STRUCTURAL MODELS FOR HOST-PATHOGEN PROTEIN-PROTEIN INTERACTIONS: ASSESSING COVERAGE AND BIAS

ERIC A. FRANZOSA

Bioinformatics Program
Boston University, 24 Cummington Street
Boston, MA 02215, USA
Email: franzosa@bu.edu

YU XIA

Bioinformatics Program, Department of Chemistry, Department of Biomedical Engineering
Boston University, 24 Cummington Street
Boston, MA 02215, USA
Email: yuxia@bu.edu

Recently, we applied structural systems biology to host-pathogen interaction and constructed the human-virus structural interaction network (SIN) based on a combination of solved structures and homology models. Subsequent analysis of the human-virus SIN revealed significant differences between antagonistic human-virus and cooperative within-human protein-protein interactions (PPIs). Although the SIN approach is advantageous due to the complementary nature of 3D structure and network data, integration of these data sources is associated with two potential issues: reduced coverage of the full human-virus PPI network, and the introduction of specific biases from structure determination. In this work, we evaluate the impact of these issues by comparing the growth and properties of human-virus and within-human PPI networks with and without structural models. We find that although the human-virus SIN is small in size, it is largely depleted for false positives, which are common in the full network. In addition, the SIN shows potential for major growth in the near future. Furthermore, compared to the full network, the coverage of viral species in the human-virus SIN is large relative to its size, suggesting that it may be a less biased sampling of the universe of human-virus PPIs. Next, we systematically compare structural versus full networks of human-virus and within-human PPIs in terms of functional, physicochemical, and network properties. We find that although there exist biases inherent to the structural approach, such biases tend to affect both human-virus and within-human PPIs equally. As a result, the significant differences between structured human-virus and within-human PPI networks are never contradicted by the full networks. Collectively, these results suggest that a structural approach to host-pathogen systems biology is not only justified, but also highly complementary to previous approaches. In particular, conclusions drawn from direct comparisons of host-virus and host-host PPIs within the SIN are minimally confounded by the inherent biases of the structural approach.

1. Introduction

Systems biology approaches have been recently applied to immunology [1] and pathogen research [2] following significant progress in the genome-wide mapping of host-pathogen protein-protein interactions (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 involved in characteristic cellular processes, such as cell cycle regulation [2]. Concurrently, a structural approach to general problems in systems biology has started to gain momentum [9-11]. Structural systems biology focuses on the analysis of structural interaction networks (SINs), a class of PPI network in which every interaction is supported by a solved 3D structure of the interacting proteins. These two data types are highly complementary: network information provides a holistic but low-resolution view of cellular events, and 3D structure information provides a mechanistic and high-resolution view of interfaces between specific biomolecules. Indeed, the structural systems biology approach has already been applied to answer interesting questions in network biology, such as the physical nature of party- and date-type protein hubs [12], and the role of native disorder in protein-protein interaction [13].

In a recent work we combined these two trends, applying a structural systems biology approach to the analysis of host-pathogen interaction networks [14]. We combined 3D structure and traditional network data for human-virus (exogenous) and within-human (endogenous) PPIs to construct a combined human-virus SIN. Our analysis of the human-virus SIN revealed that viral proteins tend to bind to and mimic existing endogenous interfaces, and that they often do so without structural homology to an endogenous binding partner. Furthermore, we found that the endogenous interfaces targeted by viruses tend to be otherwise occupied by multiple, transiently bound ("date-like") endogenous regulators, and tend to evolve significantly faster than generic protein surface sites. Interface similarity in the absence of structural homology, date-like binding, and fast evolutionary rate are comparatively rare among cooperative endogenous interactions and interfaces, and hence our discoveries highlighted the distinct consequences of antagonistic host-virus interaction.

While a SIN provides many advantages over a traditional PPI network, it also associated with several known or potential drawbacks. For example, by enforcing that all PPIs in our human-virus SIN be supported by a solved 3D structure, our coverage of the universe of known human-virus and within-human PPI is dramatically reduced. In addition to incompleteness, experimental and investigator bias remain as potential limitations for the structural systems biology approach given the inherent difficulties associated with protein structure determination. For example, the human-virus SIN will tend to under-represent interactions involving highly disordered proteins, and may be biased toward well-studied proteins. Given that our comparisons and contrasts between exogenous and endogenous interactions are carried out within the SIN, the hope is that our conclusions are minimally confounded by the inherent biases from the structural approach. In this work, we provide evidence in support of this assertion, and present additional support for a structural approach to the analysis of host-pathogen interaction networks.

We begin by analyzing the size and rate of growth of the currently known universe of structured and binary human-virus and within-human PPIs. In this work, the term *binary* interaction refers to a physical association between two proteins that is supported by

experimental evidence [15]. A *structured* interaction refers to a direct, physical association between two proteins that is supported by a 3D model. Naturally, structured interactions are a subset of known binary interactions. We demonstrate that, while small, the set of structured PPIs constituting the human-virus SIN is more accurate and more representative of the true universe of human-virus interaction than the full binary network. We then compare sequence, structural, and functional properties within these networks, revealing that, although biases from the structural approach exist, they tend to affect exogenous and endogenous interactions equally, and hence do not have a confounding effect on trends interpreted from the SIN. With the number of structured exogenous interactions increasing dramatically, structural systems biology is rapidly becoming a key route for exploring principles of host-pathogen interactions.

2. Constructing Binary and Structural Human-Virus Interaction Networks

We constructed structural and binary networks of human-virus (exogenous) and within-human (endogenous) protein-protein interactions (PPIs) for analysis and comparison. Endogenous PPI data were assembled from the IntAct database [15]; there are a total of 35,303 endogenous interactions. Human-virus PPI data were taken principally from VirusMINT [16] and supplemented with additional reports from IntAct; this produced an initial set of 1,814 exogenous interactions. Exogenous interactions were filtered to remove redundancy (e.g., orthologous viral proteins from different strains with the same human protein target). Based on sequence data sequence from the UniProt database [17], two exogenous interactions were said to be redundant if the human proteins involved were the same or highly similar (BLAST [18] 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 independent reports was considered. Endogenous interactions are considered to be non-redundant by definition.

Structural models were assigned to interactions following previous work [14, 19]. Briefly, a 3D structure of interacting proteins from the Protein Databank [20] acts as a model for an exogenous interaction if (i) the structured proteins are a human protein and a virus protein, (ii) the structured proteins are a human ortholog and a virus protein, or (iii) the structured proteins have very strong sequence homology (BLAST E-value $< 10^{-10}$) to human and virus proteins that are independently known to interact. Models satisfying the latter two conditions are referred to as *homology models*. Structural models of exogenous interactions were filtered for redundancy following the procedure described above; there are a total of 53 structured exogenous interactions. Structural models of endogenous interactions were assigned following conditions (i) and (iii) for pairs of interacting human proteins; there are a total of 3,039 structured endogenous interactions.

3. Sizes and Growth Rates of Human-Virus Interaction Networks

From the period of 1989 to 2009, the number of experimentally determined interactions in each of the four classes (binary endogenous, structured endogenous, binary exogenous, and structured

exogenous) has increased by orders of magnitude (Figure 1A). As a result of large-scale functional genomics techniques such as yeast two-hybrid and affinity capture, it is now possible to rapidly enumerate binary protein-protein interactions within a species. The number of experimentally determined human endogenous interactions (27,646) dwarfs all other counts by the end of 2009, and appears to be growing dramatically (Figure 1B). It is worth noting, however, that although the number of *reported* endogenous interactions is large, the vast majority of these (95.2% as of the end of 2009) have never been *confirmed* by a second experiment, and hence their accuracy remains largely unknown.

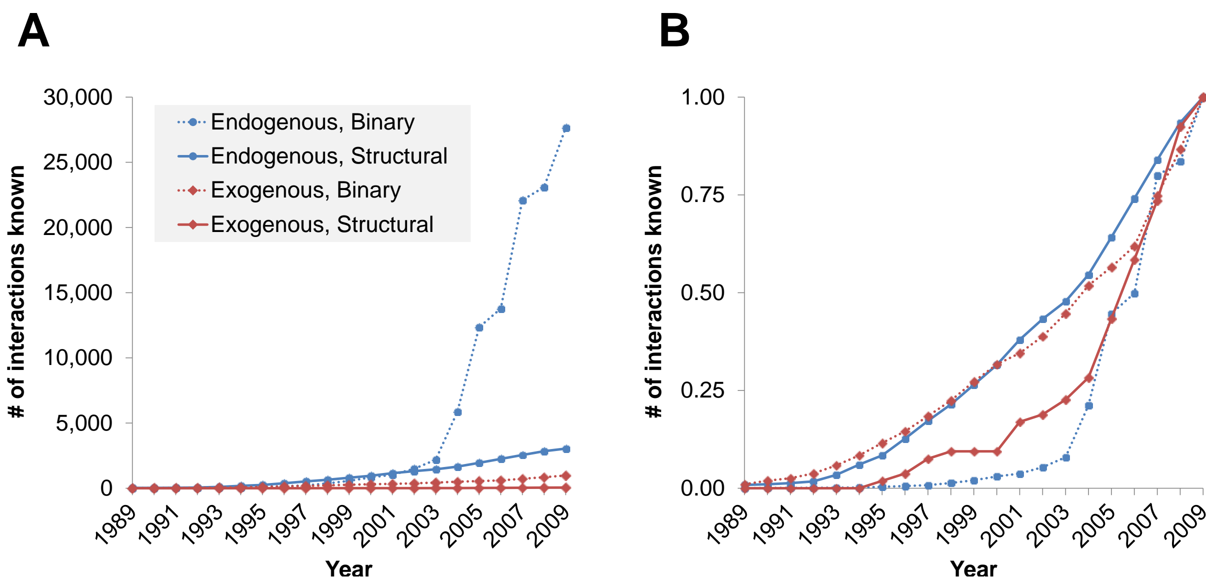

Figure 1. Growth of exogenous (human-virus) and endogenous (within-human) interaction networks, binary and structural. (**A**) Absolute numbers of interactions known, 1989-2009. (**B**) Relative numbers of interactions known, compared to end-of-2009 totals, 1989-2009.

Structured endogenous interactions are the second most abundant class, but their number (3,039) remains far below that of their binary counterparts (Figure 1A). Unlike binary endogenous interactions, the techniques for generating structures of protein complexes remain mostly low-throughput and labor-intensive, and so these interactions have been accumulating at a slower, more linear pace (Figure 1B). Roughly 10% of the structured endogenous class is composed of homology models—i.e., pairs of proteins that are known to interact and have strong sequence similarity to a solved structure (which serves as a 3D "scaffold" for their interaction). In order to maintain a high degree of accuracy, we employed a stringent sequence similarity cutoff when assigning homology models to proteins (BLAST *E*-value $< 10^{-10}$; 80% average alignment coverage). Relaxing this cutoff or employing alternative homology modeling techniques could increase the number of models substantially, and the size of the structured endogenous class by extension. Compared with binary interactions derived from noisy, high-throughput experiments, structured interactions—including homology models—will be dominated by true positives.

Exogenous interactions are the next-most abundant class, numbering 976 at the end of 2009 (Figure 1A). Compared with endogenous interactions, the number of experimentally determined exogenous interactions is growing at a more linear pace (Figure 1B), and a larger (but still small) fraction of these interactions (17.5%) have been confirmed by multiple experiments. Only 53 interactions fall in the structured exogenous class (Figure 1A), of which 11 were inferred from homology. As with structured endogenous interactions, the structured exogenous list could be expanded using additional homology modeling techniques or relaxed homology criteria. For example, alternative methods could be used to build models of viral proteins which present short linear motifs to human linear motif-binding domains [21]. Curiously, although determining the structure of exogenous interactions is subject to the same challenges and time constraints of endogenous structure determination, their relative numbers have recently begun increasing dramatically (Figure 1B). This suggests to us that structural biologists are rapidly turning their attention toward problems in host-pathogen interaction. We believe that the growth of the exogenous and endogenous networks will further accelerate as structural genomics projects naturally shift their focus from single proteins to protein-protein interactions in the near future.

4. Viral Diversity in Human-Virus Interaction Networks

The size of the total universe of exogenous interactions remains unknown. We assert here that pooled human-virus interactions from the VirusMINT and IntAct databases represent a thorough catalog of the *known* universe of binary human-virus interactions, although these are probably only a small sample of the total universe. Even if we discard the approximately half of these interactions which are redundant, the fraction of the remaining exogenous interactions having known structures is small (5.3%; Figure 2A). We therefore conclude that the human-virus SIN represents a very small sample of the total universe of exogenous interactions. This does not imply that it is inappropriate to make statistical inferences based on the structured sample, as the statistical significance of a trend depends upon its effect size (weak or strong) and the sample's absolute size, but not its relative coverage of the population.

On the other hand, the structured exogenous interactions could be a *biased* sample of the universe of human-virus PPIs, which could influence measured effect sizes. However, the structured exogenous interactions cover 21.7% of the viral proteins in the non-redundant exogenous dataset (Figure 2B), and 50.0% of the viral species (as determined by NCBI taxonomic ID; Figure 2C). These differences in interactions-to-protein and interactions-to-species ratios between structured and non-redundant exogenous interactions result from the latter's overrepresentation of interactions from viruses which have been the focus of high-throughput surveys, such as HIV and Epstein-Barr virus [5, 7]. As a result, with respect to viral diversity, the structured exogenous network may represent a less biased sample of the total universe of exogenous PPIs than the binary exogenous PPI network, and could therefore be a more appropriate sample to consider for the purposes of statistical hypothesis testing. This finding also demonstrates that, while we are principally concerned here with biases inherent to

the structural interaction network, traditional networks are also vulnerable to experimental and investigator biases.

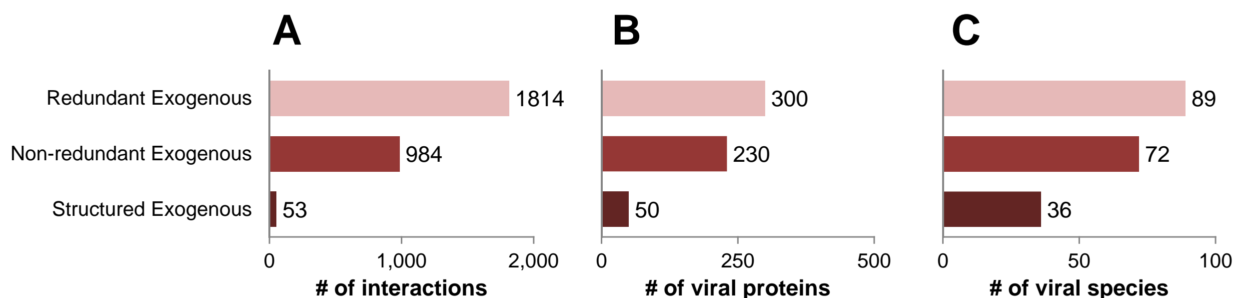

Figure 2. Properties of exogenous interaction networks. (**A**) Number of exogenous interactions represented in the redundant, non-redundant, and structured human-virus interaction networks. (**B**) Number of viral proteins represented in these networks. (**C**) Number of viral species represented in these networks (based on NCBI taxonomic ID). The structural interaction network covers 5% of known non-redundant interactions but 50% of their source species.

5. Evaluating Potential Biases from the Structural Systems Biology Approach

Although we assert that the human-virus SIN includes a relatively unbiased sampling of viral diversity, it may still be biased in other ways. As discussed in the Introduction, structured interactions will tend to reflect the experimental constraints of structure determination. For example, proteins with large disordered regions or extensive transmembrane domains are difficult to crystallize, and hence their structures may elude discovery by X-ray crystallographic methods. Given the targeted rather than high-throughput nature of structured interaction determination, investigator bias may also play a role in skewing the sample of interactions in a structural interaction network. For example, having determined the structure of a given protein in complex with one of its binding partners, an investigator may tend to continue to study the structural interactions of this protein (or similar proteins), rather than moving on to a completely different system.

Such biases from the structural approach are likely to exist in the human-virus SIN. What remains to be shown is whether or not a measured difference between the exogenous and endogenous interactions in the human-virus SIN can be *explained* by such biases. Compared to within-host PPI interfaces, we observed in our previous work that host-virus PPI interfaces tend to be more transient, targeted by more host proteins, more regulatory in function, faster evolving, and more reliant on convergent evolution to achieve interface mimicry. If we checked each of these trends within the total binary exogenous and endogenous interaction networks, several scenarios are possible (Figure 3). The first possibility is that the binary networks follow the same trend observed in the SIN (Figure 3A). In this case, there is no bias from the structural approach, and the absolute and relative level of a certain property for exogenous and endogenous interactions does not depend on whether or not the interactions are structured. The second possibility is that there is a bias introduced by the structural approach, but that it affects the

exogenous and endogenous interactions uniformly (Figure 3B). In this case, the absolute level of a certain property for exogenous and endogenous interactions *does* depend on whether or not the interactions are structured, but the relative difference between them does not. In both of these cases (Figure 3 A and B), the exogenous levels are uniformly different from the endogenous levels, indicating that the trend is biologically relevant. This may not always be the case. In Figure 3C, a trend difference between the structured exogenous and structured endogenous interactions does *not* recur between the total binary exogenous and total binary endogenous interactions, and in Figure 3D the trend is reversed. In these cases, the trend from the structural interaction network appears to be driven solely by biases from the structural approach, and is not biologically relevant.

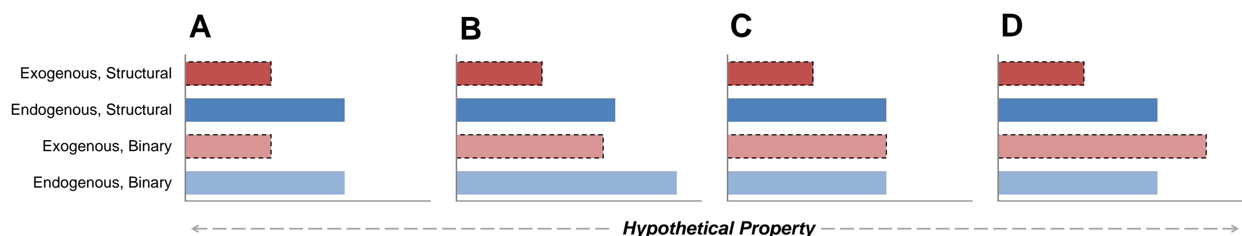

Figure 3. Illustration: The influence of structural bias on trends from the human-virus SIN. In each scenario, there is strong trend between structured exogenous and structured endogenous interactions with respect to some hypothetical property (*x*-axis); assume all differences are significant. (**A**) An example with no structural bias: the structured and full binary human-virus networks show the same trend. (**B**) An example with uniform structural bias: the structured and full binary human-virus networks show the same trend. (**C** & **D**) Examples with confounding structural bias: a trend observed in the human-virus SIN is not reflected (C) or is reversed (D) in the full binary network.

Many of the properties in which we are interested can only be measured in a structural interaction network—an issue which motivates the structural approach. For example, we cannot determine from binary interactions alone the average degree of interface overlap between exogenous and endogenous interfaces (interface mimicry), a trend we know to be significant from the structural network [14]. In order to demonstrate that trends from the human-virus structural interaction network hold for the larger human-virus binary interaction network, we must employ properties that can be measured with high coverage in both networks. We overcame this obstacle by focusing on the human proteins in these networks (enumerated in Figure 4A), and by selecting properties that were (i) available as high coverage annotations, e.g., major Gene Ontology (GO) Slim Process terms [22]; (ii) based on large-scale experiments, e.g., coexpression with binding partners [23]; or (iii) directly predicted from sequence, e.g., hydropathy index [24]. The complete set of properties that we considered is listed in Table 1. Alpha helix, beta sheet, and disorder propensities were based on previously published amino acid scales [25, 26]; domains and predicted transmembrane helices were assigned using the *Pfam* database [27] and *TMHMM* server [28], respectively; the remaining properties were directly

calculated from our previously described network and sequence data. Recall that human proteins involved in the exogenous networks are viral target proteins.

GO Slim Process Terms	Structured vs. Binary	Binary Exo. vs. Binary Endo.	Struct. Exo. vs. Struct. Endo.
Biosynthesis	>	>	
Catabolism	>>	>	
Cell differentiation	>	>	
Cellular process	>	>	
Development	>	>	
Macromolecule metabolism	>	>	
Metabolism	>	>	
Nucleic acid metabolism	>	>	
Regulation of biological process*	>	>	>
Response to stimulus*	>>	>>	>>
Transport	>	>	

Other Properties	Structured vs. Binary	Binary Exo. vs. Binary Endo.	Struct. Exo. vs. Struct. Endo.
Alpha helix propensity	>		
Beta sheet propensity	<		
Disorder propensity	<		
Extent of *Pfam* domain coverage	>	>	
Hydropathy index	<	>	
Predicted transmembrane helices*	<<	<<	
Size	<	<	
Clustering coefficient		<	
Coexpression w/ binding partners*	>>	>>	
Degree in endogenous network*	>>	>>	
Extent of gene duplication	>		

KEY: Sample is...

3/2 fold enriched	>>
significantly enriched	>
not significantly different	
significantly depleted	<
2/3 fold depleted	<<

Table 1. Properties of human proteins in structured and binary human-virus PPI networks. We explored enrichments for 22 protein properties in three sets of comparisons: structured endogenous vs. binary endogenous, binary exogenous vs. binary endogenous, and structured exogenous vs. structured endogenous. Statistical significance was determined from 10^4 rounds of random sampling and subjected to FDR correction with $\alpha = 0.1$ following the Benjamini-Hochberg method. See the inset key to interpret the significance, sign, and magnitude of the enrichments in each comparison. Starred (*) comparisons are highlighted in Figure 4.

Table 1 further summarizes the results of the comparisons made within and between the structured and binary networks with respect to our chosen protein properties. Statistical significance was determined from 10^4 rounds of random sampling and subjected to FDR correction with $\alpha = 0.1$ following the Benjamini-Hochberg method to account for multiple hypothesis testing. Many of the properties surveyed differ significantly between human proteins found in structured interactions versus those found in binary interactions, and in five cases the differences are very strong (Table 1). In most cases the effect size is quite small, and the statistical significance is mainly due to large sample size. This confirms our intuition that the challenges of protein structure determination impose experimental and investigator biases on the structural interaction network. For example, structured proteins are strongly depleted for predicted transmembrane helices (0.5-fold depletion, $p < 0.0001$), a form of experimental bias, and are consistently enriched for all major GO Slim Process terms, which may be a sign of investigator bias toward well-studied proteins.

Two of the 22 properties are significantly enriched among the structured exogenous group relative to the structured endogenous group: the GO Slim Process terms "regulation of biological process" (1.3-fold enrichment, $p = 0.0074$; Figure 4B) and "response to stimulus" (1.8-fold enrichment, $p < 0.0001$; Figure 4C). These are examples of significant structural exogenous-endogenous trends as illustrated in Figure 3. In both cases we observe a bias from the structural

approach: structured endogenous interactors are more enriched for these GO terms than generic human interacting proteins (Figure 4 B and C). However, as in the example from Figure 3B, this bias appears to be uniform: i.e., structured viral targets are also more enriched for regulatory activity and stimulus response than generic viral targets (Figure 4 B and C). As a result, these significant trends in the human-virus SIN also appear in the binary human-virus interaction network. The idea that viral targets are enriched for regulatory activity, in particular regulation of the cell cycle, has been previously suggested [2]. Enrichment for the broad term "response to stimulus" is not surprising given that this term is inclusive of the more specific "immune response" term.

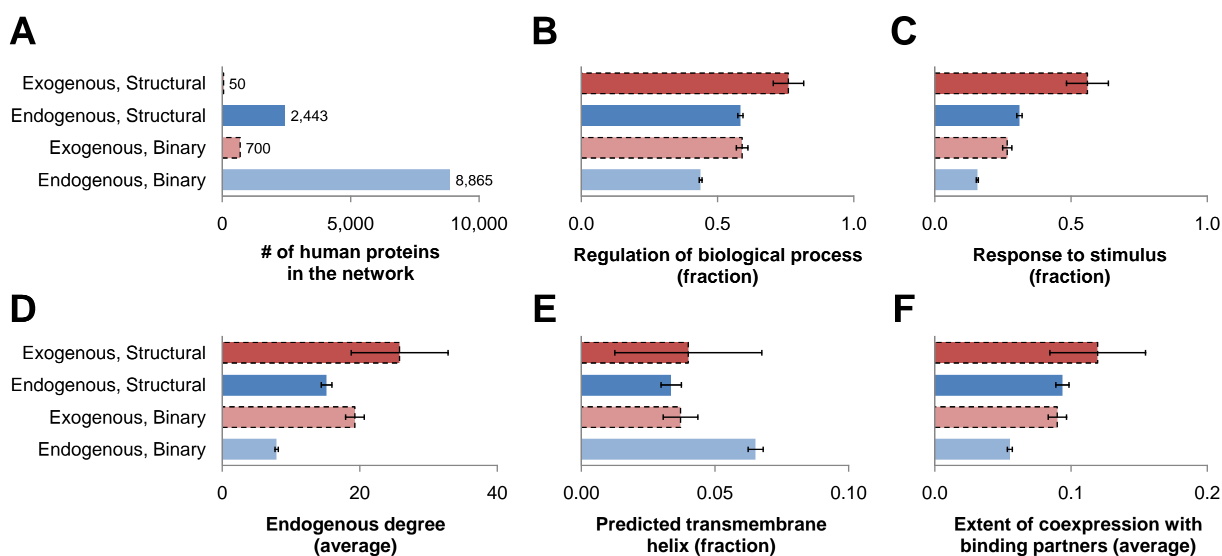

Figure 4. Significant trends in human-virus PPI networks. (**A**) The number of human proteins in each of the four networks; human proteins involved in the exogenous networks are viral target proteins. (**B** & **C**) For these properties we observe a significant exogenous-endogenous trend in the structural networks. In both cases the trend is supported by the binary interaction networks, indicating that it cannot be explained by structural biases, although they are present. (**D**, **E**, & **F**) For these properties we observe a significant exogenous-endogenous trend in the binary networks, but the structural trend is borderline significant (D) or not significant (E & F). Such discrepancies may result from reduced sample size in the structural networks (A), but do not undermine the structural approach. Error bars reflect 95% bootstrap confidence intervals.

There are other properties which differ in a strong, significant manner between the binary exogenous and binary endogenous networks, but not between their structured counterparts (Table 1). For example, as previously demonstrated by Dyer et al. [2], generic viral targets tend to have higher average endogenous degree than generic human proteins, and are said to be more hub-like (2.4-fold increase, $p < 0.0001$; Figure 4D). This trend achieves borderline significance within the human-virus SIN (1.7-fold increase, $p = 0.0560$; Figure 4D). Generic viral targets also tend to be depleted for predicted transmembrane helices (0.6-fold decrease, $p = 0.0013$; Figure 4E) and are more likely to be coexpressed with their endogenous binding partners (1.6-fold increase,

$p < 0.0001$; Figure 4F). Notably, in none of these instances does a significant difference between generic viral targets and human interacting proteins *contradict* a significant difference between their structured counterparts, as postulated in Figure 3D (Table 1). Although there are cases where a significant difference is *absent* in the structural interaction network, in both cases where a significant difference is present, a complementary significant difference is also observed in the binary networks (Table 1; Figure 4 B and C).

The fact that some differences are only observed as significant in the full network speaks to one of the strengths of binary host-pathogen interaction networks: scale. Due to the small size of the human-virus SIN (Figure 4A), and in particular the small number of structured exogenous interactions, a structural systems biology approach may lack the power to detect a weak exogenous-endogenous interaction trend. To understand the ramifications of such a trend in the structural interaction network, including its precise physical basis, will require the generation of a larger number of structured exogenous interactions. Fortunately, interest in determining the structures of host-virus interactions appears to be growing rapidly (Figure 1B). It is also worth noting that a significant trend observed only in the binary network could be explained by experimental and investigator biases that are not found in the human-virus SIN, such as over-representation of certain viruses that have been targeted by large-scale screens.

6. Conclusions

There are many advantages to a structural approach to studying the systems biology of host-pathogen interaction. By previously applying such an approach [14], we demonstrated that, compared to within-host protein-protein interaction (PPI) interfaces, host-virus PPI interfaces tend to be more transient, targeted by more host proteins, more regulatory in function, faster evolving, and more reliant on convergent evolution to achieve interface mimicry. These types of high resolution trends would be difficult or impossible to ascertain using traditional systems biology techniques that rely on binary interaction data.

In this work, we addressed two potential drawbacks inherent to a structural approach to systems biology: reduced coverage, and the introduction of biases from structure determination. With respect to coverage, we argued that—while the human-virus SIN is indeed small compared to current binary networks—the number of exogenous interactions of known structure is growing rapidly, and that these numbers can be further bolstered by the application of homology modeling techniques. Moreover, we emphasized that relative coverage of the total universe of human-virus PPIs is not critical to statistical hypothesis testing, provided that the sample size is sufficiently large to detect a trend of a given effect size. In addition, compared with binary interaction networks, interactions within the human-virus SIN are highly accurate (contain few false positives) and provide a less biased representation of viral diversity, both of which facilitate statistical hypothesis testing.

By analyzing a variety of properties that can be calculated for both structural and binary human-virus interaction networks, we demonstrated that, although strong structural biases exist, they tend to uniformly affect viral target proteins and generic human interacting proteins. In

other words, if there is a significant exogenous-endogenous trend in the human-virus structural interaction network, the trend will also be significant in the binary human-virus interaction network. For example, we confirmed here that viral target proteins (both structured and non-structured) tend to be enriched for regulatory activities [2]. This finding supports our original assertion that conclusions based on comparisons and contrasts made within the SIN ought to be minimally confounded by inherent biases from the structural approach. Conversely, it is sometimes possible to observe a significant trend in the binary human-virus interaction network but not in the human-virus SIN. For example, we confirmed here that viral target proteins in the binary network tend to be hubs [2], and showed further that they tend to be co-expressed with their binding partners. In these cases, a discrepancy between the structured and traditional approach is acceptable: due to the reduced size of the SIN relative to the full binary network, the structural approach will sometimes lack the statistical power to detect weaker trends.

Previously, we demonstrated that the interfaces mimicked by viruses tend to be otherwise used by multiple, transiently bound (date-like) human interactors [14]. However, in this work, we found that viral target proteins in the binary interaction network tend to be coexpressed with their endogenous binding partners ($p < 0.0001$), a hallmark of constitutive, "party-like" protein-protein interaction—the opposite of transient, date-like behavior. These two results seem contradictory, but are in fact compatible. In the structural systems biology approach, a protein is a 3D structure, and we are able to make statements about its individual interfaces. In the traditional systems biology approach, a protein is a node, and so we can only speak of the protein's average behavior (or the behavior of an "average interface" of that protein). Hence, while the average endogenous interaction involving a generic viral target protein tends to involve coexpression, the specific interactions and interfaces targeted by viruses may still tend to be transient. In other words, party-like hubs may contain date-like interfaces, and these interfaces may be preferentially targeted by viruses (perhaps due to regulatory function). Such contrasts point to the fact that traditional systems biology research is highly complementary to the structural approach, and we expect both methods to continue to contribute to our understanding of host-pathogen interactions in the near future.

References

1. Gardy, J.L., et al. *Trends Immunol*, **30**:249-262 (2009).
2. Dyer, M.D., T.M. Murali, and B.W. Sobral. *PLoS Pathog*, **4**:e32 (2008).
3. Maxwell, K.L. and L. Frappier. *Microbiol Mol Biol Rev*, **71**:398-411 (2007).
4. Tan, S.L., et al. *Nat Biotechnol*, **25**:1383-9 (2007).
5. Calderwood, M.A., et al. *Proc Natl Acad Sci U S A*, **104**:7606-11 (2007).
6. von Schwedler, U.K., et al. *Cell*, **114**:701-13 (2003).
7. Brass, A.L., et al. *Science*, **319**:921-6 (2008).
8. Uetz, P., et al. *Science*, **311**:239-42 (2006).
9. Aloy, P. and R.B. Russell. *Nat Rev Mol Cell Biol*, **7**:188-97 (2006).
10. Beltrao, P., C. Kiel, and L. Serrano. *Curr Opin Struct Biol*, **17**:378-84 (2007).
11. Kiel, C., P. Beltrao, and L. Serrano. *Annu Rev Biochem*, **77**:415-441 (2008).

12. Kim, P.M., et al. *Science*, **314**:1938-41 (2006).
13. Kim, P.M., et al. *Mol Syst Biol*, **4**:179 (2008).
14. Franzosa, E.A. and Y. Xia. *Proc Natl Acad Sci U S A*, **108**:10538-43 (2011).
15. Aranda, B., et al. *Nucleic Acids Res*, **38**:D525-31 (2010).
16. Chatr-aryamontri, A., et al. *Nucleic Acids Res*, **37**:D669-73 (2009).
17. The UniProt Consortium. *Nucleic Acids Res*, **38**:D142-8 (2010).
18. Altschul, S.F., et al. *Nucleic Acids Res*, **25**:3389-402 (1997).
19. Franzosa, E.A. and Y. Xia. *Mol Biol Evol*, **26**:2387-95 (2009).
20. Berman, H.M., et al. *Nucleic Acids Res*, **28**:235-42 (2000).
21. Davey, N.E., G. Trave, and T.J. Gibson. *Trends Biochem Sci*, **36**:159-69 (2011).
22. Barrell, D., et al. *Nucleic Acids Res*, **37**:D396-403 (2009).
23. Lee, H.K., et al. *Genome Res*, **14**:1085-94 (2004).
24. Kyte, J. and R.F. Doolittle. *J Mol Biol*, **157**:105-32 (1982).
25. Koehl, P. and M. Levitt. *Proc Natl Acad Sci U S A*, **96**:12524-9 (1999).
26. Campen, A., et al. *Protein Pept Lett*, **15**:956-63 (2008).
27. Finn, R.D., et al. *Nucleic Acids Res*, **36**:D281-8 (2008).
28. Krogh, A., et al. *J Mol Biol*, **305**:567-80 (2001).