reveals new features enabling Eph-ephrin binding promiscuity. *J Biol Chem* 285:644–654.
45. Bowden TA, et al. (2008) Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2. *Nat Struct Mol Biol* 15:567–572.