Prediction of host-virus interaction networks

Janet M. Doolittle

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Bioinformatics and Computational Biology.

Chapel Hill
2012

Approved by:

Tim Elston, PhD

Shawn Gomez, Eng.Sc.D.

Klaus Hahn, PhD

Keith Burridge, PhD

Anne Taylor, PhD

Jennifer Webster-Cyriaque, PhD

ABSTRACT

JANET DOOLITTLE: Prediction of host-virus interaction networks
(Under the direction of Shawn Gomez)

As with other viral pathogens, HIV-1 and dengue virus (DENV) are dependent on their hosts for the bulk of the functions necessary for viral survival and replication. Thus, successful infection depends on the pathogen's ability to manipulate the biological pathways and processes of the organism it infects, while avoiding elimination by the immune system. Protein-protein interactions are one avenue through which viruses can connect with and exploit these host cellular pathways and processes.

We developed a computational approach to predict interactions between HIV and human proteins based on structural similarity of 9 HIV-1 proteins to human proteins having known interactions. Using functional data from RNAi studies as a filter, we generated over 2,000 interaction predictions between HIV proteins and 406 unique human proteins. Additional filtering based on Gene Ontology cellular component annotation reduced the number of predictions to 502 interactions involving 137 human proteins. We find numerous known interactions as well as novel interactions showing significant functional relevance based on supporting Gene Ontology and literature evidence.

We then applied this approach to predict interactions between (DENV) and both of its hosts, *Homo sapiens* and the insect vector *Aedes aegypti*. We predict over 4,000 interactions between DENV and humans, as well as 176 interactions between DENV and

A. aegypti. Additional filtering based on shared Gene Ontology cellular component annotation reduced the number of predictions to approximately 2,000 for humans and 18 for *A. aegypti*. Of 19 experimentally validated interactions between DENV and humans extracted from the literature, this method was able to predict nearly half. Our results suggest specific interactions between virus and host proteins relevant to interferon signaling, transcriptional regulation, stress, and the unfolded protein response.

Viruses manipulate cellular processes to their advantage through specific interactions with the host's protein interaction network. The interaction networks presented here provide a set of hypothesis for further experimental investigation into viral life cycles and potential therapeutic targets.

ACKNOWLEDGMENTS

I am very grateful to the many people who have made this dissertation possible. First, I would like to sincerely thank my co-advisors Shawn Gomez and Klaus Hahn, who have supported me in countless ways over the years. They have given me advice, encouragement, scientific opportunities, financial support, and taught me how to think about science. I am also indebted to Keith Burridge, Erica Wittchen, and Tim Elston for their support and advice on various projects. Thank you to the members of my committee for your guidance and suggestions.

I would also like to thank all the members of the Gomez and Hahn labs. From the Hahn lab, I would especially like to recognize Evan Trudeau, Marie Rougie, Jason Yi, and Ellen O'Shaughnessy for training me in "wet bench" biology and helping me with collaborative projects. From the Gomez lab, special thanks to Jen Staab, Matt Berginski, and Kyla Lutz for their friendship and suggestions on my work. I am grateful to the administrative staff of the Pharmacology, Biomedical Engineering, and Genetics departments, who have helped me in many different ways, especially Sausyty Hermreck. I acknowledge the Bioinformatics and Computational Biology training program at UNC for financial support during 2009-2010.

I am grateful to all my friends that have not been mentioned yet, especially Alicia Midland, for their moral support. Thanks to Richard Hall for his constant support and love and many good times together. I would also like to thank Vinny and Celia for giving me companionship and entertainment during the past couple years and for "helping" me

write. Last but not least, I would like to thank my parents, Frank and Deb, and brother, John. My family has always given me unconditional love and support, and they made me who I am today. I dedicate this dissertation to them.

PREFACE

Parts of this dissertation have been previously published in Open Access journals:

Doolittle, J. M., & Gomez, S. M. (2010). Structural similarity-based predictions of protein interactions between HIV-1 and Homo sapiens. *Virology Journal*, *7*, 82. Doi:10.1186/1743-422X-7-82

Doolittle, J. M., & Gomez, S. M. (2011). Mapping protein interactions between Dengue virus and its human and insect hosts. *PloS Neglected Tropical Diseases*, *5*(2), e954. Doi:10.1371/journal.pntd.0000954

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACTN2: alpha actinin 2

ACTN4: alpha actinin 4

ARFIP2: arfaptin 2

C: DENV capsid protein

CA: HIV-1 capsid protein

CC: Cellular Component

DAVID: Database for Annotation, Visualization and Integrated Discovery

DCTN1: dynactin 1

dDENV-similar: D. melanogaster proteins with structural similarity to a DENV protein

DENV: dengue virus

DF: Dengue Fever

DHF: Dengue Hemorrhagic Fever

DroID: Drosophila Interactions Database

DSS: Dengue Shock Syndrome

E: DENV envelope protein

EBV: Epstein-Barr virus

EEA1: early endosomal autoantigen 1

eIF2a: eukaryotic initiation factor 2a

ESCRT: endosomal sorting complex

GO: Gene Ontology

GSEA: Gene Set Enrichment Analysis

GST: Glutathione S-transferase

GTP: guanosine triphosphate

GTPase: guanosine triphosphatase

HCV: hepatitis C virus

hDENV-similar: human proteins with structural similarity to a DENV protein

HHPID: HIV-Human protein interaction database

HIV-1: Human immunodeficiency virus type 1

HPRD: Human Protein Reference Database

Hrs: hepatocyte growth factor regulated tyrosine kinase substrate

IFN: interferon

IN: HIV-1 integrase

KSHV: Kaposi Sarcoma-associated virus

M: DENV matrix protein

MA: HIV-1 matrix protein

MSN: moesin

n-butyrate: sodium butyrate

NC: HIV-1 nucleocapsid protein

NCK adaptor proteins 1 and 2 (NCK1/2)

NS1: DENV non-structural protein 1

NS2A: DENV non-structural protein 2A

NS2B: DENV non-structural protein 2B

NS3: DENV non-structural protein 3

NS4A: DENV non-structural protein 4A

NS4B: DENV non-structural protein 4B

NS5: DENV non-structural protein 5

PDB: Protein Data Bank

PERK: PKR-like ER kinase

PIG: Pathogen Interaction Gateway

PKC: Protein kinase C

PR: HIV-1 protease

prM: DENV matrix protein before cleavage of the pr peptide

RASA1: RAS GTPase activating protein 1

RNAi: Ribonucleic Acid Interference

ROCK1: Rho-associated protein kinase 1

RT: HIV-1 reverse transcriptase

SIFTS: Structure integration with function, taxonomy and sequence

SPTA1: erythrocytic spectrin alpha

SPTB: erythrocytic spectrin beta

TAP-MS: Tandem Affinity Purification-Mass Spectrometry

TPA: 12-O-tetradecanoylphorbol-13-acetate

UPR: Unfolded Protein Response

UTR: Untranslated region

Y2H: yeast 2-hybrid

CHAPTER ONE

INTRODUCTION

IMPORTANCE OF HOST-VIRUS INTERACTIONS

The viral lifecycle cannot be completed without extensive aid from a host cell. Viruses encode very few proteins and must subvert the biological machinery and processes of their hosts to replicate. The strikingly few proteins encoded by viruses and their heavy dependence on their hosts raise the questions: How can a pathogen encoding so few proteins evade the sophisticated and multi-pronged human immune system and how can we stop it?

Identifying the ways in which a virus manipulates a human cell will lead to the development of novel anti-viral drugs. Many antiviral drugs target viral enzymes. For example, successful anti-HIV drugs include inhibitors of the viral protease and reverse transcriptase enzymes, which are often used in combination therapy due to the fast mutation rate of the viral genome and rapid development of resistance. Many clinically relevant viruses encode very few enzymes; most targets that are considered "druggable" are enzymes. However, the viral lifecycle could also be disrupted by targeting the host pathways required for replication, a strategy that has been successful for cellular kinases (Schang, 2006). In addition, despite the challenges, several small molecules blocking protein interactions, rather than enzyme activity, have been developed in recent years

(Arkin & Wells, 2004). The entry inhibitor class of anti-HIV drugs prevents interactions between HIV and cell surface receptors. One example is maraviroc (Selzentry, Pfizer), an FDA approved drug that blocks viral entry by preventing the HIV gp120 protein from binding to human CCR5 (Dorr et al., 2005). Further understanding of protein interactions that mediate essential steps of the viral lifecycle will yield novel drug targets.

Understanding the interface through which the pathogen connects with and manipulates its host requires knowledge of the molecular points of interaction between them. Specifically, knowledge of the protein interactions between pathogen and host is of particular value. While the prediction of protein interactions within species such as *S. cerevisiae* and *H. sapiens* has been pursued for some time (Jansen et al., 2003; Qi, Klein-Seetharaman, & Bar-Joseph, 2007; Yamanishi, Vert, & Kanehisa, 2004; L. V. Zhang, Wong, King, & Roth, 2004), it is only recently that host-pathogen interactions have come under greater scrutiny. Computational approaches are of significant value in the host-pathogen context as large-scale experimental characterization of these interactions is non-trivial (Calderwood et al., 2007; De Chassey et al., 2008; Tan, Ganji, Paeper, Proll, & Katze, 2007; Uetz et al., 2006).

EXPERIMENTAL HOST-VIRUS INTERACTOMES

Little is known about the network of protein interactions between pathogens and their hosts. Most protein interaction studies are small scale, determining only one or a few interactions at once. However, large-scale screens, mostly using yeast 2-hybrid (Y2H), have been conducted for several viruses, providing large datasets of protein-

protein interactions that are valuable for constructing the host-pathogen network despite the high rate of false positives and false negatives in Y2H studies.

A combination of Y2H screening and literature mining were used to construct an hepatitis C virus (HCV)-human interactome involving 481 interactions between 11 HCV proteins and 421 human proteins (De Chassey et al., 2008). Both direct and functional interactions were found for an H1N1 strain of influenza virus by Y2H and microarray analysis, with the functional importance of human proteins validated by siRNA (Shapira et al., 2009). For vaccinia virus, the virus used as a vaccine for smallpox, a Y2H screen found 109 interactions between viral and human proteins, with 63% of selected interactions confirmed by GST-tagged pulldown (L. Zhang et al., 2009).

In addition, several studies have been conducted on herpesviruses. Y2H screens were conducted to test for interactions between Epstein-Barr Virus (EBV) proteins and for interactions between EBV and human proteins. Fifty-two EBV proteins were shown to participate in 60 interactions, either with other EBV proteins or with themselves (homodimerization). In addition, 40 EBV proteins interacted with 112 human proteins in the screen, with some human proteins interacting with multiple viral proteins (Calderwood et al., 2007). The network of interactions between Kaposi Sarcoma-associated herpesvirus (KSHV) proteins was determined by Y2H and merged into the human interactome based on predicted interactions. The intraviral network was highly interconnected and exhibited properties differing from that of the human interactome. However, the merged host-pathogen network recovered many properties of the human network, suggesting that the viral network is incomplete without considering interactions with the host (Uetz et al., 2006).

Several general principals emerge from these large-scale studies. First, networks of interactions between viral proteins are very highly interconnected, whereas most cellular proteins have few interaction partners (Shapira et al., 2009; Uetz et al., 2006). In addition, a number of viral proteins were found to interact with multiple human proteins, suggesting that one viral protein may have many functions in hijacking the cell (Calderwood et al., 2007; De Chassey et al., 2008). A common finding in several screens is that viruses tend to interact with human proteins that are highly connected in the human interactome. Two ways to measure connectivity in a network are degree, or the number of edges that directly connect the node to other nodes, and betweenness, or the number of shortest paths between any two nodes that pass through a particular node. Viral proteins have a tendency to interact with human proteins that have a significantly higher degree (hubs) and betweenness than other human proteins (Calderwood et al., 2007; De Chassey et al., 2008).

PREVIOUS METHODS TO PREDICT HOST-PATHOGEN INTERACTIONS

Even large scale experimental methods, such as Y2H, do not provide full coverage of the host-pathogen network. Therefore, several recent computational methods have been developed and applied to host-pathogen interactions, suggesting additional potential interactions in different host-pathogen systems. For instance, Dyer et al. used Bayesian statistics to first calculate the probability of protein interactions between proteins from the same species based on their domain composition. Then they predicted interactions between the eukaryotic parasite *Plasmodium falciparum* and human proteins by the presence of domains with a high probability of interacting (Dyer, Murali, &

Sobral, 2007). Also focusing on malaria, Lee and colleagues generated predictions based on interactions between orthologous proteins from eukaryotes (Lee et al., 2008). However, these methods are difficult to apply to viruses. Especially for viruses with small genomes, viral proteomes have few domains, with even fewer shared between the host and virus, and few orthologs in mammalian species.

In the context of HIV-human interactions, two previous computational methods were applied. In the first study, Tastan et al. used a computational approach based on the random forest method to predict protein interactions using features taken from human proteins and the human interactome (Tastan, Qi, Carbonell, & Klein-Seetharaman, 2009). This approach requires a large training set of already known interactions, and can only be applied to well-studied viruses, like HIV. In the second study, Evans et al. predicted possible interactions using short sequence motifs found in both HIV-1 and human proteins and the human domains that bind these motifs (Evans, Dampier, Ungar, & Tozeren, 2009). However, not all interactions are mediated by short linear motifs and it is unclear whether the same relationships between domains and motifs that are found in human interactions are relevant to host-virus interactions.

Most approaches have not utilized the significant amount of protein structure information that is increasingly available. Specifically, rapid progress in structure determination technologies has led to the establishment and deposition of massive numbers of protein structures into the Protein Data Bank, with over 60,000 protein structures currently deposited (Berman, Henrick, & Nakamura, 2003). In combination with documented protein-protein interactions, the use of protein structure information provides another means for the prediction of possible protein interactions (Aloy &

Russell, 2003; Davis et al., 2006; Lu, Lu, & Skolnick, 2002). The central premise in such approaches is that, given a set of proteins with defined structures and associated interactions, proteins with similar structures or substructures will tend to share interaction partners. In the context of host-pathogen interactions, Davis et al. used sequence and structural comparisons to known complexes to predict potential protein interactions for non-viral pathogens responsible for several neglected diseases (Davis, Barkan, Eswar, McKerrow, & Sali, 2007). Many of the pathogen proteins had no known crystal structure and were predicted by homology modeling.

Here, we present a novel method for predicting host-pathogen interactions using similarities between protein structures. Our approach is broadly applicable to any virus for which crystal structures exist or reliable structural models can be made. A large set of previously known interactions is not required to train the classifier, which is important when studying neglected viruses such as dengue virus. Furthermore, sequences or domain classifications do not need to be conserved between the virus and its host. Predictions were made for HIV-1, a well-studied virus, and dengue virus, a poorly understood virus for which there is no vaccine or specific treatment. No previous large scale interaction networks have been created for dengue virus and humans, and few small scale protein interaction studies have been performed. Knowledge of virus-host protein interactions will lead to new drug targets for treatment of human disease. In addition, interactions were predicted for the mosquito vector of dengue virus, providing insights into this phase of the viral lifecycle that may aid vector control efforts.

CHAPTER TWO

STRUCTURAL SIMILARITY-BASED PREDICTIONS OF PROTEIN INTERACTIONS BETWEEN HIV-1 AND HOMO SAPIENS

INTRODUCTION

Pathogen invasion and survival requires that the pathogen interact with and manipulate its host. Human immunodeficiency virus type 1 (HIV-1) encodes only 15 proteins and must therefore rely on the host cell's machinery to accomplish vital tasks such as the transport of viral components through the cell and the transcription of viral genes (Frankel & Young, 1998; Goff, 2007). HIV-1 infects human cells by binding to CD4 and a co-receptor, fusing with the cell membrane and uncoating the virion core in the cytoplasm (Frankel & Young, 1998). The genomic RNA is then reverse transcribed and the DNA enters the nucleus as part of a viral pre-integration complex, containing both viral and host proteins. Afterwards, the viral DNA is inserted into the genome by viral integrase (IN) (Goff, 2007). The integrated provirus is transcribed by host RNA polymerase II from a promoter located in the provirus long terminal repeat, and the RNA is exported to the cytoplasm (Frankel & Young, 1998; Goff, 2007). Host machinery translates HIV-1 mRNA, and several of the resulting proteins are transported to the cell membrane to be packaged into the virion along with the genomic RNA and multiple host proteins. The virus then buds from the cell and undergoes a maturation process, which

enables it to infect other cells (Frankel & Young, 1998). Throughout this process, host proteins play an indispensable role.

Here, we develop a map of interactions between HIV-1 and human proteins based on protein structural similarity. In this approach, we first retrieve structural similarity between host and pathogen proteins identified by an established method that compares known crystal structures. Human proteins identified as having a region of high structural similarity to an HIV protein are referred to as "HIV-similar." Next, we identify known interactions for these HIV-similar proteins, with the one or more human proteins that they interact with referred to as "targets." We then assume that HIV proteins have the same interactions as their human, HIV-similar counterparts, allowing HIV to plug into the host cell protein network at these points (Figure 2.1). Using data from recent RNAi screens and cellular co-localization information, we refine this interaction map to enrich for those interactions having the greatest potential to be correct based on the available information. Evaluation of these predictions shows a statistically significant enrichment of known interactions and numerous novel interactions with potential functional relevance. These predictions provide an additional tool for further investigations into the lifecycle of HIV-1 and identification of potential clinical targets.

RESULTS AND DISCUSSION

Identification of HIV-similar human proteins

To construct a map of interactions between HIV-1 and human proteins, we established a multi-step protocol that begins with the identification of human proteins having significant structural similarity to HIV-1 proteins (Figure 2.2). We used the Dali Database (Holm, Kääriäinen, Rosenström, & Schenkel, 2008; Holm & Sander, 1993), which contains 3D structure comparisons for all protein structures in the Protein Data Bank (PDB); all publicly available crystal structures for HIV-1 and *H. sapiens* are contained within PDB. While the crystal structure for many human proteins is unknown, most HIV-1 proteins have been at least partially resolved. Specifically, crystal structures exist for PR, RT, IN, CA, MA, NC, Gag p2, gp120, gp41, Nef, Tat, Vpr, and Vpu (Table 2.1). The three enzymes encoded by HIV-1, protease (PR), reverse transcriptase (RT), and integrase (IN) are the best characterized structurally, having at least 25 structures each in the PDB, with PR having over 300. CA, gp41, and gp120 are also well studied. We note, however, that many of these structures represent only part of the full-length protein. HIV-1 proteins having regions of high similarity to at least one human protein include: gp41, gp120, CA, MA, Gag p2, PR, IN, RT, and Vpr (selected similarities to IN are given in Appendix 1). Therefore, predictions were made for nearly every HIV-1 protein that has a published structure.

Selected examples of structural similarities between the HIV-1 proteins IN, RT, and gp41 and human proteins determined by Dali are shown in Figure 2.3. The structural similarities frequently involve only part of each protein. However, because in most cases

the precise location of protein interaction sites is not known, we used the entire structure in our investigation.

Protein interaction prediction

After identifying which specific HIV-1 and human proteins have high structural similarity, we extract all known interactions for human proteins from the Human Protein Reference Database, which contains over 37,000 documented protein interactions (Mishra et al., 2006). Again, the central premise is that given a network of protein interactions, proteins with similar structures or substructures will tend to have similar interaction partners. Thus, our hypothesis is that HIV-1 proteins having similar structure to one or more human proteins are also likely to participate in the same set of protein interactions (Figure 2.1). Under these assumptions, we directly mapped HIV-1 proteins to their high-similarity matches within this network.

To reduce the number of predictions and provide an additional line of functional evidence for interactions and their possible biological relevance, we filtered these results using two types of datasets on host proteins involved in HIV-1 infection, collectively referred to as "Literature Filters" hereon. The first filter type represents host proteins that have been shown to impair HIV-1 infection or replication when knocked down by siRNA or shRNA. Three genome-scale siRNA screens have been conducted in HeLa or 293T cells (Brass et al., 2008; König et al., 2008; Zhou et al., 2008). A fourth study with a similar goal was conducted using shRNA in Jurkat T-cells, a more realistic model of HIV-1 infection (Yeung, Houzet, Yedavalli, & Jeang, 2009). Each of the four screens found over 250 host proteins involved in HIV-1 infection. Remarkably, very little overlap exists between these studies, perhaps due to differences in methods, including the cell

lines and stages of the HIV-1 life cycle investigated. The second type of data used to filter predictions identifies human proteins present in the HIV-1 virion. During budding, host proteins from both the cell surface and the cytoplasm, including some involved in the cytoskeleton, signal transduction, metabolism, and chaperone functions, may be incorporated into the virion (Chertova et al., 2006). While some of these proteins may be taken up by the budding virus simply by chance, others are known to be specifically incorporated into the virion and may play key roles in viral life cycle or pathogenesis. For example, TSG101 may be incorporated due to its interaction with Gag and facilitates budding (Cantin, Méthot, & Tremblay, 2005; Chertova et al., 2006).

We considered only predicted interactions where the target protein was observed in at least one of the previously described Literature Filters. The resulting predicted HIV-human interaction network consists of 2,143 interactions, considering all unique combinations of Uniprot accessions for an HIV-1 protein and a predicted human interactor (Figure 2.2). Of the predictions that were made, 62 were verified as true interactions based on data from two databases of known host-pathogen interactions, HHPID and PIG (Appendices 2 and 3). There were 347 human proteins predicted to have structural similarities with an HIV-1 protein, and the predictions implicate a total of 406 unique human proteins as potentially interacting with HIV-1 (Table 2.2).

We visually examined some of the structural similarities that led to predictions that were already known. SMN2 is structurally similar to IN (Figure 2.3A, Appendix 1) and both SMN2 and IN are known to interact with SIP1 (Gemin2) (Fu et al., 2009; Mishra et al., 2006). SIP1, part of the large SMN complex involved in the assembly of snRNPs, may also be part of the pre-integration complex during HIV-1 infection and may

aid viral reverse transcription (Hamamoto, Nishitsuji, Amagasa, Kannagi, & Masuda, 2006). There are also several predicted interactions between IN and host proteins that interact with SMN2 that have not yet been tested (Appendix 1). The structural similarities shown in Figure 2.3B-D also led to predictions of known interactions, even though only parts of the proteins are structurally similar.

Protein co-localization

To further narrow the list of likely interactions, we refined these results by requiring both the HIV-1 protein and the target human protein to be present in the same location within the cell, based on GO cellular component (CC) annotation. The refined set of predictions is shown in Figure 2.4. Including this filtering step reduced the number of interaction predictions to 502, involving 189 HIV-similar proteins having 137 known different binding partners. There are 31 predictions corresponding to already known HIV-human interactions (Table 2.2, Appendix 4). Using the criterion that interacting proteins must have some evidence of co-localization not only reduced the size of the predicted interactome, but also increased the percentage of true positive predictions from ~3% true positives before filtering to over 6% after filtering (Table 2.2).

Taking localization into account, gp41 has many more predicted interactors than any other HIV-1 protein. This is most likely due to the relatively large number of GO cellular component terms that were annotated to gp41 and also relevant to the host cell. Since gp41 is found in more parts of the cell than other HIV-1 proteins, a larger number of human proteins were able to meet the co-localization criterion.

The interaction predictions made by this method are specific for structures, and we note that different structures for a single protein may lead to different predictions

about its interactions. Therefore, some information is lost if predictions are described at a gene level. Nevertheless, it may be of interest to consider interactions on a gene basis (See Appendix 5 for the mapping of HIV-1 IDs). When counted according to the HIV-1 protein names and human target Entrez Gene IDs, we made 883 interaction predictions, 56 of which were true positives according to HHPID and PIG. Following CC filtering, we had 22 true positive predictions among 265 total predictions (~10% of known true positives). While these results tend to suggest higher rates of predictive accuracy when using our method, we report our more conservative Uniprot-based accuracy values as our best estimates.

Properties of human proteins predicted to interact with HIV-1

Using the CC-filtered predictions, we next examined the function of human proteins predicted to interact with HIV-1 during infection. In this instance, we sought biological process and molecular function GO terms that were enriched among these target proteins. Examining the functions of the human proteins found in our filtered list of interactions, significant enrichment is observed in the processes of protein transport, nucleic acid transport, signaling, cell death, and post-translational modifications (Figure 2.5A); all of these processes are known to be manipulated or altered by HIV-1 during infection. During the course of the HIV-1 lifecycle, viral protein and nucleic acids must be transported from one part of the cell to another to ensure viral replication. The Pre-Initiation Complex, consisting of a number of viral and host proteins and the viral genome, must be transported from the site of viral entry to the nucleus for integration of the provirus. In addition, Env and Vpr are known to play both pro- and anti-apoptotic roles by manipulating host signaling. For instance, there is evidence that HIV-1 may

inhibit apoptosis in infected cells to prevent the cell from dying before the virus can replicate and assemble. On the other hand, HIV-1 can also promote apoptosis of immune cells using several pathways; the progressive destruction of CD4+ T cells is a well-known indication of AIDS (Gougeon, 2003).

Interestingly, all of the significantly enriched molecular function GO terms relate to GTP binding or hydrolysis (Figure 2.5B). GTPases are involved in a number of host processes that HIV-1 may take advantage of, including nuclear transport and cytoskeletal rearrangements that facilitate viral entry and cellular motility. Statins, a class of drugs that lowers cholesterol levels in the blood, have also been shown to inhibit HIV-1 infection by preventing viral fusion with the cell membrane through a mechanism that involves inhibition of Rho GTPases (Del Real et al., 2004). In addition, p115-RhoGEF inhibits HIV-1 gene expression through the activation of RhoA (L. Wang et al., 2000). Furthermore, both Rho and ROCK play a role in the cellular motility that allows HIV-1 infected monocytes to cross the blood-brain barrier, leading to HIV-1 encephalitis (Persidsky et al., 2006).

Actin microfilaments of the cytoskeleton are regulated by actin-binding proteins as well as Rho family small GTPases including Rho, Rac, and Cdc42 (Matarrese & Malorni, 2005). IN, RT, and gp41 were all predicted to interact with RhoA, Rac1, and Cdc42 (Figure 2.4). We found that gp41 has regions of structural similarity with many cytoskeleton related proteins, including erythrocytic spectrin alpha (SPTA1), erythrocytic spectrin beta (SPTB), alpha actinin 4 (ACTN4), alpha actinin 2 (ACTN2), moesin (MSN), Rho-associated protein kinase 1 (ROCK1), and arfaptin 2 (ARFIP2). IN resembles NCK adaptor proteins 1 and 2 (NCK1/2), dynactin 1 (DCTN1), and RAS

GTPase activating protein 1 (RASA1), among others. The cytoskeleton is manipulated by HIV-1 during virion fusion, assembly, and budding (Matarrese & Malorni, 2005). Drugs that cause depolymerization of microtubules and actin filaments can block HIV-1 movement through the cell. Actin has also been found within HIV-1 virions, and is incorporated through binding with NC (Wilk, Gowen, & Fuller, 1999). Thus, our predictions may aid further investigation into the ways in which HIV-1 manipulates the cytoskeleton.

By integrating a variety of high-quality functional data sets in the Literature Filter, we created a smaller interaction map that has the potential to provide a physical interaction context for a number of experimental findings. As an example, retroviral budding is known to involve members of the endosomal sorting complexes (ESCRTs). The ESCRT complexes normally induce the formation of multivesicular bodies in the endosome, but can be recruited to the plasma membrane by Gag to aid in viral budding. Many members of the ESCRT machinery appear in our results, including VPS4A, STAM2, EEA1, RAB5A, and TSG101 (Goff, 2007). Early endosomal autoantigen 1 (EEA1) is recruited to early endosomes by Rab5 and phosphatidylinositol 3-phosphate (Raiborg et al., 2001). Our results show that gp41 and Gag p2 may interact with RAB5A because they are structurally similar to EEA1 (Figure 2.4). EEA1 contains a FYVE domain and colocalizes with human hepatocyte growth factor regulated tyrosine kinase substrate (Hrs) protein (Deretic et al., 2004; Raiborg et al., 2001). Gp41 is also known to interact with AP1G2, an important component of clathrin-coated vesicles. AP1G2 interacts with RAB5A and provides further support for the possibility that gp41 interacts physically with RAB5A, but through a potentially different structural motif (Mattera,

Ritter, Sidhu, McPherson, & Bonifacino, 2004). The Gag p6 protein is a known mimic of Hrs, and like Hrs can recruit TSG101, which is required for the formation multivesicular bodies and viral budding (Pornillos et al., 2003). Gag p2 and a model of gp41 show structural similarity to the human protein CEP55, which acts during cytokinesis to recruit TSG101 to the thin membrane between the daughter cells, where it is needed for the final separation of two cells (Carlton & Martin-Serrano, 2007). Our results suggest that gp41, IN, and the p2 region of Gag may all be able to interact with TSG101 (Figure 2.4). Overall, interaction predictions are supported by a variety of studies implicating host mechanisms of vesicle formation in HIV-1 infection.

Additional method assessment

To further assess our predictions, we determined how many known interactions, curated within either HHPID or PIG, could have possibly been predicted using our method and the available data. First, in order for our approach to suggest a possible HIV-human interaction, the HIV protein must be represented among the crystal structures from PDB that are included in the Dali Database. In addition, any host factors predicted to interact with HIV-1 must have at least 1 known interaction with another human protein, and to be considered further, each of these must also have representative structures within PDB and Dali. Finally, in this work we included only those proteins that have been implicated in HIV-1 infection through RNAi studies or studies of the protein composition of the virion. Because we removed any human target proteins that did not pass the Literature Filter, we did not make predictions for human proteins not mentioned in previous studies.

A total of 319 known host-pathogen interactions satisfied these criteria. Sixty-two of these interactions (~19%) were predicted by our methodology, and are the set of predictions considered to be true positives (shown in Table 2.3). We also investigated how many of these possible interactions could have been found after using the cellular component filter, and determined that only 166 known interactions met the additional criterion of being annotated to the same cellular component. Within this set, our method found 31 of these (~19%). This result suggests that while considering cellular localization decreased the number of interactions considered, the number of true positive predictions did not improve. Obviously, without experimental validation we cannot determine whether the CC filter led to better prediction accuracy within the set of predictions not previously described in the literature or elsewhere. It is clear, however, that GO cellular component annotation is incomplete and the lack of shared annotation does not completely exclude the possibility that two proteins may interact; inclusion of the CC filter did double the percentage of true positives predicted when considering unknown potential interactions as well as those previously known.

As an additional form of assessment, we investigated how often we could expect to find previously known interactions by chance alone. Starting from proteins in HPRD, we found that ~0.17% of the known interactions could be found at random (see Methods). CC filtering of these random predictions gave a slight improvement with an average of 0.29% true positives (Table 2.4). Using only HPRD human target proteins that pass the Literature Filter increased the true positive accuracy of random predictions to 0.57%. This value can be compared to the 2.89% accuracy of our method, indicated in Table 2.2. When these random predictions were also run through the CC Filter, an

average of 1.03% true positives were found (Table 2.4) versus 6.18% when using our method (Table 2.2). Thus the Literature Filter and the CC Filter improved the accuracy of the true positive predictions individually, and to an even greater extent when combined. However, even with both filters, at best ~1% of the random predictions were found to be true positives, further indicating that incorporating structural information generates predictions with enhanced accuracy and biological validity.

Overlap with other studies

We also compared our predictions to those made by two previous computational studies predicting protein-protein interactions between HIV-1 and humans, namely the studies by Evans et al. and Tastan et al. (Evans et al., 2009; Tastan et al., 2009). Since these investigations reported their results in terms of genes, we compared them to our predictions as counted by gene, to find interactions predicted by multiple methods (Figure 2.6). We did not find a high degree of overlap between the predictions made by the various studies. This was not surprising, as even large-scale experimental protein interaction studies typically show little overlap in their results. Furthermore, the methodology used to generate the predictions differed significantly between studies. Our method used structural similarity to predict interactions, whereas Evans et al. looked for the presence of sequence motifs and counter domains and Tastan et al. integrated a variety of information, including information from GO, properties of the human interactome, and sequence motifs (Evans et al., 2009; Tastan et al., 2009). There are a greater total number of shared predictions between Evans et al. and Tastan et al. than between our results and either one of the others. This may be due to the fact that Tastan et al. incorporated Eukaryotic Linear Motifs and binding domains, the key predictor used in

the work of Evans et al., as one of the features used in their prediction method. In addition, the other two studies had a larger number of predictions overall. Approximately 7% of the predictions by Tastan et al. were found in the study by Evans et al. Approximately 5% of our predictions (Literature and CC filtered) were found by Evans et al. and 10% were shared with Tastan et al.

There were a few predictions that were shared between all methods. For our results before CC filtering, we found that there were 9 interactions predicted by all three methods (Figure 2.6A). Of these, four were determined to be true positives in our results: RT and MAPK1, gp41 and LCK, gp41 and PTPRC, and IN and PRKCH. The other five interactions (RT and PIN1, p2 and MAPK1, p2 and YWHAZ, gp41 and PLK1, gp41 and MAPK1, gp41 and CLTC, IN and XPO1, and IN and YWHAZ) are not known to occur, and may be good candidates for further investigation since they were predicted by three diverse methods. After we filtered our predictions by shared cellular components, three predictions were still common between all three studies, gp41 and LCK, gp41 and PLK1, IN and XPO1, one of which is a known interaction (Figure 2.6B). In summary, although few predictions were shared by all three studies, a large proportion of them are already known to occur, suggesting that the others may be worthy of high priority in future experimental efforts.

CONCLUSIONS

We have generated a map of potential protein-protein interactions between HIV-1 and its human host. The computational methodology used to create this map is based on the assumption that proteins with similar structures will share similar interaction partners. Thus HIV-1 proteins having a structure similar to one or more human proteins may potentially "plug in" to the host protein interactome at these points, providing the interface through which manipulation of downstream host processes can occur. From previous literature, many human proteins are known to play some role in HIV-1 infection. However, in most cases the nature of this role is unknown. Here, we provide specific predictions of how these human proteins may influence viral infection, namely by interacting with certain HIV-1 proteins.

In principle, our approach is applicable to any host-pathogen system with known protein structures. HIV-1 has a small proteome, with most of its protein structures at least partially determined. In addition, HIV-1 also has a large set of identified interactions that can be used for model validation. While few pathogens currently have such rich data sets, continued progress will help to remedy such deficiencies.

Identification of points of modulation between a host and pathogen requires multiple lines of evidence. Computational methods can help integrate these data, providing a promising avenue for the discovery of novel host pathogen interactions mediated by structural similarities and enhancing our understanding of functional relationships characterized through modern screening methods such as siRNA. Knowledge of the protein interaction network between the pathogen and human will not

only further our basic understanding of pathogen survival mechanisms, but may also provide clinical targets to combat infectious disease.

METHODS

Data Sources

We used the Dali Database for structure comparisons (downloaded in January 2009), and the Human Protein Reference Database (HPRD), HHPID, and PIG for protein interactions (downloaded February, July and June 2009, respectively) (Driscoll, Dyer, Murali, & Sobral, 2009; Fu et al., 2009; Holm et al., 2008; Holm & Sander, 1993; Mishra et al., 2006). The literature sources and various databases each have their own system of identifiers. PDB codes obtained from Dali were mapped to their corresponding taxonomy and Uniprot accessions using data from the SIFTS initiative (Berman et al., 2003; Tagari et al., 2006; Velankar et al., 2005). Other identifier mappings were carried out using DAVID Gene ID Conversion or Uniprot ID mapping (Bairoch et al., 2005; Dennis et al., 2003; Huang, Sherman, & Lempicki, 2009). Network diagrams were created in Cytoscape (Shannon et al., 2003). Images of protein structures were created in MacPyMol (DeLano, 2002).

Determination of structural similarity between HIV-1 and host proteins

We used the Dali database to ascertain structural similarity. Dali compares the 3D structural coordinates of two PDB entries by alignment of alpha carbon distance matrices, allowing for differences in domain order, and produces a structural similarity score (Berman et al., 2003; Holm et al., 2008; Holm & Sander, 1993). The Dali Database includes structural comparisons of proteins from PDB90, a subset of the PDB where no two proteins share more than 90% sequence similarity, were used as queries against the full PDB (Holm & Sander, 1998). For this study, we took into consideration all human proteins that were listed in the Dali database as being similar to an HIV-1 protein (NCBI

Taxonomy ID: 11676) and having a z score above 2.0, with the HIV-1 protein being either the query or the hit. We refer to these human proteins as "HIV-similar" proteins. No proteins of unknown structure were considered.

Interaction Prediction

We found known interactions between HIV-similar proteins and target human proteins, using data from the HPRD database, which contains literature curated interactions between pairs of human proteins (Mishra et al., 2006). For each HIV-similar protein, we predict that the target proteins, which are known to both interact with the HIV-similar protein and pass the Literature Filter, might also interact with the corresponding HIV-1 protein. Therefore, interactions between the HIV-similar and the human target proteins were mapped directly to the corresponding HIV protein.

Filtering

To reduce the number of predictions and add information from functional studies, predictions were filtered based on previous implication of the human protein in the HIV-1 infection process. One criterion was presence of the host protein in the HIV-1 virion. Host proteins known to be incorporated into or onto HIV-1 during budding were taken from several literature sources (Cantin et al., 2005; Chertova et al., 2006; Saphire, Gallay, & Bark, 2006). The presence of host proteins in or on HIV-1 may be a result of specific recruitment and serve a functional role, may result from localization of the protein near the site of budding, or may simply occur by chance. Predicted interactions between HIV-1 proteins and human proteins that are incorporated into the HIV-1 virion were retained. In addition, any human protein that is incorporated into the virion and is itself structurally similar to an HIV-1 protein was also included as a possible interaction.

Another filtering criterion was the host protein's essentiality for HIV-1 infection. Recently, several large-scale experiments using siRNA or shRNA knockdowns to identify host proteins involved in the HIV-1 life cycle have been published (Brass et al., 2008; König et al., 2008; Yeung et al., 2009; Zhou et al., 2008). The probe ids of the genes implicated by Yeung et al. were mapped to their Entrez Gene IDs using the appropriate Affymetrix annotation file (http://www.affymetrix.com/products_services/arrays/specific/hgu133plus.affx#1_4) (Yeung et al., 2009). This filter is referred to as the "Literature Filter." Host proteins that were implicated in at least one of these studies as having a possible role in HIV-1 infection or replication and are also known to interact with an HIV-similar protein were predicted to interact with an HIV-1 protein in the final predicted network.

To create a smaller and potentially more reliable list for further experimental validation, we filtered the predictions based on shared sub-cellular localization. The CC Filtered dataset contains interaction predictions where the two proteins share Gene Ontology (GO) cellular component annotation. Pairs of HIV-1 and human proteins predicted to interact were only included in this dataset if both proteins were annotated by DAVID as being present in the same cellular compartment (Dennis et al., 2003; Huang et al., 2009). Pairs with only the terms "cell" and "cell part" in common were excluded due to a large number of such pairs and the relative lack of specificity of these high level terms.

Validation of Predictions

Because within Dali there may be multiple PDB structures representing the same protein, there is some redundancy in the interaction predictions. In certain cases, multiple

PDB structures for the same HIV-1 protein were found to be similar to multiple PDB structures for an HIV-similar protein, leading to the same interaction predictions. Therefore, the predictions were counted as unique pairs of Uniprot accessions. In addition, for ease of viewing the predicted interactome, each node representing an HIV-1 protein is labeled with the protein name while each human protein is represented by Entrez Gene ID. To determine the correct mapping of PBD codes to HIV-1 proteins, the molecule name associated with each PDB chain was searched for keywords indicating the protein, with ambiguous cases treated on an individual basis. For example, PDB molecule names containing the word "capsid" but not "nucleocapsid" were assigned to the node "capsid." Furthermore, molecule names indicating polyproteins, such as those containing the phrase "gag-pol", were checked individually to determine which specific part of the polyprotein was represented by the entry. Two PDB structures were found to represent more than one mature HIV-1 protein: 1l6nA contains both capsid and matrix, while 1bajA contains capsid and p2 (Tang, Ndassa, & Summers, 2002; Worthylake, Wang, Yoo, Sundquist, & Hill, 1999); these structures are represented as "capsid, matrix" and "capsid, p2" respectively. When counting predictions at the gene level, we considered pairs of HIV-1 node names and human target Entrez Gene IDs.

To determine which predictions are true positives, PIG and HHPID entries for the predicted human interactors were examined to see if they contained the HIV-1 protein they were predicted to interact with (Driscoll et al., 2009; Fu et al., 2009). These interaction databases consist of PPIs curated from the literature. HHPID uses keywords to characterize the different types of interactions listed in this database. Since this work attempts to predict physical interactions, only entries with keywords representing direct

interactions were included (Tastan et al., 2009). The Uniprot accessions associated with the HIV-1 protein PDB entry, in the case of PIG, or the Entrez Gene ID mapped to that Uniprot accession, in the case of HHPID, were checked to see if it was present as an already known interaction of the human protein.

GO Term Enrichment

The Gene Ontology (GO) provides a system of terms to consistently describe and annotate gene products (Ashburner et al., 2000). GO term enrichment was performed using the DAVID Functional Annotation Chart tool. The GO is organized as a tree structure, with terms becoming more specific as distance from the root increases. Therefore, to avoid very general and uninformative GO terms, only those that are found at least 5 steps removed from the overall root of GO were considered. The p-values were corrected for multiple testing using the Bonferroni procedure and transformed by taking the $-log_{10}$ for easier visualization.

Computational evaluation

Two forms of computational validation of the method were conducted. As it is not possible to predict all known interactions due to lack of protein structures and other factors, we first determined the largest set of known interactions that it is theoretically possible to predict using our approach. To do this, we first determined the sets of all proteins that could be considered. This includes the set of all HIV-1 proteins in Dali (HIV set), the set of all human proteins that are represented in both Dali and HPRD (possible HIV-similar set), and the set of all human proteins in HPRD that are known to interact with at least one protein in the possible HIV-similar set as well as pass the literature filter (possible target set). Next, every pairwise combination of proteins in the HIV set and the

possible target set was checked to see if it represented a known interaction curated in HHPID or PIG. The resulting number of true interactions that could have been found by the method was compared to the number of true positives that were actually found, both before and after filtering by cellular components.

In the second approach, actual prediction results were compared to predictions based on randomly selected HIV-human protein pairs. The HIV-1 proteins were chosen from the 69 Uniprot accessions represented at least once by structures in the Dali Database. For human proteins, two different sets of human Uniprot proteins were created, one containing all the proteins in HPRD, and the other containing the subset of these human proteins that also passed the Literature Filter. The set of all human proteins in HPRD consisted of 8,582 proteins and was used to assess the accuracy of purely random predictions, while the second set of 830 proteins was used to observe the effect of the Literature Filter.

Since the structural similarity step was omitted, the predictions based on a human protein being similar to an HIV-1 protein and incorporated into the virion could not be simulated with the random selection procedure. We found that if we excluded this class of predictions from our real results, the number of unique predictions made was reduced to 2,139, but all 62 true positives were still included. Therefore, we randomly selected 2,139 pairs of HIV-1 proteins and human proteins from the entire HPRD, and a second set of 2,139 pairs of HIV-1 proteins and Literature Filtered human proteins for evaluation. Next, any known interactions between the randomly chosen pairs were found using HHPID and PIG. Additionally, both the unfiltered and Literature Filtered random predictions were then subjected to the CC Filter to gauge the improvement due to this

step of the method. The CC Filter reduced the number of predictions to a variable degree, depending on how many of the random predictions were annotated with the same GO cellular component term. The entire procedure was repeated 1,000 times. The mean and standard error of the mean for each of the four variously filtered random prediction sets was calculated using R. The distributions of random predictions after Literature Filtering were approximately normal, so one-sided single sample t-tests were performed to determine if the method performed significantly better than random. In addition, we performed Wilcoxon signed-rank tests that do not make assumptions about normality. When comparing our results to random predictions that had undergone the same filtering steps, either the Literature Filter or both the Literature and CC Filters, the p-values were less than 2.2e-16 for all statistical tests. In addition, even when performing the randomization procedure 10,000 times, none of the randomly selected interaction sets had a true positive rate higher than that observed in our results, suggesting a p-value of no greater than 0.0001.

To compare our predictions to those made by Evans et al. and Tastan et al., we found the intersection of the prediction sets, counted by HIV-1 protein name and human Entrez Gene ID (Evans et al., 2009; Tastan et al., 2009). Because each study used different names for the HIV-1 proteins, we had to map the naming schemes to each other to find common predictions. For example, Evans et al.'s "CA" and "GAG" and Tastan et al.'s "gag_capsid" and "gag_pr55" were mapped to our "capsid." Proteins for which we made no predictions, such as Rev, were not mapped to anything in our results, but were converted between Evans et al. and Tastan et al. to find overlap between these two studies.

FIGURES AND TABLES

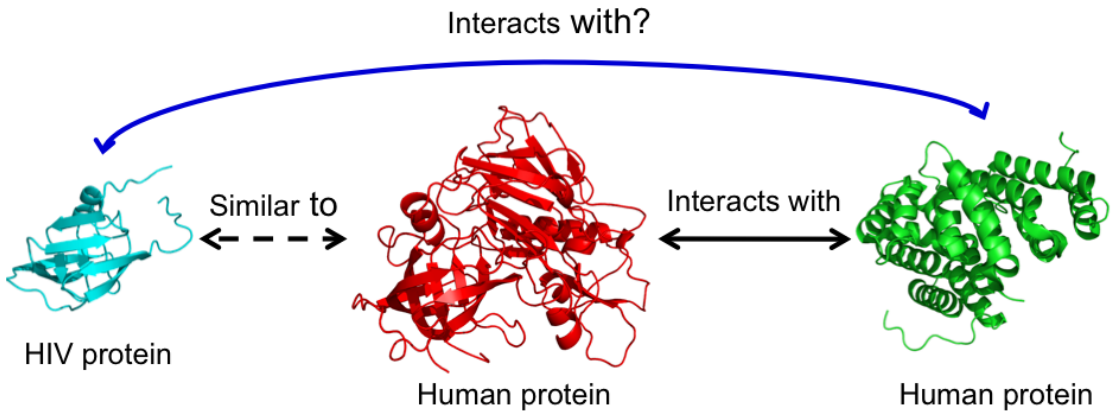


Figure 2.1. Diagram of approach. HIV-1 proteins showing structural similarity to one or more human proteins are first identified. Interactions for these "HIV-similar" proteins with other human proteins are then identified. Following appropriate filtering, this methodology predicts the existence of a physical interaction between the HIV protein and the human "target" protein(s).

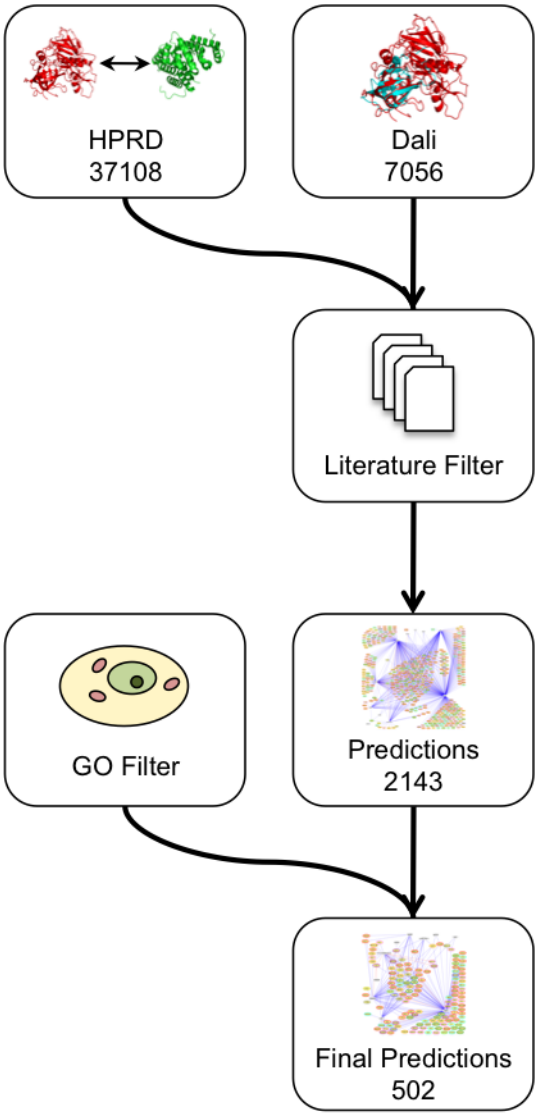


Figure 2.2. Structural prediction workflow. Structural similarities from Dali and known interactions between human proteins from HPRD are used to predict interactions between HIV-1 and human proteins. These predictions are filtered based on functional information from previous studies to make a first set of predictions. This set is further filtered using GO cellular component terms to yield a final prediction set including fewer predictions with higher confidence. Numbers represent the number of interactions, or structural similarities in the case of Dali, at each stage in the process.

Table 2.1. Coverage of HIV-1 proteins.

HIV-1 protein	PDB chains in Dali	PDB structures in Dali
Capsid	52	25
gp120	24	20
gp41	24	17
Integrase	51	26
Matrix	17	12
Nef	5	3
Nucleocapsid	3	3
p2	1	1
Protease	604	304
Reverse Transcriptase	176	85
Tat	3	3
Vpr	1	1
Vpu	**1**	**1**

The number of structures representing each HIV-1 protein in Dali.

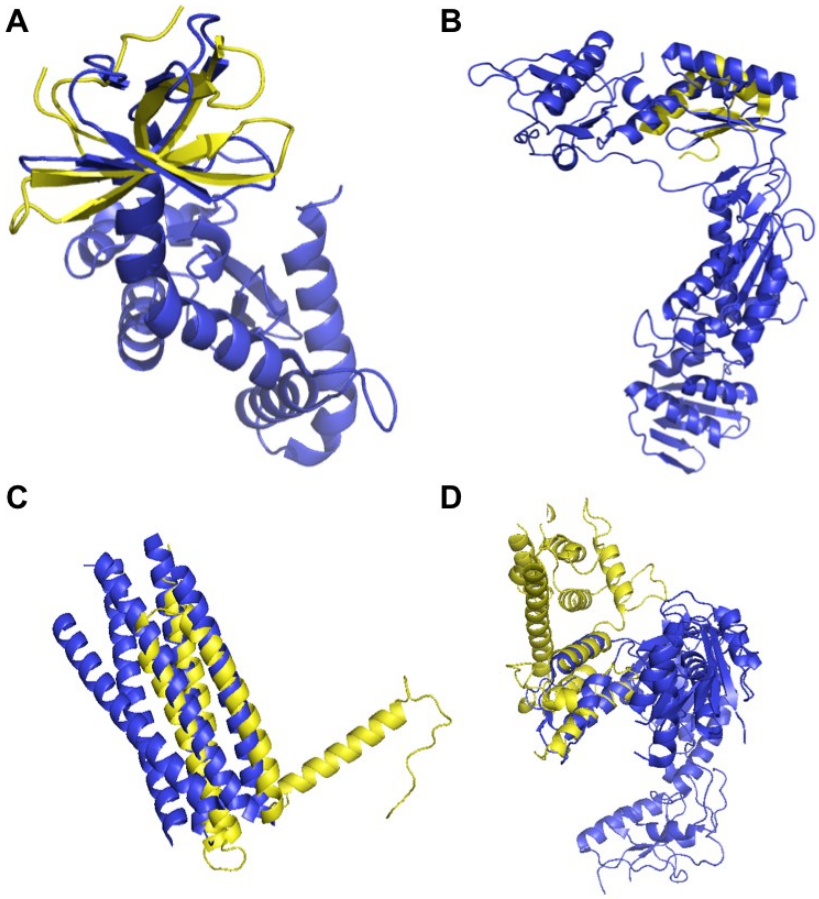


Figure 2.3. Selected structural similarities. Structures of HV-1 and human proteins aligned using Dali. (A) IN (1ex4A) aligned with SMN2 (1g5vA) (J. C. Chen et al., 2000; Selenko et al., 2001). (B) NXF1 (1ft8E) aligned with RT (1tl3A) (Hopkins et al., 2004; Liker, Fernandez, Izaurralde, & Conti, 2000) . (C) gp41 (2cmrA) aligned with PTK2 (1k04A) (Arold, Hoellerer, & Noble, 2002; Luftig et al., 2006). (D) RT (1lwcA) aligned with PLEC1 (1mb8A) (Chamberlain et al., 2002; Garcia-Alvarez, Bobkov, Sonnenberg, & de Pereda, 2003). HIV-1 proteins are in blue, human proteins are in yellow.

Table 2.2. Summary of predicted interactions.

	Before CC Filter	After CC filter
Structure Nodes	11	10
HIV-1 Uniprot	49	33
Similar Human Proteins	347	189
Predicted Human Binding Partners	406	137
True Positives	62	31
Total Predictions	2143	502
Percent True Positive	2.89%	6.18%

The number of proteins found and interactions predicted are shown. HIV-1 Structure Nodes refers to the number of HIV-1 proteins represented in Dali, while HIV-1 Uniprot refers to the number of HIV-1 Uniprot accessions present in the predictions. Human proteins and predicted interactions are counted by unique Uniprot accessions.

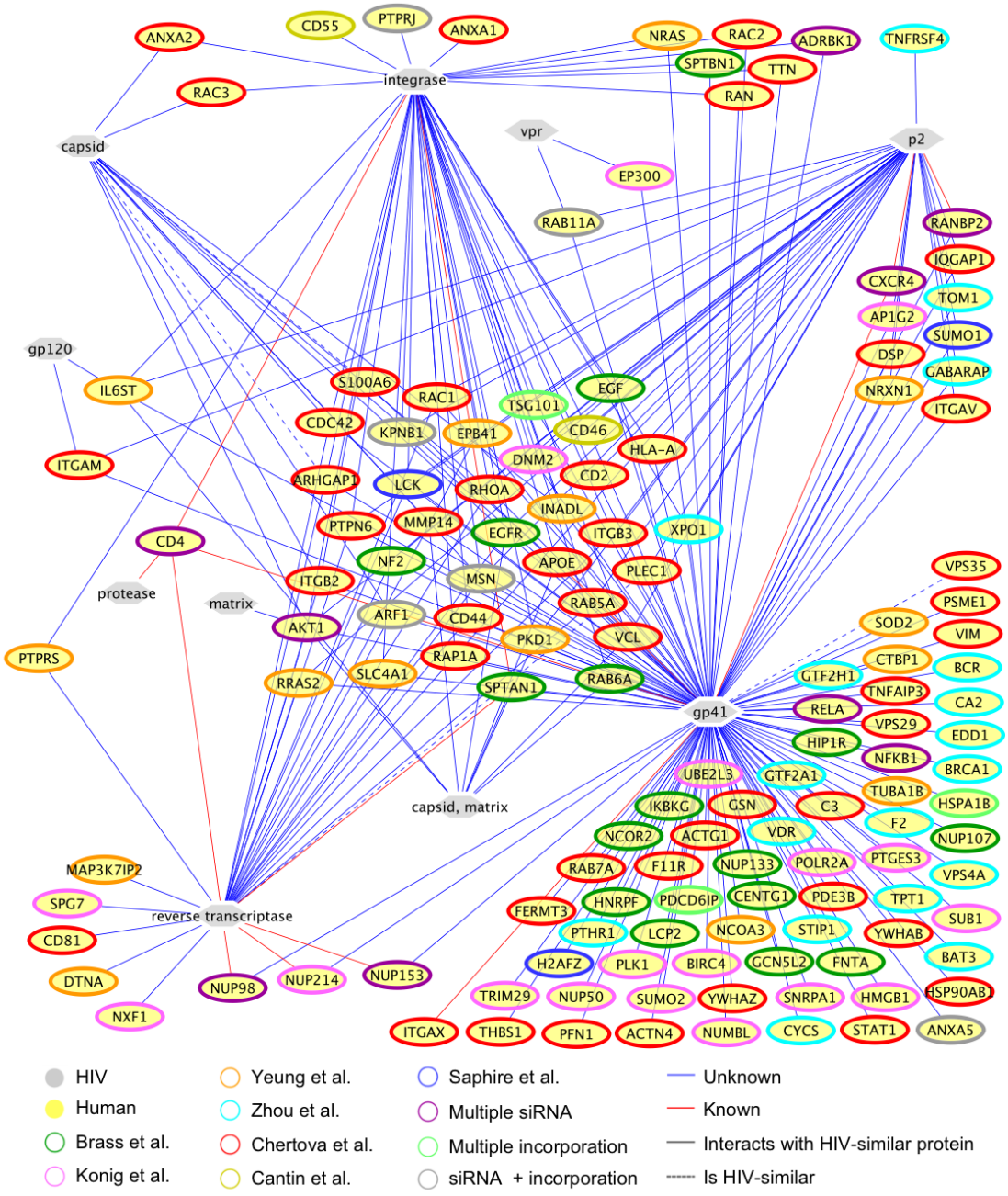


Figure 2.4. Predicted interaction network after cellular component filtering. Predictions were filtered to contain only those pairs of proteins that share at least one Gene Ontology cellular component term. In addition to the prediction of a physical interaction, the human proteins included in this prediction set are known to have a role in HIV-1 infection or replication as supported by 1) evidence of incorporation into the HIV-1 virion or 2) their reduced expression is known to prevent HIV-1 infection (node outline color corresponds to source). Red lines represent predicted interactions that are already known to occur.

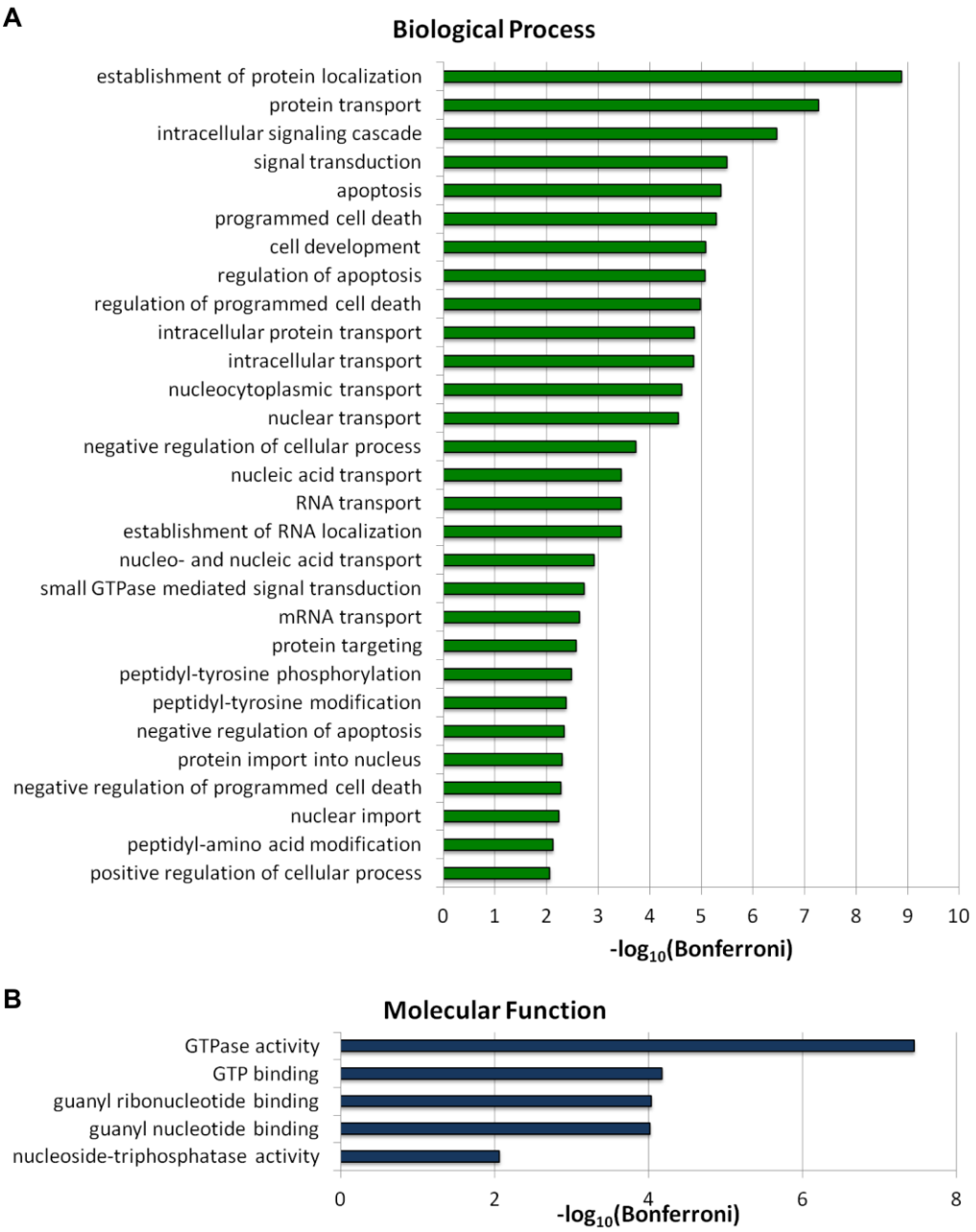


Figure 2.5. Significantly enriched Gene Ontology terms in the human-HIV-1 interaction network. GO terms removed at least 5 levels from the root for (A) Biological process and (B) Molecular function. Bonferroni corrected p-values (α = 0.01) were $-log_{10}$ transformed. "nucleo- and nucleic acid transport" is an abbreviation for "nucleobase, nucleoside, nucleotide and nucleic acid transport."

Table 2.3. Method evaluation.

	Before CC	After CC
Predicted True Positives	62	31
Possible True Interactions	319	166
Percent Found	19.44%	18.67%

Comparison of the number of known interactions predicted with the number of known interactions that could have theoretically been found using the available data.

Table 2.4. Accuracy of random predictions.

Filtering Method	Mean Accuracy	Standard Error
None	0.166%	2.79e-3%
CC	0.286%	6.09e-3%
Lit	0.567%	4.84e-3%
Lit and CC	1.030%	1.07e-2%

The mean percent of true positives and standard error of the mean for 1000 random predictions without any filtering (None), CC filtering alone (CC), Literature Filtering alone (Lit), and both Literature and CC Filtering (Lit and CC) are shown.

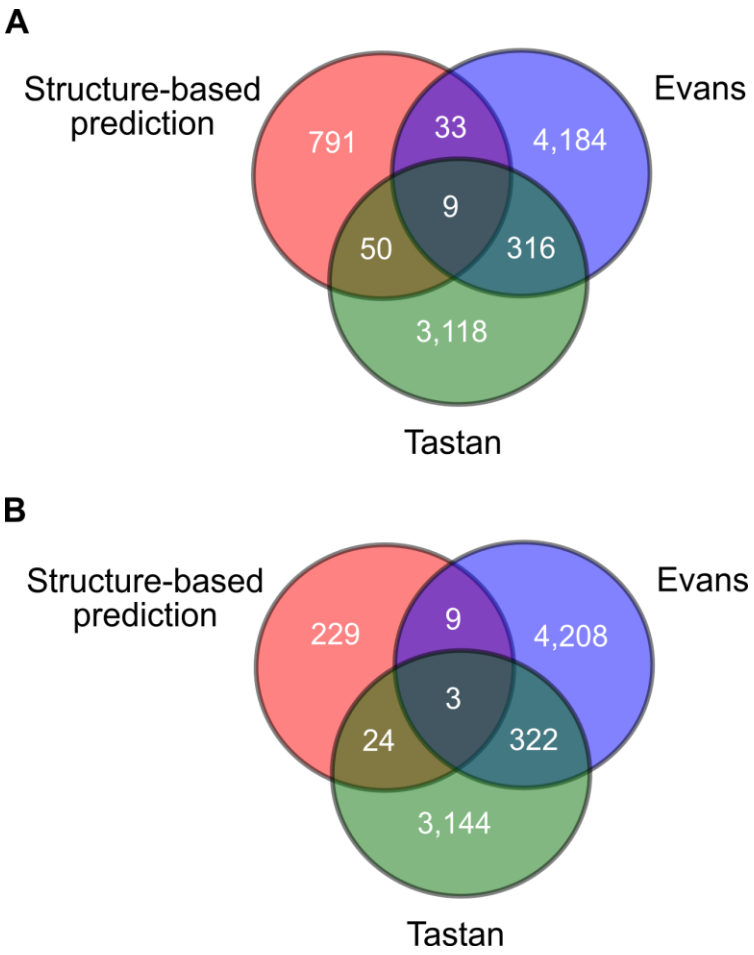


Figure 2.6. Overlap with previous studies. Venn diagrams of the overlap between our results those of previous computational studies by Evans et al. and Tastan et al. (A) with Literature Filter and (B) with Literature and CC Filters (Evans et al., 2009; Tastan et al., 2009).

CHAPTER THREE

MAPPING PROTEIN INTERACTIONS BETWEEN DENGUE VIRUS AND ITS HUMAN AND INSECT HOSTS

INTRODUCTION

With over 50 million cases per year, dengue virus (DENV) is a significant and growing threat to worldwide human health. Widespread among tropical and sub-tropical regions, this NIAID Category A pathogen consists of four serotypes, DENV1 to DENV4, and is a member of the family Flaviviridae (NIAID, 2012). DENV causes a range of diseases in humans, from the mild Dengue Fever (DF) to the more deadly Dengue Hemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS). While the average number of cases reported to the WHO and the number of countries reporting cases of DENV have increased dramatically in the past five decades, relatively little is known about this important tropical pathogen that still lacks a vaccine, specific drug treatment, and relevant animal model (Kroeger & Nathan, 2006).

DENV is and arbovirus, carried and spread to humans primarily by the mosquito vector *Aedes aegypti* and to a lesser extent *Aedes albopictus*. Thus, DENV displays the remarkable capability to survive and replicate in two very different host organisms, accomplished by a genome encoding a mere 10 proteins (Perera & Kuhn, 2008). To be successful, DENV must be able to manipulate each of its hosts at a molecular level. This manipulation must be accomplished, in part, through specific protein-protein interactions

that allow the virus to bend existing host cellular systems to the purpose of furthering the viral lifecycle. However, understanding this host-pathogen system is particularly difficult given the complexities of host-virus dynamics as well as the lack of a useful animal model system. In light of these challenges, computational approaches provide an important tool in studies of host-pathogen systems. In particular, computational approaches for predicting host-pathogen protein interactions provide opportunities for identifying specific targets for further experimental work, understanding system behavior, and determining plausible therapeutic candidates. Despite their potential value, such computational approaches have not been widely applied to the problem of predicting host-pathogen interactions. In particular, we are not aware of any studies focused on computational large-scale prediction of protein-protein interactions between DENV and humans and know of only one recent study on *Aedes* (Guo et al., 2010).

Here, we establish a network of predicted interactions between DENV proteins and proteins from its human and insect hosts. These predictions are based on protein structural similarity, where we first determine structural similarities between pathogen and host proteins using an established method for comparing 3D structures. We refer to host proteins having a region of high structural similarity to a DENV protein as "DENV-similar." Next, we identify known intra-species interactions for these DENV-similar proteins, and refer to the one or more host proteins that they interact with as "targets." We then assume that the similar structural features seen between DENV proteins and their host DENV-similar counterparts allow the DENV protein to participate in the same interactions as DENV-similar proteins; joining the host protein network at these points (Figure 3.1). We prioritize the interaction map using data from recent RNAi screens, to

create a smaller network of interactions having the greatest potential to be correct. These predictions include numerous novel interactions with potential functional relevance and we highlight predictions relevant to stress, the Unfolded Protein Response (UPR), and the interferon pathway. This computational network approach provides an additional tool for the investigation of poorly-characterized host-pathogen systems such as DENV, as well as helping to identify potential targets in both hosts that may be used in future DENV vaccination, treatment, and control efforts.

RESULTS AND DISCUSSION

Identification of Dengue-Similar Host Proteins

To develop a network of interactions between DENV and its hosts, *H. sapiens* and *A. aegypti*, we employed a method we developed previously in the prediction of protein interactions between HIV and human (Doolittle & Gomez, 2010) (see Methods for further details). First, we obtained 3D structures for the DENV proteins, from two sources. Experimentally determined structures were taken from the PDB and consist of 31 PDB entries representing the DENV2 proteins E, pr peptide, prM, C, NS2B, NS3, and NS5 (Berman et al., 2003). Because there are no experimentally determined structures for NS1, NS2A, NS4A, and NS4B, we used the I-TASSER server to predict the structure of these proteins (Y. Zhang, 2008). In this way, we investigated possible interactions for every DENV protein.

To determine structurally similar host proteins, we used DaliLite to compare DENV structures against every other structure in the PDB (Berman et al., 2003; Holm et al., 2008; Holm & Sander, 1993). We considered only significant structural matches with proteins from DENV's hosts. We found 300 human proteins with similarity to a DENV protein (hDENV-similar). However, we found no similarities between DENV proteins and *A. aegypti* proteins. This is not surprising, given that there are currently only 17 structures from *A. aegypti* in the PDB. Therefore, we looked for similarities between DENV proteins and the fly, *Drosophila melanogaster*, and found 15 proteins with structural similarity to DENV, which were then used as dDENV-similar proteins in downstream analyses.

Known DENV-Host Interactions

A particular challenge in host-pathogen studies is the general lack of interaction data. HIV is perhaps the most well characterized virus in this regard, with over 800 direct interactions documented in NCBI's HIV-Human protein interaction database (over 2,500 interactions if indirect interactions are included) (Fu et al., 2009). In contrast, a recent compilation of host-pathogen interactions from public databases describes a total of 3 DENV-human interactions (Dyer, Murali, & Sobral, 2008).

Through a more comprehensive search of the literature, we have found 20 documented interactions between DENV and human proteins (Table 3.1). Almost half of the documented protein interactions involve E protein and a receptor on the cell surface. Two of these, CD14 and HSPA5, have been shown to function as DENV receptors, although their binary interaction with E protein was not explicitly demonstrated; it may be that some other protein in complex with these receptors is the direct interaction partner of E protein (Y.-C. Chen, Wang, & King, 1999; Jindadamrongwech, Thepparit, & Smith, 2004). Furthermore, there is evidence that DENV receptor usage may be virus strain and cell type specific (Bielefeldt-Ohmann, Meyer, Fitzpatrick, & Mackenzie, 2001). Indeed, RPSA has been shown to be a DENV receptor, suggesting an interaction with E, but only for DENV1 (Thepparit & Smith, 2004). Because our predictions were focused on DENV2, this interaction was not considered for our predictions, but was included in Table 3.1 for completeness. Interactions not shown to be specific for a different serotype were included in our list of true positive interactions. Therefore, a total of 19 protein interactions were considered as known host-pathogen interactions between DENV2 and human.

There are no well-characterized protein-protein interactions between DENV and *A. aegypti*. However, in the C6/36 cell line from *A. albopictus*, tubulin is believed to interact with DENV2 E protein (Chee & AbuBakar, 2004). In addition, one protein, likely to be HSP90, has been put forward as a putative receptor for DENV2 in *A. aegypti*, having been shown to bind to the E protein (Salas-Benito et al., 2007). However, its identity has not been conclusively demonstrated. In addition, mosquito La auto-antigen is known to interact with the 3' end of DENV RNA and may play some role in RNA synthesis (Yocupicio-Monroy, Padmanabhan, Medina, & del Angel, 2007). Human La auto-antigen (SSB) is also known to interact with the ends of the viral RNA, as well as NS3 and NS5 (García-Montalvo, Medina, & del Angel, 2004). If the functions of the mosquito and human La proteins in DENV infection are similarly conserved, mosquito La may interact with NS3 and NS5 as well, although this has not been shown. It is likely that some of the protein interactions that enable DENV to manipulate the cellular pathways of two hosts are conserved between the species.

Prediction of Protein Interactions

After determining which host proteins are structurally similar to DENV proteins, we inquired into the known protein-protein interactions that each DENV-similar protein participates in. For the hDENV-similar set, we obtained known human protein interactions from the Human Protein Reference Database (HPRD), which consists of over 37,000 interactions found in the literature (Mishra et al., 2006). We predicted that DENV proteins could interact with the partners of their corresponding hDENV-similar proteins, under the hypothesis that proteins with highly similar structures are likely to be involved in similar protein interactions (Figure 3.1A). We predicted 4,273 potential host-pathogen

interactions, involving 2,321 different human proteins (Table 3.2). Of the 19 known protein-protein interactions between DENV and human, 9 are present among our predictions (Chang et al., 2001; Y.-C. Chen et al., 1999; Chiu, Shih, Yang, & Yang, 2007; J. J. E. Chua, Ng, & Chow, 2004; J. J.-E. Chua, Bhuvanakantham, Chow, & Ng, 2005; Ellencrona, Syed, & Johansson, 2009; García-Montalvo et al., 2004; Jiang, Yao, Duan, Lu, & Liu, 2009; Jindadamrongwech et al., 2004; Kurosu, Chaichana, Yamate, Anantapreecha, & Ikuta, 2007; Limjindaporn et al., 2007, 2009; Lozach et al., 2005; Noisakran et al., 2008; Reyes-Del Valle, Chávez-Salinas, Medina, & Del Angel, 2005). This method may not predict all interactions, for example those mediated by sequence motifs rather than structural features. A table of some DENV-human protein interaction predictions is provided in Appendix 6.

For the dDENV-similar proteins, we used the interactions curated in IntAct for *D. melanogaster* and potential *D. melanogaster* interactions suggested by the yeast-2-hybrid data sets in DroID (Aranda et al., 2010; J. Yu, Pacifico, Liu, & Finley, 2008). However, rather than making direct predictions using these interactions, as we did for human proteins, we determined orthologs of the *D. melanogaster* proteins in *A. aegypti*, since this is the true host of DENV. We then predict that the *A. aegypti* ortholog of a *D. melanogaster* protein that interacts with a dDENV-similar protein may also interact with the corresponding DENV protein (Figure 3.1B). As a result of this procedure, we predict that 158 *A. aegypti* proteins participate in 176 interactions with DENV proteins (Table 3.2). We note that this method did not predict interactions between E and mosquito tubulin, HSP90, or La, which have been suggested as possible interactions (Chee & AbuBakar, 2004; García-Montalvo et al., 2004; Yocupicio-Monroy et al., 2007).

However, 12 of the proteins predicted to interact with DENV were orthologs of proteins involved in DENV2 infection in humans or fly. A table of some DENV-*Aedes* protein interaction predictions is provided in Appendix 7.

GO term enrichment

Due to the sparseness of known interactions to which we can compare and evaluate our predictions, we examined the functional roles of host proteins for patterns relevant to DENV infection. To this end, we determined the GO term enrichment among the DENV-similar and target proteins for each host. We find that many of the most significantly enriched terms among these sets of proteins are for processes or functions known to be important for DENV infection (Figures 3.2 and 3.3). Our results are also consistent with a study of altered protein expression during DENV infection in which several of the proteins identified have functions related to the GO terms RNA processing, transcription, or regulation of stress response, which were enriched in our predictions (Pattanakitsakul et al., 2007).

For human target proteins (dark blue bars in Figure 3.2), processes involving signaling, cell death and apoptosis, and positive or negative regulation terms are much more frequent than would be expected if chosen at random. These processes are in agreement with processes and pathways that are altered during the course of infection. Human proteins with structural similarities to DENV (red bars) are enriched for terms describing blood coagulation and hemostasis, indicating that DENV proteins appear to have structural similarities that mimic human proteins involved in the pathways controlling the cessation of bleeding. One of the defining symptoms of DHF is hemorrhage, the pathogenesis of which has been shown to include abnormalities in levels

of cytokines, complement components, and coagulation factors (Srichaikul & Nimmannitya, 2000; Van Gorp et al., 2001). DENV proteins show structural similarity with proteins from all three of these categories. In fact, DENV mimicry of clotting factors has already been observed: antibodies against the DENV2 protein NS1 have been shown to cross-react with clotting factors and integrins on thrombocytes and endothelial cells (Falconar, 1997). Kinase, cytokine, phosphotransferase, and peptidase functional roles are also enriched in hDENV-similar and human target proteins.

The most enriched biological processes for DENV-*A. aegypti* within our interaction predictions appear to revolve around RNA processing and transcription-related processes. This is observed in enriched functions, which include "nucleotidyltransferase activity" and "sequence-specific DNA binding" GO terms (Figure 3.3). Gene expression analyses within *A. aegypti* in response to DENV infection have shown changes in several hundred genes across a range of functions, with immune and transcriptional processes being highly represented (Xi, Ramirez, & Dimopoulos, 2008). We note that one protein with mRNA processing activity, AAEL013723-PA or polypyrimidine tract binding protein, is an ortholog of human PTBP1, which is known to interact with NS4A (Jiang et al., 2009). However, we predict AAEL013723-PA interacts with NS2A in *A. aegypti* and make no predictions about PTBP1 interacting with DENV in humans (Appendix 7). In addition, there are several lines of evidence suggesting that *A. aegypti* may use the RNAi pathway as a defense against infection by DENV, and variations in the RNAi pathway in both the host and viral strains may contribute to differences in the efficiency of infection (Sanchez-Vargas et al., 2004).

Literature filtering

Our prediction of host-DENV interactions using protein structural similarity allows us to generate a list of candidate interactions with potentially significant functional relevance, forming a possible basis for further experimental and computational studies of this system. Given the large number of predictions and the scarcity of known interaction information, we wished to incorporate additional data to help refine these predictions into an "increased-confidence" set of interactions. To do this, we incorporated functional information from recent literature. Specifically, two siRNA screens have recently examined the roles of host proteins in DENV infection (Krishnan et al., 2008; Sessions et al., 2009). Sessions et al. performed a genome-wide siRNA screen for *D. melanogaster* proteins whose depletion affected the ability of DENV2 to infect the host cells. For those insect host factors with human homologues, they verified 55 as human host factors, also using siRNA (Sessions et al., 2009). In addition, Krishnan et al. identified 123 human host factors for DENV in a study primarily searching for West Nile Virus host factors, but also testing these proteins in DENV infection (Krishnan et al., 2008). In total, these two studies implicated 173 human proteins and 116 *D. melanogaster* proteins as playing a role in DENV infection. We note that similar studies in HIV together have revealed nearly 1,000 such human host factors, and thus it is unlikely that these studies have identified all such factors in DENV hosts.

While host factors may act through mechanisms other than direct protein interactions, their functional involvement in DENV infection makes them more likely to participate in host-pathogen protein interactions. Comparing the identified host factors against our predictions, we found that 48 of the human target proteins were also

identified as host factors in the siRNA screens, as well as 3 of the *D. melanogaster* targets (Figure 3.4 and Table 2).

Many host proteins are predicted to interact with both NS2A and NS4B. This is due to apparent structural similarities between these two DENV proteins; several of the hDENV-similar proteins for NS2A and NS4B are the same. In particular, many predictions for both proteins are based on regions of structural similarity to members of the 14-3-3 family, protein phosphatase 2 regulatory subunits, and beta-catenin. These hDENV-similar proteins are key signaling proteins, with many known interactions. 14-3-3 proteins have been found to interact with over 200 polypeptides involved in highly diverse cellular functions (Mhawech, 2005) and PP2 is a serine-threonine phosphatase and tumor suppressor (Janssens, Goris, & Van Hoof, 2005). Beta-catenin is a well-known member of the Wnt signaling pathway. The other two DENV proteins with no known structure, and hence have predictions based on modeled structures, do not show structural similarity to 14-3-3 or protein phosphatase proteins. The combination of a predicted interaction along with siRNA functional relevance suggests that these host-DENV relationships would form the basis of a high-priority candidate list for future investigation.

Subcellular co-localization

As an additional method of identifying the most highly supported predictions, we filtered predictions based on subcellular co-localization. The obvious assumption here is that for two proteins to directly interact, they must be physically present together within the same cellular compartment. Therefore, we used shared GO cellular component (CC) annotation to filter our predicted interactions. While this filtering should highlight a

smaller set of proteins having additional evidence of interaction, we note that CC annotation is very noisy, with the localization of proteins within a cell being often poorly characterized for even well-studied model organisms, let alone species such as DENV or *A. aegypti*.

As GO annotations taken from DAVID for DENV are linked to entries for the DENV polyprotein rather than the individual proteins, all DENV proteins receive the same annotations. To ameliorate this problem and to include proteins with modeled structures, we assigned GO terms to the sequence corresponding to each DENV protein structure using a different tool, GOanna, which finds possible annotations based on the annotations of highly significant BLAST hits (McCarthy et al., 2006). By assigning GO terms in this way, we were able to find CC terms for all of the DENV proteins.

We used this information to create a smaller list of predicted interactions, containing only those predictions where the DENV protein and the host target share at least one GO CC term. After CC filtering, there were 2,073 predicted interactions between DENV and 1,099 human proteins. Seven of the 19 known DENV-human interactions remained after filtering (reduced from 9 of 19), as well as 20 interactions involving host factors (Table 3.2). The two known interactions that were removed during CC filtering were UBE2I and HSP90AA1 interacting with E. The CC filter reduced both the total number of predictions and the ones containing host factors by about half, leaving many fewer predictions with only a small decrease in predictions already having known functional support. For the predicted interactions in *A. aegypti*, the reduction in predictions was more pronounced, with only 17 predictions and 12 targets passing the CC filter (Figure 3.5). One of these predictions involves a host factor, as compared to 12 of

the 176 predictions before filtering. While GO compartment information is incomplete, CC filtering does provide a smaller list of predictions, and in the case of human-DENV interactions, with only a slight decrease in the ability to predict known interactions. Full tables of interactions incorporating CC filtering are provided in Appendices 8 (human) and 9 (*A. aegypti*).

Prediction accuracy

It is difficult to judge the accuracy of predictions for protein interactions in host-pathogen systems, especially for those pathogens such as DENV that have received less attention than their worldwide burden deserves. However, our results include essentially half (47%) of the known interactions between DENV and human proteins, as well as a significant number of proteins from both human and fly that have been suggested to play a functional role in DENV infection (Chang et al., 2001; Chiu et al., 2007; J. J. E. Chua et al., 2004; Krishnan et al., 2008; Reyes-Del Valle et al., 2005; Sessions et al., 2009). Protein interaction prediction is a difficult problem, with even a slight improvement over random guessing considered a success. In a recent attempt to predict protein interactions in yeast, the organism with the most complete known interactome, the best method tested was able to predict 60% of the known interactions, outperforming several previous methods having accuracies as low as 38% (J. Wang, Li, Wang, & Wang, 2009). Another study evaluating the performance of three prediction methods based on protein domains showed that all three performed only slightly better than random guessing on benchmark datasets (Ta & Holm, 2009). A large set of known interactions that could serve as a gold standard for humans and DENV is lacking, but based on the small number of DENV2-human interactions that are known, we estimate that the accuracy of our method is

comparable to that of other methods applied to better-characterized systems such as yeast. Certainly there are many protein-protein interactions and host factors that remain to be discovered. As previously mentioned, HIV, another virus which encodes only a handful of proteins, is known to participate in over 800 direct interactions with human proteins, with over 2,500 interactions if ones that may be indirect are included (Fu et al., 2009). It seems reasonable to assume that DENV2 also participates in a large number of host-pathogen interactions, most of which are currently unknown.

We have also applied this methodology to the prediction of protein interactions between HIV and human, a host-pathogen system for which much is known, although the same problems still exist to some extent (See Chapter 2) (Doolittle & Gomez, 2010). In that work, we found that when we consolidate proteins to single genes, at least 10% of our predictions were estimated to be correct when comparing predictions involving HIV host factors that passed CC filtering to known HIV-human protein interactions. This represented a significant improvement over random predictions (1% of random predictions were correct). For the DENV-host systems described in this work, we similarly estimate that, at a minimum, approximately 10% of predictions provided here are correct. It is encouraging that, comparable to yeast studies, we were able to find approximately 50% of the currently known interactions between DENV and human.

Comparison to another dataset

As discussed earlier, the major challenge in predicting protein interactions for pathogens is the lack of interaction data. A further complication in this case is that little is known about protein interactions within the mosquito vector itself. Very recently, a draft of the *A. aegypti* interactome was predicted by assuming that the mosquito orthologs of

interacting proteins in *C. elegans*, *D. melanogaster*, and *S. cerevisiae* would also interact (Guo et al., 2010). Because so little is known about the insect vector, we felt it would be valuable to compare our results to another set of predictions made using our general method, but based on this predicted mosquito interactome. To do this, we found structural similarity between DENV2 proteins and fly, worm, or yeast proteins, mapped these similar proteins to their orthologs in mosquito, and then used the mosquito interactome to predict host-pathogen interactions. In comparison, we previously used only fly proteins that were similar to DENV2 proteins, linked them to interacting proteins from experimentally determined fly interactions, then found the mosquito orthologs to map our predictions.

When we expanded our search for DENV-similar proteins to include all three model organisms, we found 45 proteins (vs. 15 in our original set) that showed structural similarities to DENV2 proteins. We then mapped these proteins to orthologs in *A. aegypti*. Any *A. aegypti* proteins predicted to interact with these orthologs, according to the recently published mosquito interactome, were predicted to interact with DENV2. As a result, we found 263 mosquito targets participating in 351 interactions with DENV2 proteins (Appendix 10). This can be compared to the 158 targets and 176 interaction predictions generated by considering *Drosophila* interactions alone (Table 3.2). The larger number of interactions predicted is a result of including more protein structures from additional species. Overall, the predictions based on the new data showed similar GO term enrichment as compared to our original results (Figure 3.6).

It is difficult to directly assess prediction quality due to the absence of known interactions between DENV2 and mosquito. However, we note that despite the larger

number of predictions, fewer involved proteins that were suggested to be host factors (10 compared to 12 in the original predictions). This suggests that the larger number of predictions may represent an increase in the number of false positive predictions. Indeed, some of the within mosquito interactions used to produce this larger set of interactions may also be false positives, as the interactions of (Guo et al., 2010) are predicted from interactions in other species. The evolutionary distances between *A. aegypti*, worm and yeast may be too great to provide accurate predictions of protein interactions using orthologs. A network of the fourteen predictions made using both datasets is given in Figure 3.7.

In addition, Guo et al. predicted 22 interactions between DENV2 and *A. aegypti* by finding orthologs of proteins from any host species shown to interact with proteins from any flavivirus (Guo et al., 2010). Only 3 of the *A. aegypti* proteins they predict to interact with DENV2 (AAEL012515, AAEL014959, and AAEL013600) are present among the predictions made with our method using Drosophila experimental interactions alone. In addition, we predict them to interact with different DENV2 proteins than what is predicted in (Guo et al., 2010).

Stress and Apoptosis in DENV Pathogenesis

We note several links between DENV pathogenesis, stress responses, and apoptosis among our predictions and in the literature. The GO term "regulation of stress response" is enriched among hDENV-similar proteins, as well as several terms related to apoptosis among the human target proteins. Several potential DENV receptors are involved in stress responses, such as HSP90, HSPA4, and HSPA5 (Jindadamrongwech et al., 2004; Reyes-Del Valle et al., 2005). In particular, flaviviruses assemble within and

bud from the ER, and are known to induce the Unfolded Protein Response (UPR), which reacts to stressors of ER function. As the UPR is necessary for cell survival during infection, but also has a negative impact on viral replication, modulation of this response by DENV may be advantageous. The UPR can induce either survival or apoptosis signals depending on the strength and duration of the ER stressor (Liu & Kaufman, 2003). Three major branches, running through PERK, ATF6 and IRE1, regulate the UPR and ER homeostasis, and all three have been shown to be induced by DENV infection (Figure 3.8) (Umareddy et al., 2007).

Activation of PKR-like ER kinase (PERK), leads to phosphorylation of the eukaryotic initiation factor 2a (eIF2a). This phosphorylation inhibits the formation of translation initiation complexes, leading to translation inhibition and a reduction in the number of unfolded proteins within the ER (Liu & Kaufman, 2003). Production of ATF4 is also enhanced as a result of eIF2a phosphorylation, eventually leading to GADD34 production. GADD34 acts within a negative feedback loop, recruiting protein phosphatase 1, leading to dephosphorylation eIF2a and the restoration of normal translation efficiency. Persistent ER stress leads to CHOP expression and promotion of apoptosis. It has been suggested that DENV may be able to compensate the UPR response by inducing dephosphorylation of eIF2a to restore translation (Umareddy et al., 2007). We predict interactions between NS4B and GADD34 (PPP1R15A) (Figure 3.8). A listing of all predictions, with or without filtering, is provided in Appendices 6, 7, 8, and 9.

ATF6 is a bZIP family transcription factor that transits from the ER to the Golgi in response to ER stress. It undergoes processing in the Golgi and transits to the nucleus,

leading to upregulation of multiple apoptosis-relevant genes and eventual apoptosis. While we do not predict any direct interactions with ATF6, we do predict interactions between NS2A, NS4B, and C with associated pathway member NFYA, which forms a complex with ATF6 in response to ER stress (Yoshida et al., 2001) (Figure 3.8).

In the third branch of the UPR, the ER transmembrane protein IRE1, containing both kinase and RNase activities, becomes autophosphorylated and activated in response to ER stress, leading to XBP-1 splicing and translation of UPR relevant genes. Both DENV2 and Japanese Encephalitis Virus infection have been shown to activate XBP1 and its downstream genes in N18 mouse neuroblastoma cells, reducing the cytopathic effect of the virus (C.-Y. Yu, Hsu, Liao, & Lin, 2006). Knockdown of XBP1 expression by siRNA has also been shown to lead to greater cytotoxicity in response to infection (C.-Y. Yu et al., 2006). Persistent stress leads to apoptosis through an IRE-JNK-BCL2 pathway. Our predictions suggest potential interactions between E and BCL2. In addition, other BCL family members are also predicted to interact with DENV proteins including BCL2ll (BIM; a facilitator of apoptosis) with NS4B, BCL2L1 (BCLX; both pro- and anti-apoptotic splice variants) with E and NS3, and BCL2L10 (Boo; suppression of apoptosis induced by BAX but not BAK) with NS3 (Figure 3.8).

A recent study investigating protein interactions between DENV E protein and host proteins described direct interactions between E and BiP (HSPA5), Calnexin (CANX) and Calreticulin (CALR) (Limjindaporn et al., 2009). All three major ER stress transducers interact with BiP, which serves as a negative UPR regulator, and along with other ER chaperones, facilitates proper folding of proteins. Similarly, Calnexin and calreticulin are chaperones that bind to glycosylated proteins. Our methodology predicts

each of these interactions with the E protein. In addition, we also predict that CALR is likely to interact with NS1 as well as NS3. Overall, these results suggest multiple sites within the host network at which DENV proteins can potentially manipulate the UPR.

In addition, several studies have implicated NS3 in DENV-induced apoptosis. The ability of DENV1 to cause apoptosis in HepG2 cells differs across strains. The mouse neurovirulent strain FGA/NA d1d differs from its parental strain, FGA/89 by 4 mutations, one of which leads to a non-conservative substitution in NS3. FGA/NA d1d was shown to have a reduced capacity to induce apoptosis, although whether this was mediated by the mutation in NS3 or by one of the other mutations, which were all in the E protein, is unclear (Duarte dos Santos et al., 2000). However, Vero cells expressing DENV2 NS3 undergo apoptosis by a mechanism that is dependent on NS3 protease activity and enhanced by the presence of NS2B (Shafee, 2003). In addition, West Nile Virus NS3 is sufficient to induce caspase-8-depedent apoptosis, and is suggested to directly interact with, cleave, and activate caspase-8 in NIH 3T3 cells (Ramanathan et al., 2006). We predicted a number of interactions between NS3 and members of apoptotic pathways. A few examples of the structural similarities that led to these predictions are shown in Figure 3.9. NS3 was predicted to interact with p53 based on structural similarities with RAD51, TK1, and DDX5. Similarities with RAD51 also led to predicted interactions with ABL1, BRCA1/2, CASP3, and CASP7. Furthermore, NS3 has regions of similarity to APAF1, and is therefore predicted to interact with BCL2L1, BCL2L10, Fas, and the caspases -3, -4, -8, and -9. These results suggest that NS3 may play a role in DENV pathogenesis by influencing apoptosis in host cells, mediated by specific interactions between NS3 and host proteins.

DENV and the Interferon Response

Humans and *A. aegypti* are known to use conserved defenses against DENV infection, involving several signaling pathways of the innate immune system, which is consistent with our finding of enriched GO terms related to the immune system among the target proteins of both humans and *A. aegypti*. In particular, the JAK-STAT signaling pathway has been shown to modulate susceptibility to DENV infection, in both mosquitoes and humans (Ho et al., 2005; Souza-Neto, Sim, & Dimopoulos, 2009). In humans, the JAK-STAT pathway can be activated by the interferons (IFN), IFN- α, IFN-β, and IFN-γ, and mediates the antiviral response (Figure 3.10A). When IFN- α or IFN-β bind their receptor, IFNAR, the tyrosine kinases JAK1 and TYK2 are activated. This results in the phosphorylation and activation of STAT2 and STAT1, which then recruit IRF9 to form a transcription factor complex that transcribes IRF-7 and then the set of genes that are induced by IFN-α. The interferon response is induced upon DENV infection and high levels of IFN- α are normally present in the sera of DENV patients (Kurane et al., 1993). Furthermore, pretreatment of cells with IFN has been shown to block negative strand accumulation of DENV RNA, but this inhibition was strongly attenuated if treatment occurred even 4 hours after initial infection (Diamond & Harris, 2001).

In fact, while infection induces an interferon response, several studies have shown that DENV interferes with the signaling pathway downstream of IFN-α. In dendritic cells, DENV is known to protect itself from the antiviral effects of IFN-α by reducing the phosphorylation of TYK2 and preventing the activation of STAT1 and STAT3, although the effects of IFN-γ are not averted (Ho et al., 2005). In addition, IFN-α-, but not IFN-γ-,

dependent phosphorylation of STAT1 and STAT2 was found to be inhibited in A529 and HepG2 cells by the NGC strain of DENV2, but not by the strain TSV01, suggesting strain-dependent rather than serotype specific differences in response (Umareddy et al., 2008). Such strain-dependent differences also highlight the possibility of viral RNA sequence variations that potentially lead to changes in interaction specificity or the strength of interaction with host proteins.

A few specific proteins have been identified as modulators of the IFN response. For instance, IFN-β signaling was prevented by the viral protein NS4B, and to a lesser degree by NS4A and NS2A (Muñoz-Jordan, Sánchez-Burgos, Laurent-Rolle, & García-Sastre, 2003). Inhibition of signaling by NS4B was thought to occur through an observed reduction in the level of phosphorylated STAT1. Expression of STAT2 was also repressed following infection (Jones et al., 2005). Recently, NS5 has also been shown to bind to STAT2, resulting in reduced IFN signaling (Ashour, Laurent-Rolle, Shi, & García-Sastre, 2009). In this same study, when expressed as a proteolytically processed precursor, NS5 was also found to target STAT2 for proteasome-mediated degradation. However, while it is clear that DENV is actively involved in modulating the host interferon response, there likely remain many specific interactions by which DENV proteins inhibit IFN signaling that are not known.

Our predictions suggest many potential interactions between DENV and multiple human proteins involved in the JAK-STAT pathway (Figure 3.10B). In particular, for NS4B, we have predicted possible interactions with IFNAR2, IFNGR1, JAK1, JAK2, TYK2, PTPN11, PTPN2, PTPN6, PKR (EIF2AK2), STAT1, STAT2, and STAT3. Thus, NS4B may reduce the observed phosphorylation of STAT1 through direct interactions

with the host STAT1 protein, through interactions that affect the activities of proteins upstream of STAT1 (e.g. JAK1 or TYK2), or through interactions with at least one PTP protein, which are negative regulators of STAT activity. In addition, NS4B is predicted to interact with PKR, a key component of the IFN response in blocking virus replication. A close relative of DENV, hepatitis C virus, has been shown to inhibit interferon signaling through inhibition of PKR and by competing with eukaryotic translation initiation factor 2a as a PKR substrate (Gale et al., 1998; Taylor, 1999). NS2A is also predicted to interact with the same members of this signaling pathway as NS4B. However, NS4A is predicted to interact with IFNAR1, IFNAR2, IFNB1, JAK2, TYK2, PTPN11, and SOCS1. In summary, we have predicted specific host-pathogen protein interactions that may enable DENV to escape the antiviral response induced by IFN and which can be tested in the future to determine the precise mechanism by which DENV manipulates this host system.

Orthologous Human and *A. aegypti* Targets

To complete its lifecycle, DENV must survive in two very different hosts and must perform many of the same basic processes in each, such as transcription and translation. Since some proteins and essential processes are conserved between mosquitoes and humans, it is possible that some of the proteins that are manipulated by DENV in one host are orthologous to the proteins used in the other host. To identify potential interactions of this type, we compared our interaction network predictions before CC filtering for human and *A. aegypti* to find proteins which are orthologous between the two hosts, as well as predicted to interact with the same DENV protein in each. We found 47 pairs of orthologs that were predicted to interact with the same DENV protein (Figure 3.11, Appendix 11). Four of these predicted interactions represent host

factor targets for *A. aegypti*. We note that our method depends on known interactions within species, and may miss some orthologous host-pathogen interactions if the within species interactions are not known.

To examine the functional role of these conserved interactions, we performed GO term enrichment for biological process and molecular function. No molecular function terms were significantly enriched (Bonferroni corrected p-value, 0.01). Most of the biological processes represented by orthologous interactions in both hosts were also found enriched in the predictions for a single host. For example, at least 17 of the human proteins and their orthologs in *A. aegypti* are involved in mRNA processing or metabolism (ASCC3L1, DCP1B, DCP2, HNRPA1, HNRPD, HNRPF, HNRPM, UPF3B, WDR77, MAGOH, NCBP2, PABPN1, PAPOLA, PRMT5, SNRPA1, SF3B3, SMN1). Five of the interactions known to occur between DENV and human proteins involve the mRNA processing proteins HNRNPK, HNRNPC, PTBP1, and SSB (Chang et al., 2001; García-Montalvo et al., 2004; Jiang et al., 2009; Noisakran et al., 2008). "RNA processing" and "RNA metabolic process" were highly enriched in the mosquito predictions and in the orthologous predictions.

In addition, we found a number of enriched GO terms relating to the regulation or formation of synapses among the predictions conserved in both species. Previously, the term "brain development" was enriched in the set of dDENV-similar proteins. DENV2 virus particles have been found in vesicles near the presynaptic membrane in spinal cords of SCID mice, and it was suggested that fusion of these vesicles at the synapse might aid the spread of DENV2 from neuron to neuron (An, Zhou, Kawasaki, & Yasui, 2003). In *Culex pipiens quinquefasciatus* mosquitoes, West Nile Virus, another flavivirus, was also

found near synapses and in synaptic vesicles (Girard, Popov, Wen, Han, & Higgs, 2005). Furthermore, DENV3 is known to infect the nervous system of *A. aegypti*, altering the mosquito's feeding behavior by prolonging feeding, possibly enhancing the spread of DENV3 by making it more likely that feeding will be interrupted, and the mosquito will have to feed on additional humans (Platt et al., 1997).

The processes enriched in the predicted interactions conserved between the two hosts are consistent with the effects of DENV infection in each host. In particular, many of the terms enriched among the orthologous predictions are similar to terms enriched for mosquito predictions. This is not necessarily surprising, given that the mosquito prediction set is much smaller than the human one, but indicates that orthologous predictions in humans were made corresponding to many of the mosquito predictions. However, we found terms involving cell morphogenesis enriched among the orthologous predictions, but not within the predictions specific to either host. Therefore, the mosquito-specific predictions do not completely overlap with the human predictions, and new processes key to DENV infection in both hosts can be revealed.

CONCLUSIONS

We have created a map of potential protein-protein interactions between the host-pathogen triad DENV2, *H. sapiens,* and *A. aegypti*. The computational methodology employed to generate this map assumes that proteins with comparable structures will share interaction partners. Therefore, we predict that DENV2 proteins may merge into the host protein interactome at the points normally occupied by structurally similar host proteins, creating an interface for the manipulation of downstream host processes. From previous studies, a number of human and fly proteins have been suggested to play some role in DENV2 infection, although the nature of this role is unknown in most cases. Using this methodology, we are able to make predictions regarding which host proteins may impact viral infection through interactions with specific DENV2 proteins. We note that the structural-based methodology here provides a larger picture of the interaction network, while more subtle changes at the sequence level are likely to explain experimentally observed differences in strain effects. Given the paucity of both structural and interaction data for this system, we cannot determine fine differences between strains, but this may be elucidated by further study. The networks presented here may provide a set of hypotheses for further investigation, potential therapeutic intervention, as well as help in improving our understanding of the DENV life cycle.

METHODS

Data Sources

Structures of DENV2 proteins were taken from the PDB (downloaded on Dec. 9, 2009), and any DENV2 protein without a known structure was modeled using I-TASSER (Berman et al., 2003; Y. Zhang, 2008). The protein sequences used to create I-TASSER models were Entrez Protein 159024813, 159024814, 159024817, and 159024819. Default settings were used, with no restraints or selection/exclusion of any templates. Each of the structures for DENV proteins was run on the DaliLite v. 3 webserver (Holm et al., 2008; Holm & Sander, 1993). HPRD Release 7 was used to obtain known human protein-protein interactions, while known *D. melanogaster* interactions were taken from DroID v5.0 and IntAct (Aranda et al., 2010; Mishra et al., 2006; J. Yu et al., 2008). The *A. aegypti* orthologs of *D. melanogaster* proteins were determined using the Inparanoid ortholog annotation for the *D. melanogaster* genes in FlyBase v. FB2009 10 (Berglund, Sjölund, Ostlund, & Sonnhammer, 2008; Crosby, Goodman, Strelets, Zhang, & Gelbart, 2007). The literature sources and various databases used each have their own system of identifiers. PDB codes obtained from Dali were mapped to their corresponding taxonomy and Uniprot accessions using data from the SIFTS initiative, which aims to ease the integration of data from multiple databases (http://www.ebi.ac.uk/msd/sifts/) (Berman et al., 2003; Tagari et al., 2006). Other identifier mappings were carried out using DAVID Gene ID Conversion or Uniprot ID mapping (DAVID 6.7, Uniprot Release 15.14) (Bairoch et al., 2005; Dennis et al., 2003; Huang et al., 2009). Network diagrams were created in Cytoscape (Shannon et al., 2003). Images of protein structures were created in MacPyMol (DeLano, 2002).

Determination of Structural Similarity between DENV and Host Proteins

We investigated protein mimicry using structural similarities from Dali. Dali compares the 3D structural coordinates of two PDB entries by alignment of alpha carbon distance matrices, allowing for differences in domain order, and produces a structural similarity score (Berman et al., 2003; Holm et al., 2008; Holm & Sander, 1993). For this study, we ran each of the DENV2 protein structures, both known and predicted, through the DaliLite webserver, which searched against the entire PDB for structurally similar proteins, with a z score above 2.0. Default settings of a score cutoff of 40 bits and sequence overlap cutoff of 50% were used. We then took from these results only those structures that were from the species *H. sapiens* and *D. melanogaster*. We refer to these human proteins as hDENV-similar proteins and the fly proteins as dDENV-similar proteins.

Interaction Prediction

To predict which human proteins may interact with DENV2 proteins, we sought those target human proteins that interact with the hDENV-similar proteins during cellular processes. To this end, we determined known interactions between hDENV-similar proteins and target human proteins, using data from the Human Protein Reference Database (HPRD) database, which contains literature curated interactions between pairs of human proteins (Mishra et al., 2006). For each hDENV-similar protein, we predicted that the target proteins known to interact with the hDENV-similar protein might also interact with that DENV protein.

A similar process was used to predict interactions between DENV2 proteins and *A. aegypti* proteins, but with the added step of finding orthologs between *A. aegypti* and

D. melanogaster proteins. Known interactions between the dDENV-similar proteins and other *D. melanogaster* proteins were taken from DroID, using a cutoff confidence value of 0.4, and IntAct (Aranda et al., 2010; J. Yu et al., 2008). Then, orthologs of the *D. melanogaster* proteins were found for *A. aegypti* using FlyBase (Crosby et al., 2007). We made the prediction that the *A. aegypti* target protein interacts with the DENV protein.

GO Term Enrichment

The Gene Ontology (GO) provides a system of terms to consistently describe and annotate gene products (Ashburner et al., 2000). GO term enrichment was performed using the DAVID Functional Annotation Chart tool (Dennis et al., 2003; Huang et al., 2009). The GO is organized as a tree structure, with terms becoming more specific as distance from the root increases. Therefore, to avoid very general and uninformative GO terms, we used only GO level 4 terms. The p-values were corrected for multiple testing using the Bonferroni procedure and $-\log_{10}$ transformed.

Validation of Predictions

Since there may be multiple PDB structures present in Dali to represent the same protein, there was some redundancy in the interaction predictions. In some cases, multiple PDB structures for the same DENV protein were found to be similar to multiple PDB structures for a DENV-similar protein, leading to the same interaction predictions. Therefore, the predictions were counted as unique pairs of human Uniprot accessions and DENV protein names. In addition, for ease of viewing the predicted interactome, each node representing a DENV protein is labeled with the protein name while each human protein is represented by its Entrez GeneID.

Support for the predicted interactions was obtained from literature. As few interactions between DENV and humans are known, we looked within the literature to see if any of them were predicted by our method (Chang et al., 2001; Chiu et al., 2007; J. J. E. Chua et al., 2004; Reyes-Del Valle et al., 2005). In addition, recent studies using siRNA screens have found proteins that may play some role, either facilitating or inhibiting in DENV infection, in both humans and *D. melanogaster* (Krishnan et al., 2008; Sessions et al., 2009). *A. aegypti* orthologs of these host factors were recently curated by Guo et al. (Guo et al., 2010). We checked for the presence of these human host factors or mosquito orthologs among our predictions. Although it is not known if these proteins act through direct protein-protein interactions with DENV or indirect mechanisms, their involvement in DENV infection provides functional support for a possible interaction and gives them higher priority for further testing.

GO Cellular Components Filter

GO annotations for the human and *A. aegypti* target proteins were obtained through DAVID 6.7 (Dennis et al., 2003; Huang et al., 2009). However, since DAVID assigns all DENV proteins the same GO terms, GO annotation for the DENV proteins was obtained using the GOanna webserver, provided through AgBase v. 2.00 (McCarthy et al., 2006). This tool assigns GO terms to the input sequences by transitively assigning the GO terms of similar, already annotated sequences identified by BLAST. The most significant BLAST hits for the DENV protein sequences were DENV polyprotein sequences. However, there were multiple polyprotein sequences, each with their own annotations. The input sequences matched more significantly to some polyproteins than to others, and were therefore assigned different GO terms based on sequence similarity.

The predicted interactions were filtered so that only those predictions for which the DENV protein and host protein shared at least one GO cellular component term were retained.

Predictions Using Mosquito Interactome

Guo et al. recently generated a first draft of the mosquito interactome (Guo et al., 2010). Because their interactome was based on the three model organisms *A. aegypti*, *C. elegans*, and *S. cerevisiae*, we found proteins from all three of these species that show structural similarity with DENV2 using the Dali server (Berman et al., 2003; Holm & Sander, 1993). The *A. aegypti* orthologs of *C. elegans* and *S. cerevisiae* proteins were determined using InParanoid, and the *D. melanogaster* orthologs were taken from the InParanoid ortholog annotation for the *D. melanogaster* genes in FlyBase v. FB2009 10 (Berglund et al., 2008; Crosby et al., 2007). Then, the interactions with these orthologs taken from the mosquito interactome were used to map predicted interactions between DENV2 and mosquito target proteins. For the GO term enrichment, we used only GO terms from DAVID's GO fat set, to eliminate non-specific terms with many children.

Determining Orthologous Targets

The genome-wide set of orthologs between human and *A. aegypti* was downloaded from InParanoid 7.0 (Berglund et al., 2008). Because InParanoid lists human proteins by their Ensembl Protein IDs, mappings to Uniprot accessions were downloaded from Ensembl 57 using BioMart (Flicek et al., 2010). Orthologous human and *A. aegypti* targets that were predicted to interact with the same DENV2 protein were identified.

FIGURES AND TABLES

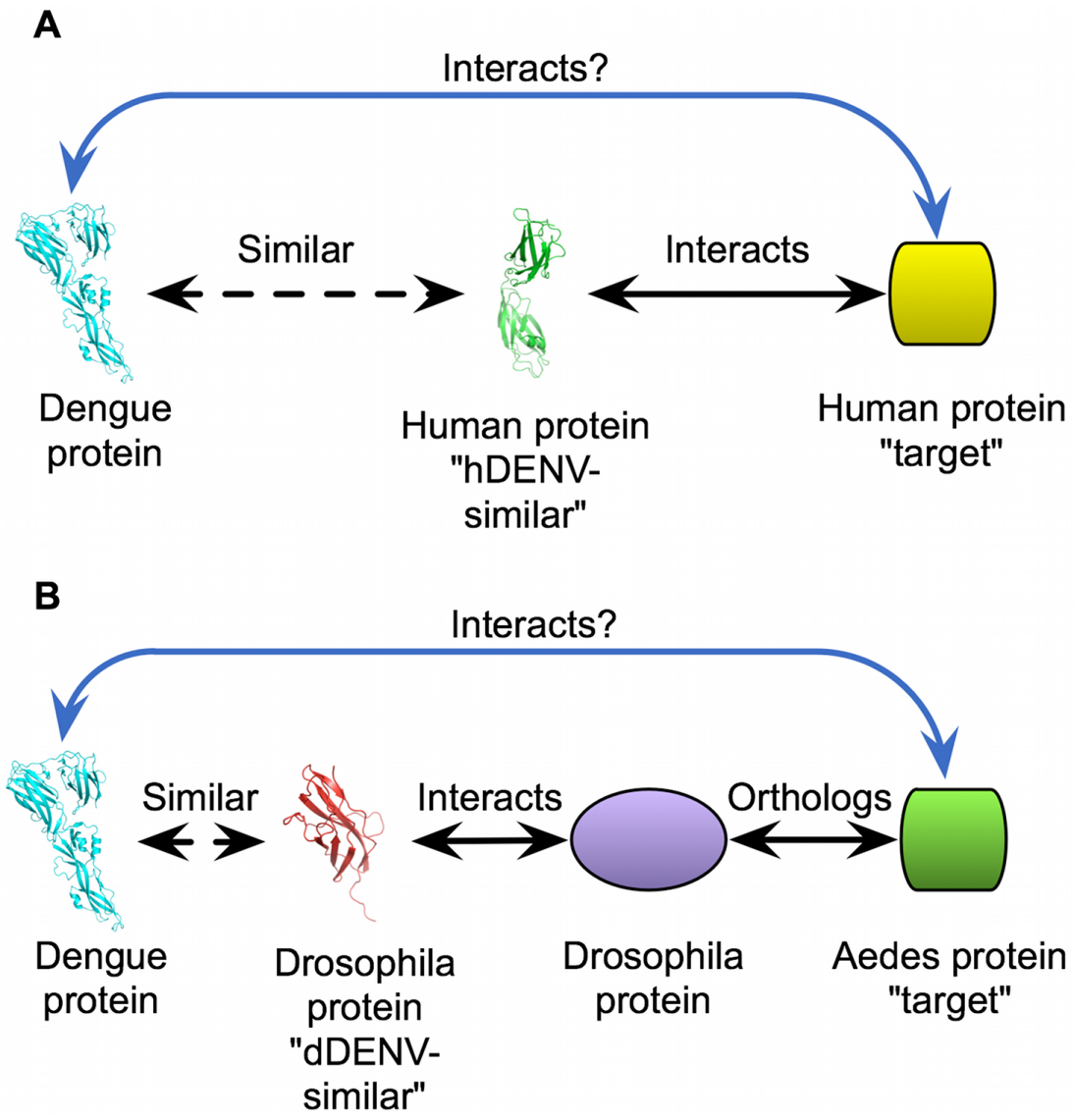


Figure 3.1. Diagram of approach. (A) Predictions for the human host. Interactions between DENV proteins and human targets are predicted on the basis of structural similarity between the DENV protein and an hDENV-similar protein and the hDENV-similar protein's known interaction with the human target. (B) Predictions for the insect host are made in a similar manner as (A), except for the additional step of finding orthologs of the *D. melanogaster* target proteins in the real host of interest, *A. aegypti*.

Table 3.1. Known interactions between *H. sapiens* and DENV.

Human	DENV	Serotype	System	Reference
UBE2I	E	2	In vitro†	(Chiu et al., 2007)
HSP90AA1	E	2	U937, SK-SY5Y, monocyte	(Reyes-Del Valle et al., 2005)
HSPA4	E	2	U937, SK-SY5Y, monocyte	(Reyes-Del Valle et al., 2005)
HSPA5	E	2	HepG2, Vero	(Jindadamrongwech et al., 2004; Limjindaporn et al., 2009)
CANX	E	2	Vero	(Limjindaporn et al., 2009)
CALR	E	2	Vero	(Limjindaporn et al., 2009)
CD14	E	2	Primary monocytes/ macrophages	(Y.-C. Chen et al., 1999)
CD209	E	1‡	BHK†	(Lozach et al., 2005)
RPSA	E	1*	HepG2	(Thepparit & Smith, 2004)
DAXX	C	2	HepG2	(Limjindaporn et al., 2007)
HNRNPK	C	2	293T†	(Chang et al., 2001)
HNRNPC	NS1	2	293T	(Noisakran et al., 2008)
CLU	NS1	2	Plasma, 293T, Vero	(Kurosu et al., 2007)
STAT3	NS1	2	BHK	(J. J.-E. Chua et al., 2005)
NRBP1	NS3	2	BHK†	(J. J. E. Chua et al., 2004)
SSB	NS3	4	U937	(García-Montalvo et al., 2004)
SSB	NS5	4	U937	(García-Montalvo et al., 2004)
TJP1	NS5	2	Epithelial cells	(Ellencrona et al., 2009)
STAT2	NS5	2	293T	(Ashour et al., 2009)
PTBP1	NS4A	2	Huh-7	(Jiang et al., 2009)

Experimentally determined interactions are listed, with the serotype and system that the interaction was demonstrated in.
†Interaction suggested in additional cell lines by functional assay.
‡Interaction suggested for other serotypes by functional assay.
*Interaction shown to be specific for this serotype.

Table 3.2. Summary of interaction predictions.

	Human	Human CC	Insect	Insect CC
DENV-similar	300	254	15	6
Targets	2,321	1,099	158	12
Known Predictions	9	7	0	0
Host Factor Predictions	48	20	12	1
Total Predictions	4,273	2,073	176	17

Counts are given for the predictions made between DENV and both the human and *A. aegypti* hosts, both before and after CC filtering.

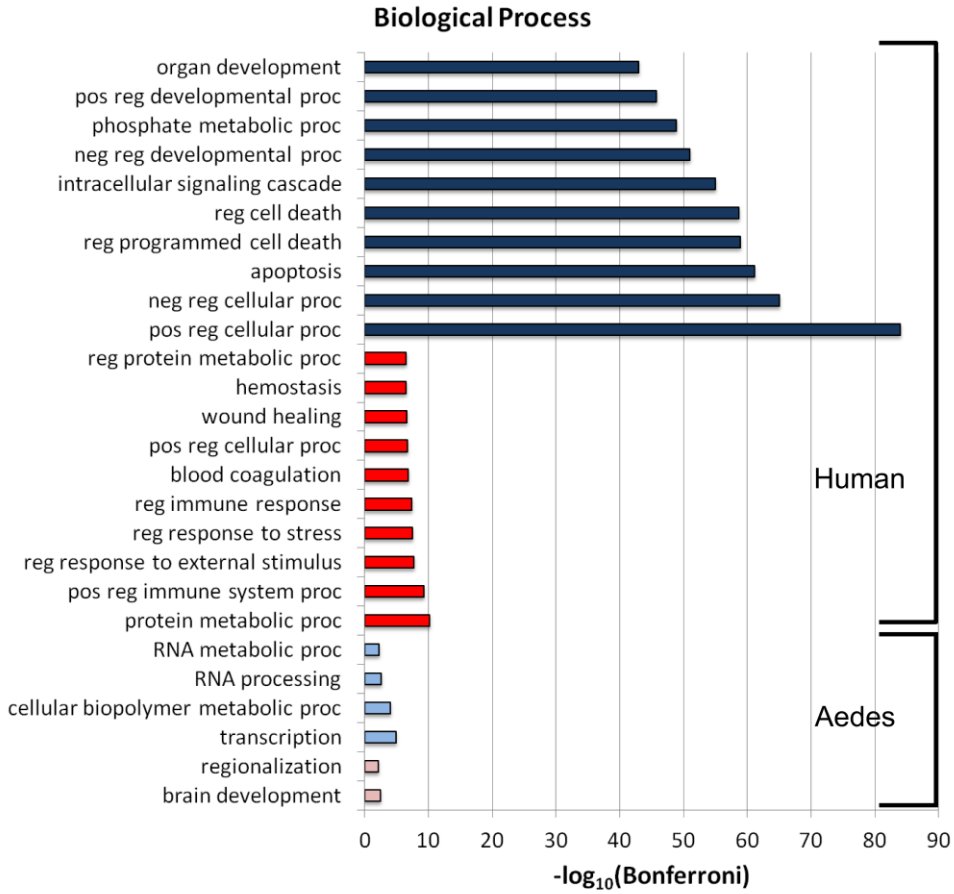


Figure 3.2. GO Biological Process term enrichment of host proteins. Blue bars represent terms enriched among human target proteins, red is terms enriched among hDENV-similar proteins, light blue is terms for *A. aegypti* targets, and pink is for terms from dDENV-similar proteins. When more than ten terms were enriched for a set of proteins, only the ten most significant terms are shown. Bonferroni corrected p-values were transformed by $-log_{10}$. The following abbreviations are used: "reg" is "regulation of," "pos" is "positive," "neg" is "negative," and "proc" is "process." Brackets delineate the two host species.

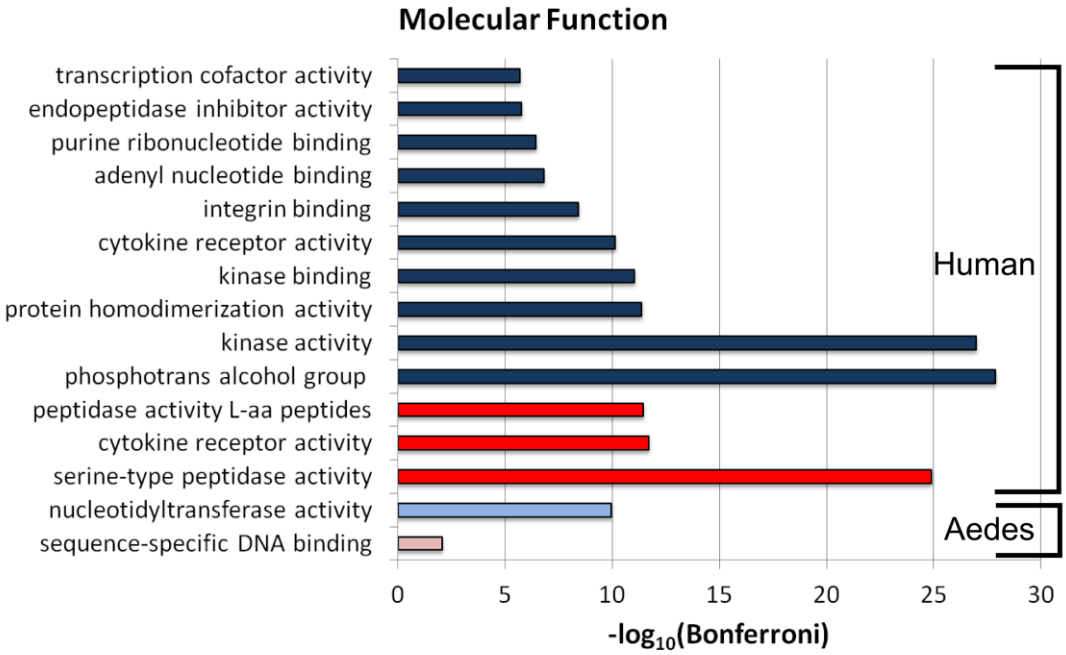


Figure 3.3. GO Molecular Function term enrichment of host proteins. Blue bars represent terms enriched among human target proteins, red is terms enriched among hDENV-similar proteins, light blue is terms for *A. aegypti* targets, and pink is for terms from dDENV-similar proteins. When more than ten terms were enriched for a set of proteins, only the ten most significant terms are shown. Bonferroni corrected p-values were transformed by $-log_{10}$. The following abbreviations are used: "peptidase activity L-aa peptides" is "peptidase activity acting on L-amino acid peptides," and "phosphotrans alcohol group" is "phosphotransferase activity alcohol group as acceptor." Brackets delineate the two host species.

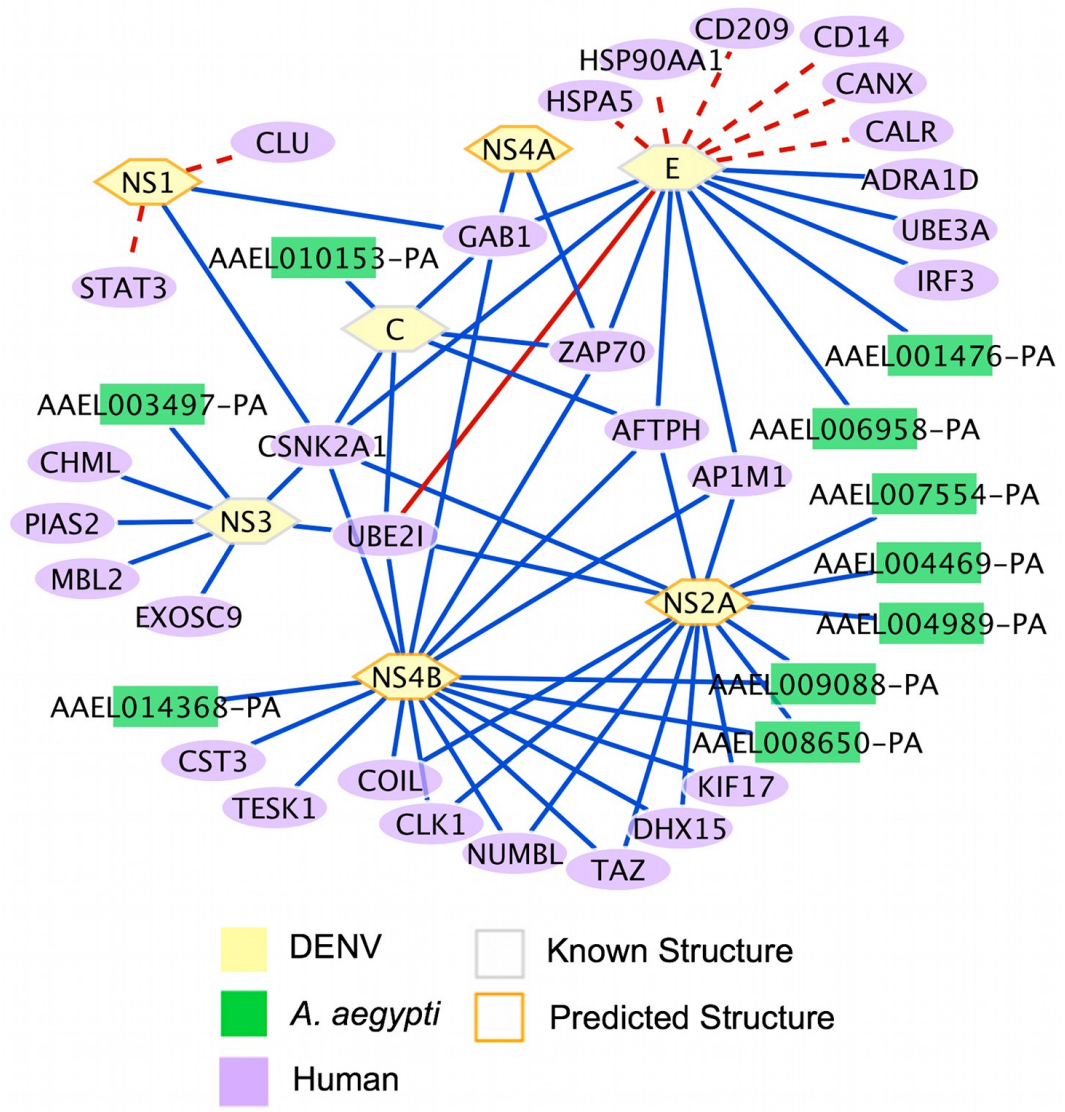


Figure 3.4. Predicted interactions with literature support. Predicted interactions between DENV and its hosts that were already known or involve host factors are shown. Solid lines represent interactions for which the host protein was found by an siRNA screen to be involved in DENV infection, while dashed lines indicate that it is not a known host factor. Red lines represent interactions already known from the literature.

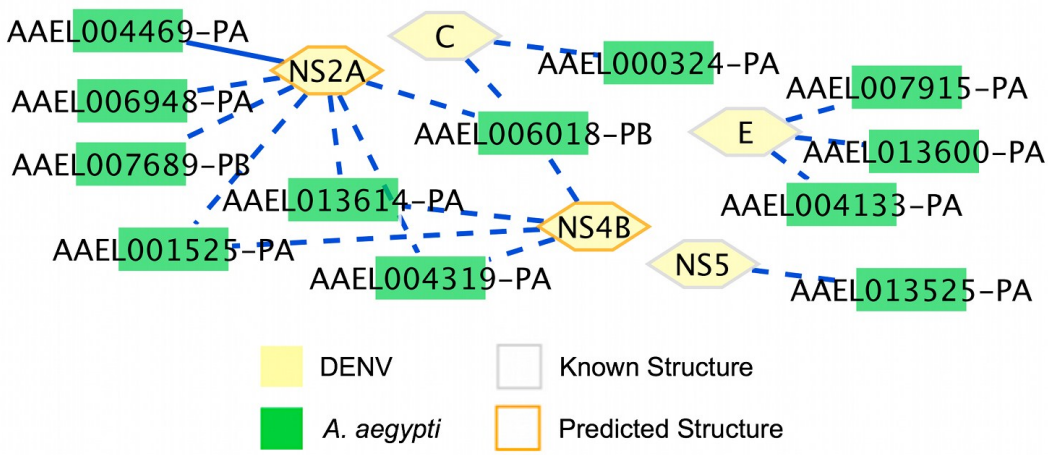


Figure 3.5. Predicted interactions in *A. aegypti* after CC filtering. Predicted interactions between DENV and *A. aegypti* where the DENV protein and its target share at least one GO CC term. Solid lines represent interactions for which the *A. aegypti* protein is a host factor.

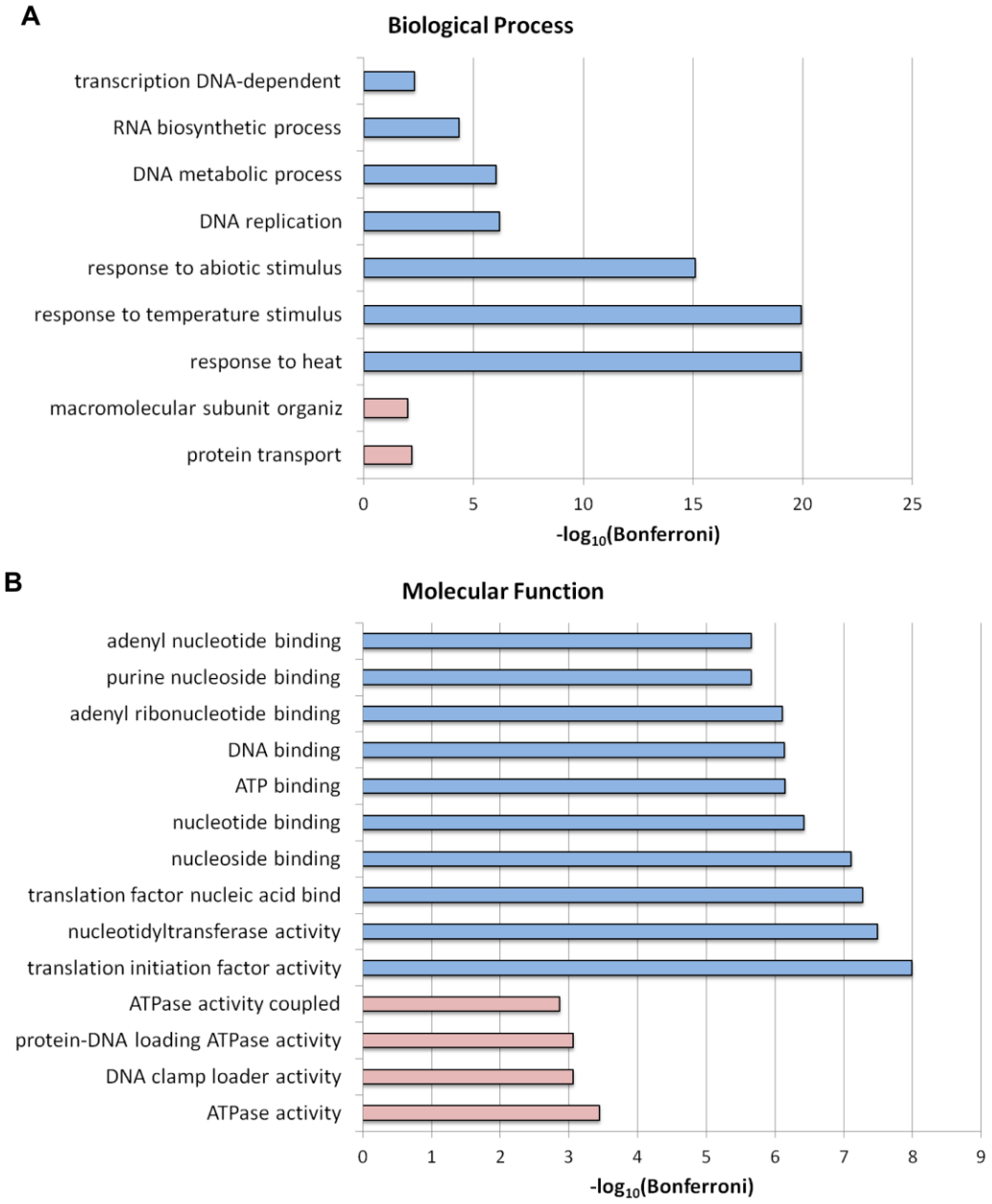


Figure 3.6. GO term enrichment of *A. aegypti* proteins based on data from Guo et al. (Guo et al., 2010). (A) Enriched GO biological process terms. (B) Enriched GO molecular function terms. Light blue bars represent terms for *A. aegypti* targets, and pink is for terms from DENV-similar proteins. When more than ten terms were enriched for a set of proteins, only the ten most significant terms are shown. Bonferroni corrected p-values were transformed by $-log_{10}$. The following abbreviations are used: "translation factor nucleic acid bind" is "translation factor activity nucleic acid binding," and "macromolecular subunit organiz" is "macromolecular complex subunit organization."

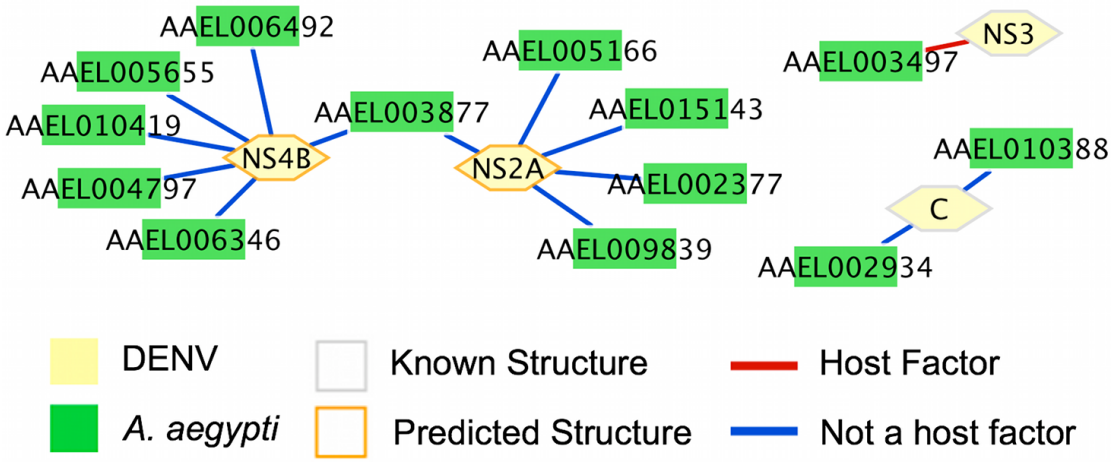


Figure 3.7. Interactions predicted using both the original fly data and the mosquito interactome.

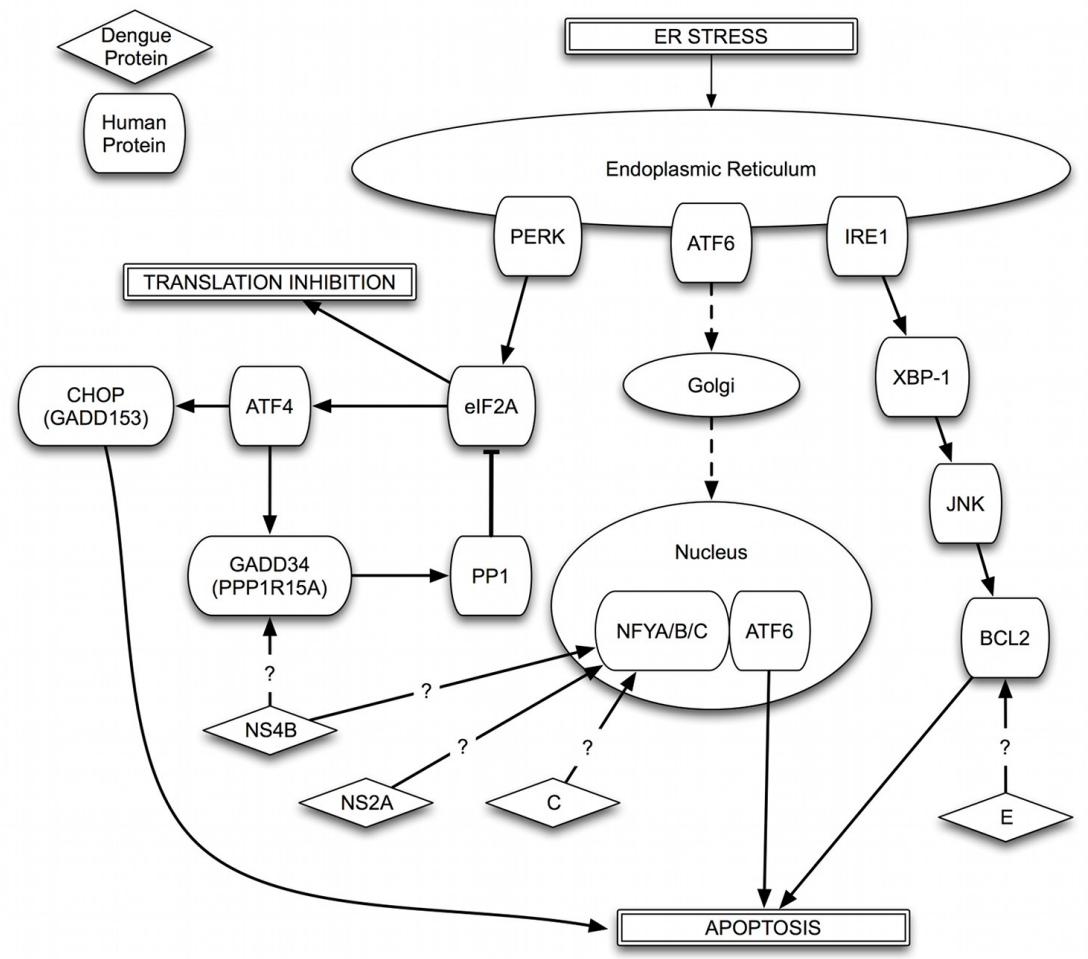


Figure 3.8. DENV and ER Stress. Potential interactions between DENV proteins and key components of the Unfolded Protein Response (UPR) and ER stress. See text for additional details.

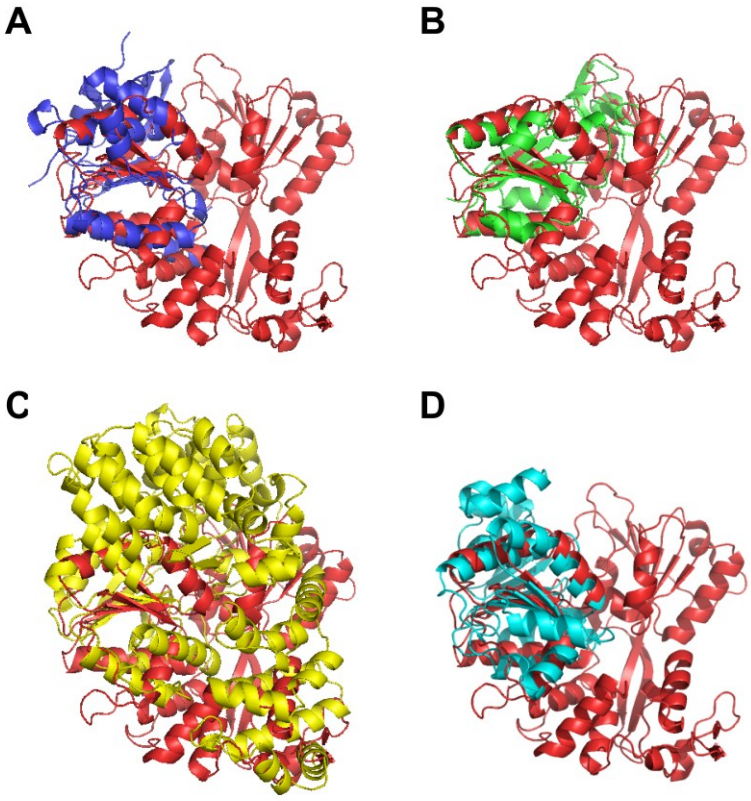


Figure 3.9. Structural similarities between NS3 and human proteins involved in apoptosis. (A) Structural similarity between NS3 and RAD51 (1n0wA) (Pellegrini et al., 2002). (B) Structural similarity between NS3 and TK1 (1w4rA) (Birringer et al., 2005). (C) Structural similarity between NS3 and APAF1 (1z6tC) (Riedl, Li, Chao, Schwarzenbacher, & Shi, 2005). (D) Structural similarity between NS3 and DDX5 (3fe2A) (Schütz et al., 2010). NS3 (2bhrA) (Xu et al., 2005) is shown in red.

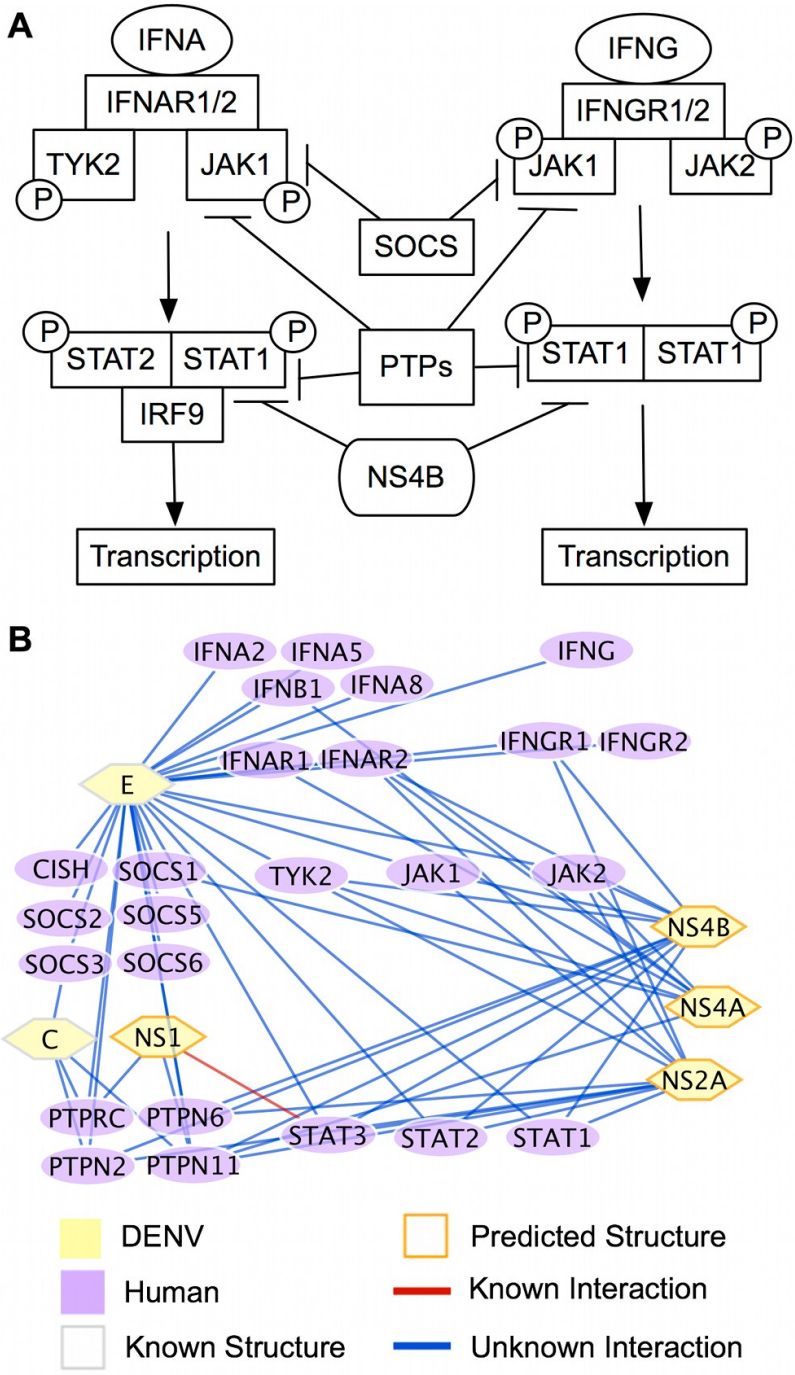


Figure 3.10. DENV influences IFN signaling. (A) Interferon signaling pathway. IFNA and IFNG bind to their respective receptors and cause the activation of Jak family tyrosine kinases. This activates STAT proteins, which form hetero- or homodimers and induce the expression of IFN response gene. SOCS proteins negatively regulate JAK1, and PTP proteins negatively regulate JAK1 and STAT1. NS4B can reduce the phosphorylation of STAT1. (B) Predicted interactions between DENV proteins and members of the IFN-induced JAK-STAT pathway.

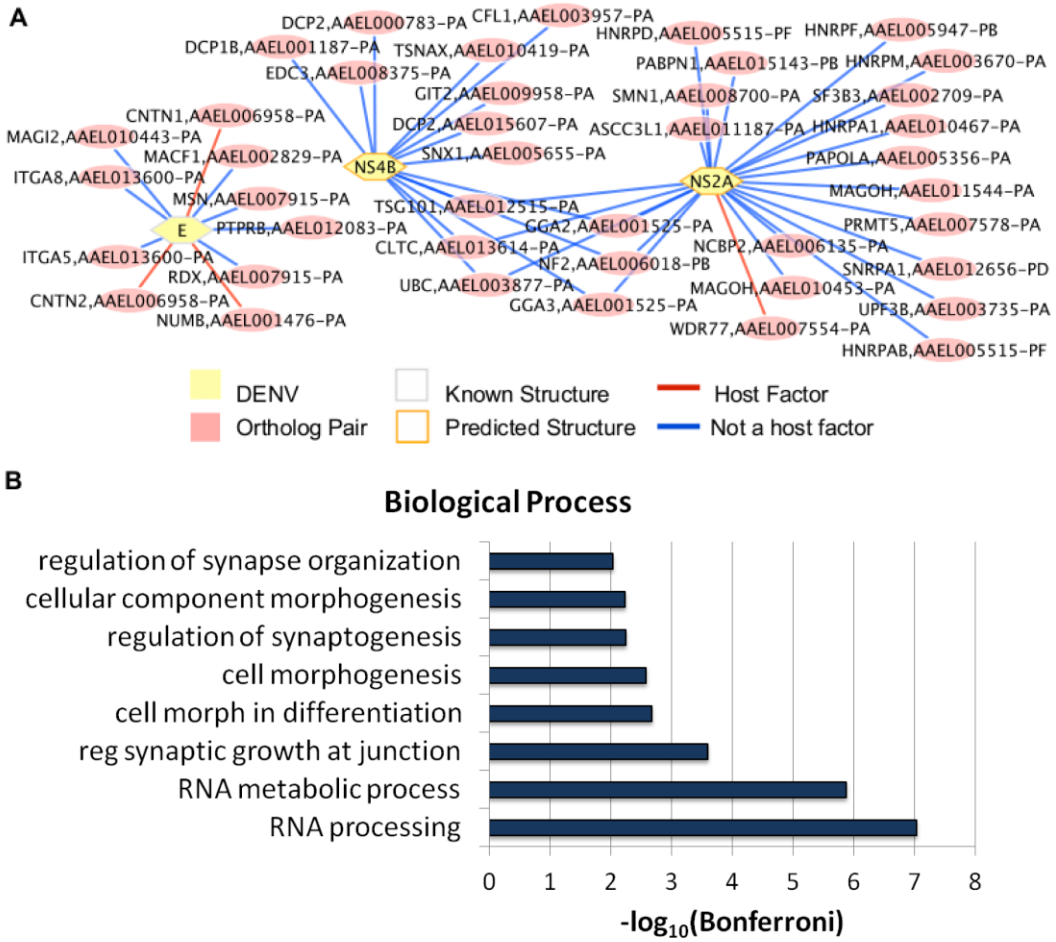


Figure 3.11. Predicted orthologous interactions. (A) Predicted interactions between DENV and orthologous pairs of *A. aegypti* and human proteins. The human protein is listed first, followed by its ortholog in *A. aegypti*, which is also predicted to interact with the DENV protein. (B) GO biological process terms enriched among the interactions predicted to be conserved between human and *A. aegypti*. The following abbreviations are used: "reg synaptic growth at junction" is an abbreviation for "regulation of synaptic growth at neuromuscular junction," and "cell morph in differentiation" stands for "cell morphogenesis involved in differentiation."

CHAPTER FOUR

CONCLUSION

SUMMARY OF FINDINGS

A novel method of predicting host-pathogen interactions based on structural similarities between proteins was developed and applied to HIV-1 and DENV2. For each virus, the predictions were filtered using GO CC terms and functional evidence from the literature to give sets of high confidence predictions. The two viruses studied had very different signatures of GO term enrichment among the target human proteins they were predicted to interact with. A significant number of HIV-1 targets were involved in signaling, transport, and GTP binding or hydrolysis, while the predictions for DENV2 centered on RNA processing, stress responses, and interferon signaling. Part of the DENV2 lifecycle occurs in mosquito vectors that spread the disease, and predictions were also made for interactions with this host, showing similar target functions to those in the human host. We estimate that 10% of predictions are accurate based on the known interactions for HIV-1 using gene identifiers.

CHALLENGES

Lack of known host-pathogen interactions remains a major hurdle in determining interspecies interactomes. Recent advances using large-scale Y2H or proteomic approaches have greatly expanded the host-pathogen interaction networks for many

pathogens. However, these approaches have high rates of false positives, making the quality of the data uncertain, and false negatives, leaving interactomes incomplete. Furthermore, many human proteins are involved in viral pathogenesis, suggesting some form of interaction with the virus, but have not been shown to directly interact with a viral protein. For pathogens that can be transmitted to humans from animals or are borne by vectors, such as mosquitoes, knowledge of the non-human host(s) can also prove limiting. Technological developments have recently lead to great advances, but studies of mosquitoes are still far behind studies on humans and model organisms (reviewed in (Severson & Behura, 2012)). Therefore, computational methods are necessary to augment experimental approaches, especially in cases where a pathogen can infect multiple diverse hosts.

Most protein interaction prediction methods are forms of supervised learning and rely on prior information about known interactions to make predictions about unknown ones. The lack of this information prohibits the use of many machine-learning algorithms for many viruses. Furthermore, because negative results are rarely reported, there are no sets of host proteins known not to interact with pathogen proteins to provide negative sets for classification methods. Random protein pairs can be chosen with the assumption that it is unlikely they will interact, but it cannot be determined if these represent true negative cases.

Even if experimental data exists for the pathogen, an additional challenge lies in collecting a comprehensive list of previously known interactions. Although several databases exist to curate protein interaction information, these databases are incomplete, lacking many known interactions that can be found in the literature and making the

determination of a gold-standard set of interactions difficult. For example, only 3 DENV-human interactions were found in databases in 2008 (Dyer et al., 2008), while we found 16 more by literature search. A recent compilation of *Salmonella*-host interactions found only 6 of 62 known interactions in automatically retrievable databases (Schleker et al., 2011). Even HHPID, a database specifically for HIV-human protein interactions, has missing data, redundant data, and data of varying quality, including interactions that may be indirect (MacPherson, Dickerson, Pinney, & Robertson, 2010). Various methods to automate literature mining specifically tailored to host-pathogen interactions are discussed in (Korkin, Thieu, Joshi, & Warren, 2011).

This work has focused on interactions between host and viral proteins. However, protein-protein interactions are only one way in which the virus can manipulate the host. For example, the La autoantigen in both humans and mosquitoes interacts with the 3'-UTR of viral RNA, in a protein-RNA interaction that may regulate viral replication (De Nova-Ocampo, Villegas-Sepúlveda, & del Angel, 2002; Yocupicio-Monroy et al., 2007). In addition, RNA-RNA interactions play an important role in viral infection; host cells can defend themselves from viruses using RNAi and several viruses have been shown to encode microRNAs that may regulate host transcripts (reviewed in (Scaria, Hariharan, Maiti, Pillai, & Brahmachari, 2006)). Maulik et al. recently predicted the possible role of microRNAs regulated by human proteins that are known or predicted to be highly connected to HIV proteins in the HIV-human protein interactome (Maulik, Bhattacharyya, Mukhopadhyay, & Bandyopadhyay, 2011). To understand fully how a pathogen hijacks its host, these and other types of interactions will need to be examined.

CONTINUING PROGRESS IN THE FIELD

Despite the challenges in determining host-pathogen protein interactomes, recent progress has been made for both HIV and DENV. A recent study used biclustering to create association rules used to predict HIV-human interactions (Mukhopadhyay, Maulik, & Bandyopadhyay, 2012). Of the 180 predictions made, 5 overlap with our study, and 4 of these also overlap with predictions from Tastan et al. (Tastan et al., 2009). In addition, Qi et al. used multi-tasked learning, taking advantage of the many HIV-human interactions that are suggested but not proven as a "weakly labeled" set to augment a gold-standard set of 158 interactions validated by experts (Qi, Tastan, Carbonell, Klein-Seetharaman, & Weston, 2010). Dyer et al. used Support Vector Machines to classify HIV-human protein pairs as interacting or not based on domain composition, protein sequence, and properties of the human protein interaction network (Dyer, Murali, & Sobral, 2011). Nouretdinov et al. predicted HIV-human protein interactions using the conformal method to assign p-values and confidence levels to predicted interactions. Confidence levels can be valuable to experimentalists when deciding how much risk of the failure is acceptable for their purposes (Nouretdinov, Gammerman, & Qi, 2012). These supervised learning approaches by Mukhopadhyay et al, Qi et al, Dyer et al., and Nouretdinov et al. require a large set of already known interactions, although the methods by Mukhopadhyay et al. and Nouretdinov et al. do not use negative sets. Therefore, they cannot be trained accurately for the many viruses about which little is known, and it is unlikely that the rules or criteria learned for HIV will apply to unrelated viruses.

Most computational methods to determine the host-pathogen interactome still focus on HIV due to the relatively vast amount known about this virus, but experimental

progress has been made for DENV in recent years. Since our predictions for the DENV interactome, three screens have been conducted, vastly increasing the size of the known DENV-human interactome. In the first study, 31 human proteins were found to interact with the DENV structural proteins, E, C, and prM, in a bacterial 2-hybrid screen, with a few proteins interacting with more than one of these viral proteins (Folly, Weffort-Santos, Fathman, & Soares, 2011). Interestingly, they found enrichment in the stress response, wound healing, and the complement and coagulation cascade, correlating very well with our predictions. Of the 31 human proteins they found to interact with DENV, 6 were among our predictions with 1 associated with the same DENV protein. Four of the 31 proteins were predicted to interact with non-structural proteins, which were not examined in the screen, rather than structural proteins (APOA2 with NS4B, ARHGEF11 with NS4A, CLU with NS1, and F8 with NS3). In addition, we predicted IPO13 interacts with C, whereas they found it to interact with E. One of our predictions was shown to be correct in the screen: PLG interacting with E. Folly et al. found PLG also interacts with prM, while we predict it additionally interacts with NS1, NS3, and NS4B (Folly et al., 2011).

In the second study, a Y2H screen found 139 unique interactions between DENV and human proteins, most of which were previously unknown (Khadka et al., 2011). Six of the interactions found were correctly predicted by our study. A third Y2H screen focused only on the NS3 and NS5 proteins of *flaviviruses*, including DENV1, finding 186 interactions with human proteins in their screen or in the literature (Le Breton et al., 2011). Three of these were predicted by our study: AMBP, CASP8, and ENO1

interacting with NS3. Despite this recent progress, it is likely that many more DENV-human interactions remain undiscovered.

In addition, several recent smaller scale experiments on DENV-human protein interactions have come to our attention. Johansson et al. found that NS5 localizes to the nucleus in DENV infected cells by interacting with the nuclear import receptor importin-β, and NS3 competes with importin-β for the same binding site on NS5 (Johansson, Brooks, Jans, & Vasudevan, 2001). NS5 shuttles out of the nucleus by interacting with the nuclear export factor CRM1 (Rawlinson, Pryor, Wright, & Jans, 2009). We predict importin-β (KPNB1) interacts with NS2A, NS4B, C, and E and that CRM1 (XPO1) interacts with NS1, NS2A, NS4B, E, C, and NS3, but not NS5. In addition, Sec3 (EXOC1) was shown to interact with DENV C protein, but was not among our predictions (Bhuvanakantham, Li, Tan, & Ng, 2010). Another study found that vacuolar ATPase interacts with DENV prM and this interaction mediates viral entry and egress (Duan, Lu, Li, & Liu, 2008). Heaton et al. found that XBP-1 and ATF6, two players in ER stress, are necessary for completion of the DENV lifecycle and uncovered an interaction between NS3 and fatty acid synthase (FASN) (Heaton et al., 2010). None of our predictions involve Sec3, vacuolar ATPase, or FASN. There has been much progress in recent years on the DENV-human interactome, but very little is yet known about the interactions between DENV and mosquitoes.

A larger base of knowledge about host-virus interactomes will aid further discovery by computational methods. Large-scale screens have been conducted for several other virus-host interactomes in recent years, including Chikungunya virus,

influenza polymerase subunits, and variola virus (Bouraï et al., 2012; Mohamed et al., 2009; Tafforeau et al., 2011).

FUTURE DIRECTIONS

Ongoing work concerns mechanisms of reactivation of latent KSHV. Two chemicals are known to induce KSHV replication, sodium butyrate (*n*-butyrate) and 12-O-tetradecanoylphorbol-13-acetate (TPA). A third viral replication inducer is spent media from a common oral gram-negative bacterium, *Porphyromonas gingivalis*, which contains short-chain fatty acids and lipopolysaccharide. Spent media from *P. gingivalis* has been shown to induce KSHV reactivation by a mechanism that differs, at least partially, from that of TPA- or *n*-butyrate- induced reactivation in the latently infected BCBL-1 cell line (Morris, Arnold, & Webster-Cyriaque, 2007; Y. Yu et al., 1999). A PKC inhibitor prevented KSVH reactivation in TPA treated cells, but not in cells treated with *n*-butyrate or spent media from *P. gingivalis*. In addition, treatment of *n*-butyrate and spent media from *P. gingivalis* caused hyperacetylation of histones 3 and 4, whereas TPA did not (Morris et al., 2007). Viral reactivation by *n*-butyrate led to higher levels of viral gene expression than TPA treatment, and unlike TPA, *n*-butyrate induces apoptosis in BCBL-1 cells (Y. Yu et al., 1999). Gene expression changes that may underlay these mechanisms and observed differences are being explored.

To investigate gene expression changes during KSHV reactivation, differential gene expression was measured by former members of the Webster-Cyriaque lab in untreated BCBL-1 cells and BCBL-1 cells treated with TPA, *n*-butyrate, or spent media from *P. gingivalis* using custom two-color microarrays. Three replicates were performed

for each of the three conditions. Using the Bioconductor package limma, I background subtracted the red and green channels, and the data were normalized within each array using the print tip loess method and between arrays by quantile normalization of the green channel, which represented the untreated reference on all arrays (Ritchie et al., 2007; G. Smyth, 2005; G. K. Smyth & Speed, 2003). M values for multiple spots representing the same gene were averaged.

Preliminary pathway analysis was performed using Gene Set Enrichment Analysis (GSEA). Gene expression from cells treated with *P. gingivalis* spent media was compared to gene expression from untreated, *n*-butyrate, or TPA treated cells and genes were ranked by correlation with treatment condition. One interesting gene set, up regulated in BCBL-1 cells treated with *P. gingivalis* spent media but not with either TPA or *n*-butyrate, consists of genes containing the motif KTGGYRSGAA in their promoter region. However, the function of this motif is currently unknown.

Future work will continue to take a systems approach to the interactions between viruses and their hosts. Direct protein-protein interactions are only one means by which a virus can manipulate its host, and other microbes present in the human body can also influence host-virus interactions, as illustrated by the reactivation of KSHV by *P. gingivalis* end products. High throughput methods targeting different physical and functional interactions, such as Y2H or TAP-MS, microarray or RNA-seq, and ChIP-seq or FAIRE-seq, are needed to characterize the global cellular response to clinically relevant viruses. My interests lay in using systems approaches to determine the mechanisms by which microbes interact with and cause disease in humans.

CONCLUDING REMARKS

Characterization of host-pathogen interactions is far from complete. This work has led to predictions of the protein interactions underlying the mechanisms that allow two diverse viruses to subvert host cell processes. The higher confidence predictions supported by several lines of evidence provide hypotheses that can be tested experimentally in the future. This method can be applied to any pathogen for which there is adequate structural information. Understanding the mechanisms by which viruses manipulate host cell proteins to complete the viral lifecycle will lead to novel drug targets and pharmacological approaches. In addition, knowledge of viral interactions with non-human hosts will aid vector control strategies. Novel treatment and control measures are needed to reduce the global health burden of infectious disease.

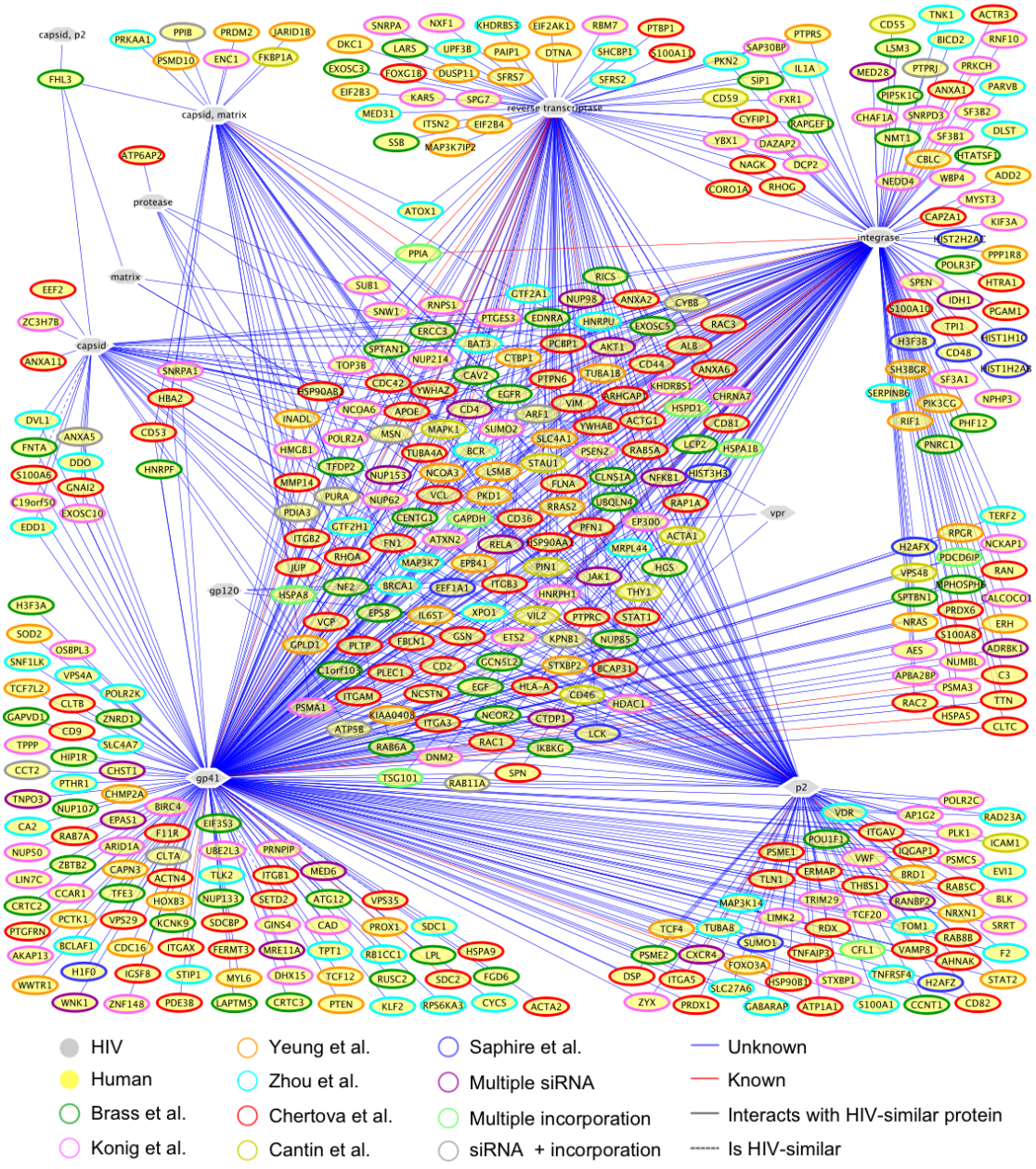
APPENDIX 1: Twenty-five selected structural similarities between HIV-1 and human proteins from the Dali Database

HIV PDB code	HIV Molecule Name	HIV GeneID	HIV Uniprot	Human PDB code	Human Gene Symbol	Human GeneID	Human Uniprot	Structural Similarity Score
1ex4A	integrase	155348	P04585	2ysqA	ARHGEF9	23229	O43307	3.6
1ex4A	integrase	155348	P04585	3dkmA	NA	25831	Q9ULT8	2.9
1ex4A	integrase	155348	P04585	3dtpB	MYH7	4625	P12883	3.9
1ex4A	integrase	155348	P04585	1g83A	FYN	2534	P06241	3.5
1ex4A	integrase	155348	P04585	1zbuB	THEX1	90459	Q8IV48	2
1ex4A	integrase	155348	P04585	2creA	C20ORF32	57091	Q9NQ75	4.2
1ex4A	integrase	155348	P04585	2diqA	TDRKH	11022	Q9Y2W6	2.5
1ex4A	integrase	155348	P04585	2dlmA	SORBS3	10174	O60504	3.2
1ex4A	integrase	155348	P04585	2egeA	RIMBP3	---	Q9UFD9	3.2
1ex4A	integrase	155348	P04585	2ew3A	SH3GL3	6457	Q99963	4.1
1ex4A	integrase	155348	P04585	2qkkA	RNASEH1	246243	O60930	4
1ex4A	integrase	155348	P04585	3eo3A	GNE	10020	Q9Y223	3.5
1ex4A	integrase	155348	P04585	1ark	NEB	---	P20929	3.7
1ex4A	integrase	155348	P04585	1aww	BTK	695	Q06187	4
1ex4A	integrase	155348	P04585	1hsq	PLCG1	5335	P19174	3.4
1ex4A	integrase	155348	P04585	1j3tA	ITSN2	50618	Q9NZM3	4.4
1ex4A	integrase	155348	P04585	1mhnA	SMN2	6607	Q16637	4.2
1ex4A	integrase	155348	P04585	1udlA	ITSN2	50618	Q9NZM3	3.7
1ex4A	integrase	155348	P04585	1ue9A	ITSN2	50618	Q9NZM3	3.9
1ex4A	integrase	155348	P04585	1uffA	ITSN2	50618	Q9NZM3	2.9
1ex4A	integrase	155348	P04585	1ugvA	ARHGAP26	23092	Q9UNA1	4.2
1ex4A	integrase	155348	P04585	1wx6A	NCK2	8440	O43639	4.1
1ex4A	integrase	155348	P04585	2bz8B	SH3KBP1	30011	Q96B97	4.4
1ex4A	integrase	155348	P04585	2cowA	KIF13B	23303	Q9NQT8	2.8
1ex4A	integrase	155348	P04585	2dl3A	SORBS1	10580	Q9BX66	4.8

APPENDIX 2: Twenty-five selected interaction predictions between HIV-1 and human

HIV PDB code	HIV protein name	HIV-similar Human PDB code	HIV-similar Human Gene Symbol	Human interactor Gene Symbol	Source Datasets	True Positive
1hniA	reverse transcriptase	2rd7A	C8A	CD59	Cantin et al.	no
1ikyB	reverse transcriptase	1j4wA	FUBP1	VIM	Chertova et al.	no
1t03A	reverse transcriptase	1khmA	HNRPK	KHDRBS1	Konig et al.	no
1dmpA	protease	1lyaA	CTSD	FN1	Chertova et al.	yes
2vg6A	reverse transcriptase	1jfiC	TBP	GTF2A1	Zhou et al.	no
1ce0B	gp41	1junA	JUN	BRCA1	Zhou et al.	no
1f23C	gp41	1w0bA	ERAF	HBA2	Chertova et al.	no
1proA	protease	2yujA	UFD1L	VCP	Chertova et al.	no
1ihwA	integrase	1zsgA	ARHGEF7	CDC42	Chertova et al.	no
1rtjA	reverse transcriptase	1qu6A	EIF2AK2	HSP90AA1	Chertova et al.	no
1ihvA	integrase	2jxbA	NCK2	EGFR	Brass et al.	no
1rpiA	protease	2qp8B	BACE1	PDIA3	Zhou et al., Chertova et al.	no
1sv5A	reverse transcriptase	1tceA	SHC1	PTPN6	Chertova et al.	no
3aidB	protease	1lyaA	CTSD	FN1	Chertova et al.	no
1ikwA	reverse transcriptase	1tceA	SHC1	MAPK1	Cantin et al.	no
1hihB	protease	1lyaB	CTSD	CD4	Brass et al., Zhou et al.	yes
1c1cA	reverse transcriptase	1qu6A	EIF2AK2	RAC1	Chertova et al.	no
1bdqA	protease	1lyaA	CTSD	FN1	Chertova et al.	no
1ex4A	integrase	1ycsB	TP53BP2	RELA	Brass et al., Konig et al., Zhou et al.	no
1ce0A	gp41	2rmzA	ITGB3	ITGA5	Chertova et al.	no
1rtiA	reverse transcriptase	1tceA	SHC1	MAPK1	Cantin et al.	yes
2cmrA	gp41	1qjbA	YWHAZ	BCR	Zhou et al.	no
3cyoA	gp41	3brtD	IKBKG	HSP90AB1	Chertova et al.	no
4hvpA	protease	3braA	BACE1	PDIA3	Zhou et al., Chertova et al.	no
2fnsB	protease	2ikoA	REN	ATP6AP2	Chertova et al.	no

APPENDIX 3: Full prediction network



HIV-1 proteins that resemble human proteins are predicted to interact with the known interactors of the mimicked protein. The human proteins included in the prediction set have a supported role in HIV-1 infection or replication, either because they are incorporated into the HIV-1 virion or their reduced expression is known to prevent HIV-1 infection (node line color corresponds to source). Red lines represent predicted interactions that are already known to occur.

APPENDIX 4: Twenty-five selected interactions after the CC filter

HIV PDB code	HIV protein name	HIV-similar Human PDB code	HIV-similar Human Gene Symbol	Human interactor Gene Symbol	Source Datasets	True Positive
1ihwA	integrase	1zsgA	ARHGEF7	CDC42	Chertova et al.	no
1ihvA	integrase	2jxbA	NCK2	EGFR	Brass et al.	no
1sv5A	reverse transcriptase	1tceA	SHC1	PTPN6	Chertova et al.	no
1hihB	protease	1lyaB	CTSD	CD4	Brass et al., Zhou et al.	yes
1c1cA	reverse transcriptase	1qu6A	EIF2AK2	RAC1	Chertova et al.	no
1u57A	p2	1gk7A	VIM	DSP	Chertova et al.	no
2cmrA	gp41	2c1jA	YWHAZ	EGFR	Brass et al.	no
1suqA	reverse transcriptase	1ft8E	NXF1	NUP98	Konig et al., Yeung et al.	yes
1htfB	protease	1lyaB	CTSD	CD4	Brass et al., Zhou et al.	yes
1f23A	gp41	1usdA	VASP	ACTG1	Chertova et al.	no
1a8o	capsid	1r8qF	PSCD2	ITGB2	Chertova et al.	no
1f23C	gp41	1urfA	PKN1	CD44	Chertova et al.	no
3cp1A	gp41	3e1rA	CEP55	TSG101	Chertova et al., Cantin et al.	no
1qbtA	protease	1lyaA	CTSD	CD4	Brass et al., Zhou et al.	yes
1ihwA	integrase	1zsgA	ARHGEF7	RAC1	Chertova et al.	no
1rt1A	reverse transcriptase	1mb8A	PLEC1	SPTAN1	Brass et al.	yes
1dlbA	gp41	1hynQ	SLC4A1	SLC4A1	Yeung et al.	no
1qmcA	integrase	2jw4A	NCK1	RAC1	Chertova et al.	no
1a8o	capsid	1grnB	ARHGAP1	ARHGAP1	Chertova et al.	no
2vg5A	reverse transcriptase	1qu6A	EIF2AK2	RAC1	Chertova et al.	no
2b6aA	reverse transcriptase	1pbuA	EEF1G	ARF1	Brass et al., Chertova et al.	no
1a94B	protease	1lyaB	CTSD	CD4	Brass et al., Zhou et al.	yes
1u57A	p2	2q13A	APPL1	AKT1	Brass et al., Zhou et al.	no
1f23A	gp41	3brtD	IKBKG	TNFAIP3	Chertova et al.	no
2i5jA	reverse transcriptase	1ft8E	NXF1	NUP214	Konig et al.	yes

APPENDIX 5: HIV-1 protein identifiers

HIV protein name	HIV PDB code	HIV Uniprot
reverse transcriptase	1uwbB, 1ep4B, 1s6pB, 1s6pA, 1uwbA, 3hvtA, 1hniA, 1hvuJ, 1rt2A, 1hqeA, 1hysB, 1t03A, 1t03B, 1bqnB, 1lw2A, 1lw2B, 1c1bB, 1hnvA, 1har, 1c1bA, 1s1vB, 1s1vA, 1tkxA, 1s9eA, 1hnvB, 1tkxB, 1rt2B, 1n6qB, 2hmiA, 1rthA, 1tl3A, 2be2B, 2vg7B, 2be2A, 1dttA, 2i5jA, 1hvuD, 1hvuG, 1hniB, 2vg5A, 1ikyB, 1ikyA, 2vg5B, 1revA, 2vg6B, 1revB, 1hpzB, 1hvuK, 1hpzA, 1c0uB, 1c0uA, 1rtiA, 1rtiB, 1bqmB, 1vruA, 1lw0A, 1lw0B, 1s1tA, 1s1tB, 1dloA, 1dloB, 1s6qA, 1s6qB, 1sv5A, 1klmB, 1klmA, 1fk9B, 1fk9A, 1lweA, 1lweB, 1c0tA, 1c0tB, 1fkpA, 1rthB, 1ep4A, 1fkpB, 1rtjA, 1rtjB, 1r0aA, 1tktA, 1hquB, 1n6qA, 2hmiB, 1rt1A, 2b6aA, 1rt1B, 1vruB, 2vg7A, 1hquA, 1ikwA, 1ikwB, 2b5jA, 2banA, 1rtdB, 1bqnA, 1fkoA, 1fkoB, 1rtdC, 1ikxA, 1ikxB, 1dttB, 1lwcB, 1lwcA, 2zd1A, 1c1cA, 1c1cB, 1s1wB, 1vrtA, 1vrtB, 2hnzA, 1hysA, 1hmvB, 1hmvA, 1eetB, 1bqmA, 1rt3A, 1eetA, 1t05B, 1t05A, 1lwfA, 1lwfB, 2rkiA, 2rkiB, 2ze2B, 2ze2A, 1n5yA, 1n5yB, 1s9gA, 1dtqB, 1dtqA, 1s9gB, 1s9eB, 3bgrA, 3bgrB, 1j5oB, 1hvuA, 1j5oA, 1rtdD, 1tkzA, 1s1xA, 1rtdA, 1ikvB, 2vg6A, 1suqA, 1ikvA, 1tl1A	P03366, Q8UTX6, P04585, P03367
p2	1u57A	Q70622
gp41	3cp1A, 1ce0A, 1ce0B, 1qr9A, 1favA, 1k33A, 2cmrA, 1dlbA, 1df5A, 3cyoA, 1szt, 1i5xA, 1ce0C, 1qr8A, 1f23A, 1f23C, 1f23B, 1f23E, 1f23D, 1f23F, 1i5yA, 1k34A, 1df4A, 2ot5A	P03377, Q7SIH0, Q70626, Q89797, Q53I19, P04578, P04582, Q76270
protease	1aidB, 1aidA, 2fxdA, 2bpvB, 2bpvA, 2f81A, 1a9mA, 1hihB, 2qi3A, 2qi3B, 2pk5B, 2pk5A, 1hshC, 7hvpA, 1kjfA, 1hshA, 1hvjA, 1hvjB, 1upj, 2avvB, 2avvD, 2b7zA, 2r43B, 2pwcA, 2pwcB, 1dw6D, 1c6zB, 1bvgA, 1bvgB, 1ebwA, 1m0bA, 1a8gB, 1a8gA, 1b6kA, 1b6kB, 1mesB, 1hefE, 1hpsA, 1hpsB, 1hihA, 1c6xA, 2qi4A, 2qi4B, 1metB, 1a9mB, 1ytgB, 1ytgA, 2qnnA, 2qnnB, 1hosA, 1mt9A, 1mt9B, 1ztzB, 1d4yB, 1d4yA, 1hvkB, 1hpxA, 1hpxB, 1htgA, 1qbrA, 1htgB, 2i4uA, 1t3rA, 1t3rB, 1vijA, 1vijB, 2bpyB, 2bpyA, 1gnnB, 1fb7A, 1odwA, 1odwB, 5hvpB, 5hvpA, 1ythB, 1ythA, 3aidA, 3aidB, 2fgvA, 1k1uA, 1k6vA, 1k6vB, 1t7iA, 1t7iB, 1n49D, 1n49A, 1n49C, 1n49B, 1fgcD, 3bgcB, 1fgcC, 2qi0A, 1a94D, 1a94E, 1a94B, 1a94A, 1izhA, 1odxA, 1hvc, 1hvsB, 1hvsA, 1odxB, 1kj7B, 1kj7A, 1mtrB, 2fgvB, 1t7jA, 2qhzB, 1t7jB, 1d4sA, 1d4sB, 1b6jB, 1merA, 1merB, 1b6jA, 1hegE, 4phvA, 2qhzA, 2uy0A, 2qi5B, 1c6yA, 1kjfB, 2uy0B, 2fddA, 7upjB, 7upjA, 1meuA, 1meuB, 1ffiD, 1kjhB, 1kjhA, 1a30A, 1a30B, 2bpwA, 2bpwB, 2qnqA, 2qnqB, 1difB, 1difA, 1mtrA, 2fntB, 2qi7B,	Q9QM22, P35963, Q9Q288, Q903J0, O92103, Q6Q004, Q5RZ09, Q9WFL7, Q7SRY5, Q9Q2G8, Q90EB9, P04587, P04585, P04584, O38716, P12497, Q75002,

	2bqvA, 2p3aA, 1fffD, 1vikB, 1vikA, 2bpzB, 1xl2A, 2pqzA, 2pqzB, 2qmpA, 2qmpB, 2fguA, 1ohrB, 1ohrA, 2fguB, 2qi0B, 2cemA, 1hpvB, 1hpvA, 1fg6C, 1bdlA, 1bwaB, 2j9kB, 2qi1B, 2qi1A, 1qbtA, 1qbtB, 1iziA, 1bdrA, 1bdrB, 1wbkB, 1wbkA, 3hvp, 4hvpB, 4hvpA, 1qbsA, 1qbsB, 1htfB, 1htfA, 1hshD, 1bv7A, 1hshB, 7hvpB, 1kj4C, 1kj4B, 1kj4A, 1kj4D, 1tcxB, 1tcxA, 1gnmA, 1gnmB, 2a4fB, 1b6mB, 1b6mA, 2fnsB, 2fnsA, 2q63B, 2q63A, 1npvA, 1hviA, 1hviB, 1k1tA, 1hivB, 1hivA, 1axaA, 1axaB, 1c70B, 2avsA, 9hvpB, 9hvpA, 2qnpB, 2qnpA, 1k6pB, 1k6pA, 2i4wA, 2i4wB, 1mesA, 1hvrA, 1hvrB, 1bdqA, 1ec3A, 1bdqB, 1proB, 1proA, 1ec3B, 1c6xB, 2qhcA, 1bveB, 1bveA, 2qhcB, 1a8kD, 1a8kE, 1a8kB, 1a8kA, 1ebyB, 1hvlA, 1lv1A, 1t7kA, 2cenB, 1ec1B, 1ec1A, 1ajxA, 2qi6B, 2qi6A, 1ajxB, 1kjgB, 1kjgA, 1rq9A, 2fdeB, 1hosB, 1metA, 1ztzA, 1ebkF, 1ebkC, 1g2kB, 1g2kA, 1hteB, 1hxwA, 2j9kA, 1hxwB, 2aidA, 2aidB, 1bwaA, 1b6pA, 1b6pB, 1gnnA, 1sduA, 1hvkA, 1nh0B, 1nh0A, 1muiA, 1muiB, 1b6lA, 1b6lB, 1fqxB, 1fqxA, 1g35B, 2upjB, 2upjA, 1hvhB, 1mt8B, 1mt8A, 1hvhA, 2r43A, 2cejB, 1k6tB, 1k6tA, 1qbuB, 1bwbB, 1bwbA, 1qbuA, 2pk6B, 2pk6A, 1dmpA, 1dmpB, 1mrwA, 1mrwB, 1f7aB, 1f7aA, 2p3cA, 1msmB, 1msmA, 1hsgA, 1lzqA, 1k6cB, 1k6cA, 2bpxA, 2bpxB, 1rpiA, 1aaqA, 1aaqB, 1k2bA, 1ajvB, 1ajvA, 2p3dA, 1g6lA, 1hvlB, 1fejC, 1mt7B, 1mt7A, 1ec0A, 3bgbA, 3bgbB, 1hbvA, 1hbvB, 2qi7A, 2fntA, 8hvpB, 8hvpA, 2i4dA, 2i4dB, 2r3tB, 2r3tA, 2bbbA, 1mtbA, 1mtbB, 4phvB, 2pwrB, 1yt9A, 2pwrA, 1bv9B, 1bv9A, 2qhyB, 2qhyA, 2uxzB, 2uxzA, 1sguA, 1gnoB, 1gnoA, 1hpoA, 1hpoB, 1hsgB, 1hwrA, 1hwrB, 1iiqA, 2pynA, 1ebzA, 2qi5A, 2r5pA	O92139, Q8Q3J5, P12499, Q7SPG9, Q72874, P03369, P03366, P03367, Q5RTL1, O38907, O09893, Q6BB74, Q90HG9, Q8Q3H0
vpr	1vpc	Q73369
gp120	1g9nG, 2nxzA, 1rzkG	P35961, Q8QDX5
capsid, p2	1baj	P12497
capsid, matrix	1l6nA	Q72497
capsid	1aum, 1a8o, 2jygA, 2jylA, 3dphB, 1a43, 3dphA, 2k1cA, 2buoA	Q72497, P12497, P35963
integrase	1bi4C, 1bi4A, 1wjbB, 1bisB, 1bisA, 1wjbA, 1biuA, 1biuB, 1biuC, 1wjfA, 1wjfB, 1wjdB, 1wjdA, 1ihvA, 1ihvB, 1bl3C, 1bl3A, 2b4jB, 2b4jA, 1bizB, 1qmcB, 1qmcA, 1b9dA, 1wjcB, 1wjcA, 1wjaA, 1wjaB, 1wjeA, 1wjeB, 1qs4A, 1qs4B, 1qs4C, 1ihwB, 1ex4B, 1ex4A, 1ihwA, 1k6yA, 1k6yB, 1k6yC, 1k6yD, 1exqA, 1exqB	P35963, Q72498, P03366, Q76353, P12497, P04587, P04586, P04585
matrix	2jmgA, 2h3vA, 2h3zA, 1hiwQ, 1hiwR, 1hiwS, 2golA, 1uphA, 2h3iA, 2h3fA, 2hmx, 1hiwA, 1hiwB, 1hiwC, 2h3qA, 2nv3A	Q72497, P12493, P12497

APPENDIX 6: Twenty-five selected interaction predictions between DENV2 and human

Dengue Structure	Dengue protein	Similar Human PDB	Similar Human Gene Symbol	Interactor Gene Symbol	Host Factor	True Positive
1df9B	NS3	1danH	F7	F10	no	no
1df9B	NS3	1md7A	C1R	C1S	no	no
1df9B	NS3	1md7A	C1R	C1R	no	no
1df9B	NS3	1md7A	C1R	C1QB	no	no
1df9B	NS3	1md7A	C1R	C1QA	no	no
1df9B	NS3	1md7A	C1R	SERPING1	no	no
1df9B	NS3	1md7A	C1R	CSNK2A1	yes	no
1df9B	NS3	1c5mD	F10	MGST3	no	no
1df9B	NS3	1c5mD	F10	PLG	no	no
1df9B	NS3	1c5mD	F10	F3	no	no
1df9B	NS3	1c5mD	F10	F7	no	no
1df9B	NS3	1c5mD	F10	PROS1	no	no
1df9B	NS3	1c5mD	F10	APOH	no	no
1df9B	NS3	1c5mD	F10	PRKAB1	no	no
1df9B	NS3	1c5mD	F10	GAD2	no	no
1df9B	NS3	1c5mD	F10	SERPINB6	no	no
1df9B	NS3	1c5mD	F10	PLAT	no	no
1df9B	NS3	1c5mD	F10	HIST1H1C	no	no
1df9B	NS3	1c5mD	F10	F8	no	no
1df9B	NS3	1c5mD	F10	TFPI	no	no
1df9B	NS3	1c5mD	F10	GGCX	no	no
1df9B	NS3	1c5mD	F10	F5	no	no
1df9B	NS3	1c5mD	F10	EPR1	no	no
1df9B	NS3	1c5mD	F10	SERPINA10	no	no
1df9B	NS3	1c5mD	F10	F10	no	no

APPENDIX 7: Twenty-five selected interaction predictions between DENV2 and *A. aegypti*

Dengue Structure	Dengue protein	Similar fly PDB	Interactor Gene Symbol	Aedes Ortholog VectorBase	Host Factor	Predicted by Guo et al.
159024814	NS2A	1oxjA	YPS	AAEL001375-PA	no	no
159024814	NS2A	1oxjA	STC	AAEL001636-PA	no	no
159024814	NS2A	1oxjA	CG2807	AAEL003605-PA	no	no
159024814	NS2A	1oxjA	HEPH	AAEL013723-PA	no	no
159024814	NS2A	1oxjA	SC35	AAEL010340-PA	no	no
159024814	NS2A	1oxjA	CG10375	AAEL005070-PA	no	no
159024814	NS2A	1oxjA	CG4896	AAEL004989-PA	yes	no
159024814	NS2A	1oxjA	U2AF38	AAEL006713-PA	no	no
159024814	NS2A	1hx8A	EYA	AAEL005166-PA	no	no
159024814	NS2A	1hx8A	CHC	AAEL013614-PA	no	no
159024814	NS2A	1hx8A	CG3259	AAEL002173-PA	no	no
159024814	NS2A	1hx8A	CG5608	AAEL011389-PA	no	no
159024814	NS2A	1hx8A	SEC5	AAEL009926-PA	no	no
159024814	NS2A	1hx8A	SL	AAEL004431-PA	no	no
159024814	NS2A	1hx8A	BX42	AAEL014528-PA	no	no
159024814	NS2A	1hx8A	α-ADAPTIN	AAEL004469-PA	yes	no
159024814	NS2A	1hx8A	TOMOSYN	AAEL006948-PA	no	no
159024814	NS2A	1hx8A	CG33298	AAEL007689-PB	no	no
159024819	NS4B	2vxgA	DCP2	AAEL000783-PA AAEL015607-PA	no	no
159024819	NS4B	2vxgA	DCP1	AAEL001187-PA	no	no
159024819	NS4B	2vxgA	CG16728	AAEL009958-PA	no	no
159024819	NS4B	2vxgA	EDC3	AAEL008375-PA	no	no
159024819	NS4B	2vxgA	SUCB	AAEL005552-PA	no	no
159024819	NS4B	2vxgA	NMO	AAEL004797-PA	no	no
159024819	NS4B	2qvaC	IK2	AAEL005359-PA	no	no

APPENDIX 8: Twenty-five selected interaction predictions between DENV2 and human after the CC Filter

Dengue Structure	Dengue Protein	Similar Human PDB	Similar Human Gene Symbol	Interactor Gene Symbol	Host Factor	True Positive
159024813	NS1	1s1dB	CANT1	CANT1	no	no
159024813	NS1	2h2nB	CANT1	CANT1	no	no
159024813	NS1	1s1dA	CANT1	CANT1	no	no
159024813	NS1	1s18B	CANT1	CANT1	no	no
159024813	NS1	1v04A	PON1	CLU	no	yes
159024813	NS1	1v04A	PON1	APOA1	no	no
159024813	NS1	1v04A	PON1	ALB	no	no
159024813	NS1	2h2nA	CANT1	CANT1	no	no
159024813	NS1	2h2uA	CANT1	CANT1	no	no
159024813	NS1	1s18A	CANT1	CANT1	no	no
159024813	NS1	2h2uB	CANT1	CANT1	no	no
159024813	NS1	1ijqB	LDLR	LDLR	no	no
159024813	NS1	1ijqB	LDLR	APOB	no	no
159024813	NS1	1ijqB	LDLR	LRPAP1	no	no
159024813	NS1	1ijqB	LDLR	PF4	no	no
159024813	NS1	1ijqB	LDLR	AP1M2	no	no
159024813	NS1	1ijqB	LDLR	SNX17	no	no
159024813	NS1	1ijqB	LDLR	HSPA5	no	no
159024813	NS1	1ijqB	LDLR	LDLRAP1	no	no
159024813	NS1	2vdoA	ITGA2B	CIB1	no	no
159024813	NS1	2vdoA	ITGA2B	CALR	no	no
159024813	NS1	2vdoA	ITGA2B	FGA	no	no
159024813	NS1	1u6dX	KEAP1	MYO7A	no	no
159024813	NS1	2qnsA	GNB1	KCNJ3	no	no
159024813	NS1	2qnsA	GNB1	GNG3	no	no

APPENDIX 9: Twenty-five selected interaction predictions between DENV2 and *A. aegypti* after the CC Filter

Dengue Structure	Dengue Protein	Similar fly PDB	Similar fly Gene Symbol	Interactor Gene Symbol	Aedes Ortholog VectorBase	Host Factor	Predicted by Guo
159024814	NS2A	1dvpA	HRS	EGFR	AAEL004319-PA	no	no
159024814	NS2A	1dvpA	HRS	CHC	AAEL013614-PA	no	no
159024814	NS2A	1dvpA	HRS	GGA	AAEL001525-PA	no	no
159024814	NS2A	1dvpA	HRS	MER	AAEL006018-PB	no	no
159024814	NS2A	1hx8A	LAP	CHC	AAEL013614-PA	no	no
159024814	NS2A	1hx8A	LAP	α-ADAPTIN	AAEL004469-PA	yes	no
159024814	NS2A	1hx8A	LAP	TOMOSYN	AAEL006948-PA	no	no
159024814	NS2A	1hx8A	LAP	CG33298	AAEL007689-PB	no	no
159024819	NS4B	1dvpA	HRS	EGFR	AAEL004319-PA	no	no
159024819	NS4B	1dvpA	HRS	CHC	AAEL013614-PA	no	no
159024819	NS4B	1dvpA	HRS	GGA	AAEL001525-PA	no	no
159024819	NS4B	1dvpA	HRS	MER	AAEL006018-PB	no	no
1oanA	E	1cfbA	NRG	ED	AAEL004133-PA	no	no
1oanA	E	1cfbA	NRG	MOE	AAEL007915-PA	no	no
1oanA	E	1cfbA	NRG	IF	AAEL013600-PA	no	no
1oanB	E	1cfbA	NRG	ED	AAEL004133-PA	no	no
1oanB	E	1cfbA	NRG	MOE	AAEL007915-PA	no	no
1oanB	E	1cfbA	NRG	IF	AAEL013600-PA	no	no
1okeA	E	1cfbA	NRG	ED	AAEL004133-PA	no	no
1okeA	E	1cfbA	NRG	MOE	AAEL007915-PA	no	no
1okeA	E	1cfbA	NRG	IF	AAEL013600-PA	no	no
1okeB	E	1cfbA	NRG	ED	AAEL004133-PA	no	no
1okeB	E	1cfbA	NRG	MOE	AAEL007915-PA	no	no
1okeB	E	1cfbA	NRG	IF	AAEL013600-PA	no	no
1r6rA	C	1fjlA	PRD	MER	AAEL006018-PB	no	no

APPENDIX 10: Twenty-five selected interaction predictions between DENV2 and *A. aegypti* made using yeast orthologs

Dengue Structure	Dengue Protein	Similar PDB	Similar host	Aedes Ortholog Gene	Interactor Gene	Host Factor	Predicted by Guo
159024813	NS1	2hesX	yeast	AAEL002912	AAEL000270	no	no
159024813	NS1	2hesX	yeast	AAEL002912	AAEL006731	no	no
159024813	NS1	2hesX	yeast	AAEL002912	AAEL000440	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL001516	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL008391	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL006311	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL014534	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL004394	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL007484	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL003908	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL013464	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL002102	no	no
159024813	NS1	2pm7D	yeast	AAEL012240	AAEL007586	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL001516	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL008391	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL006311	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL014534	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL004394	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL007484	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL003908	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL013464	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL002102	no	no
159024813	NS1	2pm6B	yeast	AAEL012240	AAEL007586	no	no
159024813	NS1	2pm6D	yeast	AAEL012240	AAEL001516	no	no
159024813	NS1	2pm6D	yeast	AAEL012240	AAEL008391	no	no

APPENDIX 11: Predictions involving orthologous human and *A. aegypti* target proteins

Dengue Structure	Dengue Protein	Aedes Ortholog VectorBase	Fly Host Factor	Predicted by Guo	Human Gene Symbol	Human Host Factor	Human True Positive
159024814	NS2A	AAEL012515-PA	no	no	TSG101	no	no
159024814	NS2A	AAEL013614-PA	no	no	CLTC	no	no
159024814	NS2A	AAEL001525-PA	no	no	GGA3	no	no
159024814	NS2A	AAEL001525-PA	no	no	GGA2	no	no
159024814	NS2A	AAEL003877-PA	no	no	UBC	no	no
159024814	NS2A	AAEL006018-PB	no	no	NF2	no	no
159024814	NS2A	AAEL005356-PA	no	no	PAPOLA	no	no
159024814	NS2A	AAEL005515-PF	no	no	HNRPAB	no	no
159024814	NS2A	AAEL005515-PF	no	no	HNRPD	no	no
159024814	NS2A	AAEL012656-PD	no	no	SNRPA1	no	no
159024814	NS2A	AAEL007554-PA	yes	no	WDR77	no	no
159024814	NS2A	AAEL003670-PA	no	no	HNRPM	no	no
159024814	NS2A	AAEL010453-PA	no	no	MAGOH	no	no
159024814	NS2A	AAEL011544-PA	no	no	MAGOH	no	no
159024814	NS2A	AAEL002709-PA	no	no	SF3B3	no	no
159024814	NS2A	AAEL005947-PB	no	no	HNRPF	no	no
159024814	NS2A	AAEL008700-PA	no	no	SMN1	no	no
159024814	NS2A	AAEL003735-PA	no	no	UPF3B	no	no
159024814	NS2A	AAEL007578-PA	no	no	PRMT5	no	no
159024814	NS2A	AAEL015143-PB	no	no	PABPN1	no	no
159024814	NS2A	AAEL006135-PA	no	no	NCBP2	no	no
159024814	NS2A	AAEL010467-PA	no	no	HNRPA1	no	no
159024814	NS2A	AAEL011187-PA	no	no	ASCC3L1	no	no
159024819	NS4B	AAEL000783-PA	no	no	DCP2	no	no
159024819	NS4B	AAEL015607-PA	no	no	DCP2	no	no
159024819	NS4B	AAEL001187-PA	no	no	DCP1B	no	no
159024819	NS4B	AAEL009958-PA	no	no	GIT2	no	no
159024819	NS4B	AAEL008375-PA	no	no	EDC3	no	no
159024819	NS4B	AAEL005655-PA	no	no	SNX1	no	no
159024819	NS4B	AAEL003957-PA	no	no	CFL1	no	no
159024819	NS4B	AAEL010419-PA	no	no	TSNAX	no	no
159024819	NS4B	AAEL012515-PA	no	no	TSG101	no	no
159024819	NS4B	AAEL013614-PA	no	no	CLTC	no	no
159024819	NS4B	AAEL001525-PA	no	no	GGA3	no	no
159024819	NS4B	AAEL001525-PA	no	no	GGA2	no	no
159024819	NS4B	AAEL003877-PA	no	no	UBC	no	no

159024819	NS4B	AAEL006018-PB	no	no	NF2	no	no
1oanA	E	AAEL002829-PA	no	no	MACF1	no	no
1oanB	E	AAEL002829-PA	no	no	MACF1	no	no
1okeA	E	AAEL002829-PA	no	no	MACF1	no	no
1okeB	E	AAEL002829-PA	no	no	MACF1	no	no
1tg8A	E	AAEL002829-PA	no	no	MACF1	no	no
2jsfA	E	AAEL002829-PA	no	no	MACF1	no	no
2r29A	E	AAEL002829-PA	no	no	MACF1	no	no
2r69A	E	AAEL002829-PA	no	no	MACF1	no	no
3c5xA	E	AAEL002829-PA	no	no	MACF1	no	no
3c6eA	E	AAEL002829-PA	no	no	MACF1	no	no
1oanA	E	AAEL012083-PA	no	no	PTPRB	no	no
1oanB	E	AAEL012083-PA	no	no	PTPRB	no	no
1ok8A	E	AAEL012083-PA	no	no	PTPRB	no	no
1okeA	E	AAEL012083-PA	no	no	PTPRB	no	no
1okeB	E	AAEL012083-PA	no	no	PTPRB	no	no
1tg8A	E	AAEL012083-PA	no	no	PTPRB	no	no
2jsfA	E	AAEL012083-PA	no	no	PTPRB	no	no
2r29A	E	AAEL012083-PA	no	no	PTPRB	no	no
2r69A	E	AAEL012083-PA	no	no	PTPRB	no	no
3c5xA	E	AAEL012083-PA	no	no	PTPRB	no	no
3c6eA	E	AAEL012083-PA	no	no	PTPRB	no	no
1oanA	E	AAEL010443-PA	no	no	MAGI2	no	no
1oanB	E	AAEL010443-PA	no	no	MAGI2	no	no
1okeA	E	AAEL010443-PA	no	no	MAGI2	no	no
1okeB	E	AAEL010443-PA	no	no	MAGI2	no	no
1tg8A	E	AAEL010443-PA	no	no	MAGI2	no	no
2r29A	E	AAEL010443-PA	no	no	MAGI2	no	no
3c5xA	E	AAEL010443-PA	no	no	MAGI2	no	no
3c6eA	E	AAEL010443-PA	no	no	MAGI2	no	no
1tg8A	E	AAEL001476-PA	yes	no	NUMB	no	no
3c5xA	E	AAEL001476-PA	yes	no	NUMB	no	no
1oanA	E	AAEL006958-PA	yes	no	CNTN2	no	no
1oanB	E	AAEL006958-PA	yes	no	CNTN2	no	no
1okeA	E	AAEL006958-PA	yes	no	CNTN2	no	no
1okeB	E	AAEL006958-PA	yes	no	CNTN2	no	no
1tg8A	E	AAEL006958-PA	yes	no	CNTN2	no	no
2jsfA	E	AAEL006958-PA	yes	no	CNTN2	no	no
2r29A	E	AAEL006958-PA	yes	no	CNTN2	no	no
2r69A	E	AAEL006958-PA	yes	no	CNTN2	no	no
3c5xA	E	AAEL006958-PA	yes	no	CNTN2	no	no
3c6eA	E	AAEL006958-PA	yes	no	CNTN2	no	no
1oanA	E	AAEL006958-PA	yes	no	CNTN1	no	no

1oanB	E	AAEL006958-PA	yes	no	CNTN1	no	no
1ok8A	E	AAEL006958-PA	yes	no	CNTN1	no	no
1okeA	E	AAEL006958-PA	yes	no	CNTN1	no	no
1okeB	E	AAEL006958-PA	yes	no	CNTN1	no	no
1tg8A	E	AAEL006958-PA	yes	no	CNTN1	no	no
2jsfA	E	AAEL006958-PA	yes	no	CNTN1	no	no
2r29A	E	AAEL006958-PA	yes	no	CNTN1	no	no
2r69A	E	AAEL006958-PA	yes	no	CNTN1	no	no
3c5xA	E	AAEL006958-PA	yes	no	CNTN1	no	no
3c6eA	E	AAEL006958-PA	yes	no	CNTN1	no	no
1oanA	E	AAEL007915-PA	no	no	MSN	no	no
1oanB	E	AAEL007915-PA	no	no	MSN	no	no
1ok8A	E	AAEL007915-PA	no	no	MSN	no	no
1okeA	E	AAEL007915-PA	no	no	MSN	no	no
1okeB	E	AAEL007915-PA	no	no	MSN	no	no
1tg8A	E	AAEL007915-PA	no	no	MSN	no	no
2jsfA	E	AAEL007915-PA	no	no	MSN	no	no
2r29A	E	AAEL007915-PA	no	no	MSN	no	no
2r69A	E	AAEL007915-PA	no	no	MSN	no	no
3c5xA	E	AAEL007915-PA	no	no	MSN	no	no
3c6eA	E	AAEL007915-PA	no	no	MSN	no	no
1oanA	E	AAEL007915-PA	no	no	RDX	no	no
1ok8A	E	AAEL007915-PA	no	no	RDX	no	no
1tg8A	E	AAEL007915-PA	no	no	RDX	no	no
2jsfA	E	AAEL007915-PA	no	no	RDX	no	no
2r29A	E	AAEL007915-PA	no	no	RDX	no	no
2r69A	E	AAEL007915-PA	no	no	RDX	no	no
3c5xA	E	AAEL007915-PA	no	no	RDX	no	no
3c6eA	E	AAEL007915-PA	no	no	RDX	no	no
1oanA	E	AAEL013600-PA	no	no	ITGA8	no	no
1oanB	E	AAEL013600-PA	no	no	ITGA8	no	no
1ok8A	E	AAEL013600-PA	no	no	ITGA8	no	no
1okeA	E	AAEL013600-PA	no	no	ITGA8	no	no
1okeB	E	AAEL013600-PA	no	no	ITGA8	no	no
1tg8A	E	AAEL013600-PA	no	no	ITGA8	no	no
2jsfA	E	AAEL013600-PA	no	no	ITGA8	no	no
2r29A	E	AAEL013600-PA	no	no	ITGA8	no	no
2r69A	E	AAEL013600-PA	no	no	ITGA8	no	no
3c5xA	E	AAEL013600-PA	no	no	ITGA8	no	no
3c6eA	E	AAEL013600-PA	no	no	ITGA8	no	no
1oanA	E	AAEL013600-PA	no	no	ITGA5	no	no
1oanB	E	AAEL013600-PA	no	no	ITGA5	no	no
1ok8A	E	AAEL013600-PA	no	no	ITGA5	no	no

1okeA	E	AAEL013600-PA	no	no	ITGA5	no	no
1okeB	E	AAEL013600-PA	no	no	ITGA5	no	no
1tg8A	E	AAEL013600-PA	no	no	ITGA5	no	no
2jsfA	E	AAEL013600-PA	no	no	ITGA5	no	no
2r29A	E	AAEL013600-PA	no	no	ITGA5	no	no
2r69A	E	AAEL013600-PA	no	no	ITGA5	no	no
3c5xA	E	AAEL013600-PA	no	no	ITGA5	no	no
3c6eA	E	AAEL013600-PA	no	no	ITGA5	no	no

REFERENCES

Aloy, P., & Russell, R. B. (2003). InterPreTS: protein Interaction Prediction through Tertiary Structure. *Bioinformatics*, *19*(1), 161-162. doi:10.1093/bioinformatics/19.1.161

An, J., Zhou, D.-S., Kawasaki, K., & Yasui, K. (2003). The pathogenesis of spinal cord involvement in dengue virus infection. *Virchows Archiv*, *442*(5), 472-81. doi:10.1007/s00428-003-0785-3

Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., Bridge, A., Derow, C., Feuermann, M., et al. (2010). The IntAct molecular interaction database in 2010. *Nucleic Acids Research*, *38*(Database issue), D525-31. doi:10.1093/nar/gkp878

Arkin, M. R., & Wells, J. a. (2004). Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nature Reviews. Drug Discovery*, *3*(4), 301-17. doi:10.1038/nrd1343

Arold, S. T., Hoellerer, M. K., & Noble, M. E. . (2002). The Structural Basis of Localization and Signaling by the Focal Adhesion Targeting Domain. *Structure*, *10*(3), 319-327. doi:10.1016/S0969-2126(02)00717-7

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*(1), 25-9. doi:10.1038/75556

Ashour, J., Laurent-Rolle, M., Shi, P.-Y., & García-Sastre, A. (2009). NS5 of dengue virus mediates STAT2 binding and degradation. *Journal of Virology*, *83*(11), 5408-18. doi:10.1128/JVI.02188-08

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., et al. (2005). The Universal Protein Resource (UniProt). *Nucleic Acids Research*, *33*(Database issue), D154-9. doi:10.1093/nar/gki070

Berglund, A.-C., Sjölund, E., Ostlund, G., & Sonnhammer, E. L. L. (2008). InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Research*, *36*(Database issue), D263-6. doi:10.1093/nar/gkm1020

Berman, H., Henrick, K., & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, *10*(12), 980.

Bhuvanakantham, R., Li, J., Tan, T. T. T., & Ng, M.-L. (2010). Human Sec3 protein is a novel transcriptional and translational repressor of flavivirus. *Cellular Microbiology*, *12*(4), 453-72. doi:10.1111/j.1462-5822.2009.01407.x

Bielefeldt-Ohmann, H., Meyer, M., Fitzpatrick, D. R., & Mackenzie, J. S. (2001). Dengue virus binding to human leukocyte cell lines: receptor usage differs between cell types and virus strains. *Virus Research*, *73*(1), 81-89. doi:10.1016/S0168-1702(00)00233-1

Birringer, M. S., Claus, M. T., Folkers, G., Kloer, D. P., Schulz, G. E., & Scapozza, L. (2005). Structure of a type II thymidine kinase with bound dTTP. *FEBS Letters*, *579*(6), 1376-82. doi:10.1016/j.febslet.2005.01.034

Bouraï, M., Lucas-Hourani, M., Gad, H. H., Drosten, C., Jacob, Y., Tafforeau, L., Cassonnet, P., et al. (2012). Mapping of Chikungunya virus interactions with host proteins identified nsP2 as a highly connected viral component. *Journal of Virology*, *86*(6), 3121-34. doi:10.1128/JVI.06390-11

Brass, A. L., Dykxhoorn, D. M., Benita, Y., Yan, N., Engelman, A., Xavier, R. J., Lieberman, J., et al. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. *Science*, *319*(5865), 921-6. doi:10.1126/science.1152725

Calderwood, M. A., Venkatesan, K., Xing, L., Chase, M. R., Vazquez, A., Holthaus, A. M., Ewence, A. E., et al. (2007). Epstein-Barr virus and virus human protein interaction maps. *Proceedings of the National Academy of Sciences*, *104*(18), 7606-11. doi:10.1073/pnas.0702332104

Cantin, R., Méthot, S., & Tremblay, M. J. (2005). Plunder and stowaways: incorporation of cellular proteins by enveloped viruses. *Journal of Virology*, *79*(11), 6577-87. doi:10.1128/JVI.79.11.6577-6587.2005

Carlton, J. G., & Martin-Serrano, J. (2007). Parallels between cytokinesis and retroviral budding: a role for the ESCRT machinery. *Science*, *316*(5833), 1908-12. doi:10.1126/science.1143422

Chamberlain, P. P., Ren, J., Nichols, C. E., Douglas, L., Lennerstrand, J., Larder, B. A., Stuart, D. I., et al. (2002). Crystal Structures of Zidovudine- or Lamivudine-Resistant Human Immunodeficiency Virus Type 1 Reverse Transcriptases Containing Mutations at Codons 41, 184, and 215. *Journal of Virology*, *76*(19), 10015-10019. doi:10.1128/JVI.76.19.10015-10019.2002

Chang, C. J., Luh, H. W., Wang, S. H., Lin, H. J., Lee, S. C., & Hu, S. T. (2001). The heterogeneous nuclear ribonucleoprotein K (hnRNP K) interacts with dengue virus core protein. *DNA and Cell Biology*, *20*(9), 569-77. doi:10.1089/104454901317094981

Chee, H.-Y., & AbuBakar, S. (2004). Identification of a 48kDa tubulin or tubulin-like C6/36 mosquito cells protein that binds dengue virus 2 using mass spectrometry.

Biochemical and Biophysical Research Communications, *320*(1), 11-7. doi:10.1016/j.bbrc.2004.05.124

Chen, J. C., Krucinski, J., Miercke, L. J., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., & Stroud, R. M. (2000). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proceedings of the National Academy of Sciences*, *97*(15), 8233-8. doi:10.1073/pnas.150220297

Chen, Y.-C., Wang, S.-Y., & King, C.-C. (1999). Bacterial Lipopolysaccharide Inhibits Dengue Virus Infection of Primary Human Monocytes/Macrophages by Blockade of Virus Entry via a CD14-Dependent Mechanism. *Journal of Virology*, *73*(4), 2650-2657.

Chertova, E., Chertov, O., Coren, L. V., Roser, J. D., Trubey, C. M., Bess, J. W., Sowder, R. C., et al. (2006). Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages. *Journal of Virology*, *80*(18), 9039-52. doi:10.1128/JVI.01013-06

Chiu, M.-W., Shih, H.-M., Yang, T.-H., & Yang, Y.-L. (2007). The type 2 dengue virus envelope protein interacts with small ubiquitin-like modifier-1 (SUMO-1) conjugating enzyme 9 (Ubc9). *Journal of Biomedical Science*, *14*(3), 429-44. doi:10.1007/s11373-007-9151-9

Chua, J. J. E., Ng, M. M. L., & Chow, V. T. K. (2004). The non-structural 3 (NS3) protein of dengue virus type 2 interacts with human nuclear receptor binding protein and is associated with alterations in membrane structure. *Virus Research*, *102*(2), 151-63. doi:10.1016/j.virusres.2004.01.025

Chua, J. J.-E., Bhuvanakantham, R., Chow, V. T.-K., & Ng, M.-L. (2005). Recombinant non-structural 1 (NS1) protein of dengue-2 virus interacts with human STAT3beta protein. *Virus Research*, *112*(1-2), 85-94. doi:10.1016/j.virusres.2005.03.025

Crosby, M. A., Goodman, J. L., Strelets, V. B., Zhang, P., & Gelbart, W. M. (2007). FlyBase: genomes by the dozen. *Nucleic Acids Research*, *35*(Database issue), D486-91. doi:10.1093/nar/gkl827

Davis, F. P., Barkan, D. T., Eswar, N., McKerrow, J. H., & Sali, A. (2007). Host pathogen protein interactions predicted by comparative modeling. *Protein Science*, *16*(12), 2585-96. doi:10.1110/ps.073228407

Davis, F. P., Braberg, H., Shen, M.-Y., Pieper, U., Sali, A., & Madhusudhan, M. S. (2006). Protein complex compositions predicted by structural similarity. *Nucleic Acids Research*, *34*(10), 2943-52. doi:10.1093/nar/gkl353

De Chassey, B., Navratil, V., Tafforeau, L., Hiet, M. S., Aublin-Gex, A., Agaugué, S., Meiffren, G., et al. (2008). Hepatitis C virus infection protein network. *Molecular Systems Biology*, *4*(230), 230. doi:10.1038/msb.2008.66

De Nova-Ocampo, M., Villegas-Sepúlveda, N., & del Angel, R. M. (2002). Translation elongation factor-1alpha, La, and PTB interact with the 3' untranslated region of dengue 4 virus RNA. *Virology*, *295*(2), 337-47. doi:10.1006/viro.2002.1407

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. San Carlos, CA: DeLano Scientific.

Del Real, G., Jiménez-Baranda, S., Mira, E., Lacalle, R. A., Lucas, P., Gómez-Moutón, C., Alegret, M., et al. (2004). Statins inhibit HIV-1 infection by down-regulating Rho activity. *The Journal of Experimental Medicine*, *200*(4), 541-7. doi:10.1084/jem.20040061

Dennis, G., Sherman, B., Hosack, D., Yang, J., Gao, W., Lane, H., & Lempicki, R. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, *4*(9), R60. doi:10.1186/gb-2003-4-9-r60

Deretic, V., Vergne, I., Chua, J., Master, S., Singh, S. B., Fazio, J. A., & Kyei, G. (2004). Endosomal membrane traffic: convergence point targeted by Mycobacterium tuberculosis and HIV. *Cellular Microbiology*, *6*(11), 999-1009. doi:10.1111/j.1462-5822.2004.00449.x

Diamond, M. S., & Harris, E. (2001). Interferon inhibits dengue virus infection by preventing translation of viral RNA through a PKR-independent mechanism. *Virology*, *289*(2), 297-311. doi:10.1006/viro.2001.1114

Doolittle, J. M., & Gomez, S. M. (2010). Structural similarity-based predictions of protein interactions between HIV-1 and Homo sapiens. *Virology Journal*, *7*, 82. doi:10.1186/1743-422X-7-82

Dorr, P., Westby, M., Dobbs, S., Griffin, P., Irvine, B., Macartney, M., Mori, J., et al. (2005). Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrobial Agents and Chemotherapy*, *49*(11), 4721-4732. doi:10.1128/AAC.49.11.4721

Driscoll, T., Dyer, M. D., Murali, T. M., & Sobral, B. W. (2009). PIG--the pathogen interaction gateway. *Nucleic Acids Research*, *37*(Database issue), D647-50. doi:10.1093/nar/gkn799

Duan, X., Lu, X., Li, J., & Liu, Y. (2008). Novel binding between pre-membrane protein and vacuolar ATPase is required for efficient dengue virus secretion. *Biochemical*

and Biophysical Research Communications, *373*(2), 319-24. doi:10.1016/j.bbrc.2008.06.041

Duarte dos Santos, C. N., Frenkiel, M. P., Courageot, M. P., Rocha, C. F., Vazeille-Falcoz, M. C., Wien, M. W., Rey, F. A., et al. (2000). Determinants in the envelope E protein and viral RNA helicase NS3 that influence the induction of apoptosis in response to infection with dengue type 1 virus. *Virology*, *274*(2), 292-308. doi:10.1006/viro.2000.0457

Dyer, M. D., Murali, T. M., & Sobral, B. W. (2007). Computational prediction of host-pathogen protein-protein interactions. *Bioinformatics*, *23*(13), i159-66. doi:10.1093/bioinformatics/btm208

Dyer, M. D., Murali, T. M., & Sobral, B. W. (2008). The landscape of human proteins interacting with viruses and other pathogens. *PLoS Pathogens*, *4*(2), e32. doi:10.1371/journal.ppat.0040032

Dyer, M. D., Murali, T. M., & Sobral, B. W. (2011). Supervised learning and prediction of physical interactions between human and HIV proteins. *Infection, Genetics and Evolution*, *11*(5), 917-23. doi:10.1016/j.meegid.2011.02.022

Ellencrona, K., Syed, A., & Johansson, M. (2009). Flavivirus NS5 associates with host-cell proteins zonula occludens-1 (ZO-1) and regulating synaptic membrane exocytosis-2 (RIMS2) via an internal PDZ binding mechanism. *Biological Chemistry*, *390*(4), 319-323.

Evans, P., Dampier, W., Ungar, L., & Tozeren, A. (2009). Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs. *BMC Medical Genomics*, *2*(1), 27. BioMed Central. doi:10.1186/1755-8794-2-27

Falconar, A. K. I. (1997). The dengue virus nonstructural-1 protein (NS1) generates antibodies to common epitopes on human blood clotting, integrin/adhesin proteins and binds to human endothelial cells: potential implications in haemorrhagic fever pathogenesis. *Archives of Virology*, *142*(5), 897-916. doi:10.1007/s007050050127

Flicek, P., Aken, B. L., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., et al. (2010). Ensembl's 10th year. *Nucleic Acids Research*, *38*(Database issue), D557-62. doi:10.1093/nar/gkp972

Folly, B. B., Weffort-Santos, A. M., Fathman, C. G., & Soares, L. R. B. (2011). Dengue-2 structural proteins associate with human proteins to produce a coagulation and innate immune response biased interactome. *BMC Infectious Diseases*, *11*(1), 34. doi:10.1186/1471-2334-11-34

Frankel, A. D., & Young, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annual Review of Biochemistry*, *67*, 1-25. doi:10.1146/annurev.biochem.67.1.1

Fu, W., Sanders-Beer, B. E., Katz, K. S., Maglott, D. R., Pruitt, K. D., & Ptak, R. G. (2009). Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Research*, *37*(Database issue), D417-22. doi:10.1093/nar/gkn708

Gale, M., Blakely, C. M., Kwieciszewski, B., Tan, S.-L., Dossett, M., Tang, N. M., Korth, M. J., et al. (1998). Control of PKR Protein Kinase by Hepatitis C Virus Nonstructural 5A Protein: Molecular Mechanisms of Kinase Regulation. *Molecular and Cellular Biology*, *18*(9), 5208-5218.

Garcia-Alvarez, B., Bobkov, A., Sonnenberg, A., & de Pereda, J. (2003). Structural and Functional Analysis of the Actin Binding Domain of Plectin Suggests Alternative Mechanisms for Binding to F-Actin and Integrin β4. *Structure*, *11*(6), 615-625. doi:10.1016/S0969-2126(03)00090-X

García-Montalvo, B. M., Medina, F., & 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 Research*, *102*(2), 141-50. doi:10.1016/j.virusres.2004.01.024

Girard, Y. A., Popov, V., Wen, J., Han, V., & Higgs, S. (2005, May). Ultrastructural Study of West Nile Virus Pathogenesis in Culex pipiens quinquefasciatus (Diptera: Culicidae). *Journal of Medical Entomology*, *42*(3), 429-444. doi:10.1603/0022-2585(2005)042[0429:USOWNV]2.0.CO;2

Goff, S. P. (2007). Host factors exploited by retroviruses. *Nature Reviews. Microbiology*, *5*(4), 253-63. doi:10.1038/nrmicro1541

Gougeon, M.-L. (2003). Apoptosis as an HIV strategy to escape immune attack. *Nature Reviews. Immunology*, *3*(5), 392-404. doi:10.1038/nri1087

Guo, X., Xu, Y., Bian, G., Pike, A. D., Xie, Y., & Xi, Z. (2010). Response of the mosquito protein interaction network to dengue infection. *BMC Genomics*, *11*(1), 380. doi:10.1186/1471-2164-11-380

Hamamoto, S., Nishitsuji, H., Amagasa, T., Kannagi, M., & Masuda, T. (2006). Identification of a novel human immunodeficiency virus type 1 integrase interactor, Gemin2, that facilitates efficient viral cDNA synthesis in vivo. *Journal of Virology*, *80*(12), 5670-7. doi:10.1128/JVI.02471-05

Ho, L.-J., Hung, L.-F., Weng, C.-Y., Wu, W.-L., Chou, P., Lin, Y.-L., Chang, D.-M., et al. (2005). Dengue Virus Type 2 Antagonizes IFN-alpha but Not IFN-gamma Antiviral Effect via Down-Regulating Tyk2-STAT Signaling in the Human Dendritic Cell. *Journal of Immunology*, *174*(12), 8163-8172.

Holm, L., Kääriäinen, S., Rosenström, P., & Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics*, *24*(23), 2780-1. doi:10.1093/bioinformatics/btn507

Holm, L., & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *Journal of Molecular Biology*, *233*(1), 123-38. doi:10.1006/jmbi.1993.1489

Holm, L., & Sander, C. (1998). Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics*, *14*(5), 423-429. doi:10.1093/bioinformatics/14.5.423

Hopkins, A. L., Ren, J., Milton, J., Hazen, R. J., Chan, J. H., Stuart, D. I., & Stammers, D. K. (2004). Design of non-nucleoside inhibitors of HIV-1 reverse transcriptase with improved drug resistance properties. *Journal of Medicinal Chemistry*, *47*(24), 5912-22. doi:10.1021/jm040071z

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44-57. doi:10.1038/nprot.2008.211

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N. J., Chung, S., Emili, A., et al. (2003). A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science*, *302*(5644), 449-53. doi:10.1126/science.1087361

Janssens, V., Goris, J., & Van Hoof, C. (2005). PP2A: the expected tumor suppressor. *Current Opinion in Genetics & Development*, *15*(1), 34-41. doi:10.1016/j.gde.2004.12.004

Jiang, L., Yao, H., Duan, X., Lu, X., & Liu, Y. (2009). Polypyrimidine tract-binding protein influences negative strand RNA synthesis of dengue virus. *Biochemical and Biophysical Research Communications*, *385*(2), 187-92. doi:10.1016/j.bbrc.2009.05.036

Jindadamrongwech, S., Thepparit, C., & Smith, D. R. (2004). Identification of GRP 78 (BiP) as a liver cell expressed receptor element for dengue virus serotype 2. *Archives of Virology*, *149*(5), 915-27. doi:10.1007/s00705-003-0263-x

Johansson, M., Brooks, A. J., Jans, D. A., & Vasudevan, S. G. (2001). A small region of the dengue virus-encoded RNA-dependent RNA polymerase, NS5, confers interaction with both the nuclear transport receptor importin-beta and the viral helicase, NS3. *The Journal of General Virology*, *82*(Pt 4), 735-45.

Jones, M., Davidson, A., Hibbert, L., Gruenwald, P., Schlaak, J., Ball, S., Foster, G. R., et al. (2005). Dengue virus inhibits alpha interferon signaling by reducing STAT2

expression. *Journal of Virology*, *79*(9), 5414-20. doi:10.1128/JVI.79.9.5414-5420.2005

Khadka, S., Vangeloff, A. D., Zhang, C., Siddavatam, P., Heaton, N. S., Wang, L., Sengupta, R., et al. (2011). A physical interaction network of dengue virus and human proteins. *Molecular & Cellular Proteomics*, *10*(12), M111.012187. doi:10.1074/mcp.M111.012187

Korkin, D., Thieu, T., Joshi, S., & Warren, S. (2011). Mining Host-Pathogen Interactions. In N.-S. Yang (Ed.), *Systems and Computational Biology - Molecular and Cellular Experimental Systems* (pp. 163-184). InTech. doi:10.5772/22016

Krishnan, M. N., Ng, A., Sukumaran, B., Gilfoy, F. D., Uchil, P. D., Sultana, H., Brass, A. L., et al. (2008). RNA interference screen for human genes associated with West Nile virus infection. *Nature*, *455*(7210), 242-5. doi:10.1038/nature07207

Kroeger, A., & Nathan, M. B. (2006). Dengue: setting the global research agenda. *Lancet*, *368*(9554), 2193-5. doi:10.1016/S0140-6736(06)69873-5

Kurane, I., Innis, B. L., Nimmannitya, S., Nisalak, A., Meager, A., & Ennis, F. A. (1993). High levels of interferon alpha in the sera of children with dengue virus infection. *The American Journal of Tropical Medicine and Hygiene*, *48*(2), 222-9.

Kurosu, T., Chaichana, P., Yamate, M., Anantapreecha, S., & Ikuta, K. (2007). Secreted complement regulatory protein clusterin interacts with dengue virus nonstructural protein 1. *Biochemical and Biophysical Research Communications*, *362*(4), 1051-6. doi:10.1016/j.bbrc.2007.08.137

König, R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M. C., Irelan, J. T., Chiang, C.-Y., et al. (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*, *135*(1), 49-60. doi:10.1016/j.cell.2008.07.032

Le Breton, M., Meyniel-Schicklin, L., Deloire, A., Coutard, B., Canard, B., de Lamballerie, X., Andre, P., et al. (2011, January). Flavivirus NS3 and NS5 proteins interaction network: a high-throughput yeast two-hybrid screen. *BMC Microbiology*. doi:10.1186/1471-2180-11-234

Lee, S.-A., Chan, C.-h, Tsai, C.-H., Lai, J.-M., Wang, F.-S., Kao, C.-Y., & Huang, C.-Y. F. (2008). Ortholog-based protein-protein interaction prediction and its application to inter-species interactions. *BMC Bioinformatics*, *9*(Suppl 12), S11. doi:10.1186/1471-2105-9-S12-S11

Liker, E., Fernandez, E., Izaurralde, E., & Conti, E. (2000). The structure of the mRNA export factor TAP reveals a cis arrangement of a non-canonical RNP domain and an LRR domain. *The EMBO Journal*, *19*(21), 5587-98. doi:10.1093/emboj/19.21.5587

Limjindaporn, T., Netsawang, J., Noisakran, S., Thiemmeca, S., Wongwiwat, W., Sudsaward, S., Avirutnan, P., et al. (2007). Sensitization to Fas-mediated apoptosis by dengue virus capsid protein. *Biochemical and Biophysical Research Communications*, *362*(2), 334-9. doi:10.1016/j.bbrc.2007.07.194

Limjindaporn, T., Wongwiwat, W., Noisakran, S., Srisawat, C., Netsawang, J., Puttikhunt, C., Kasinrerk, W., et al. (2009). Interaction of dengue virus envelope protein with endoplasmic reticulum-resident chaperones facilitates dengue virus production. *Biochemical and Biophysical Research Communications*, *379*(2), 196-200. doi:10.1016/j.bbrc.2008.12.070

Liu, C. Y., & Kaufman, R. J. (2003). The unfolded protein response. *Journal of Cell Science*, *116*(Pt 10), 1861-2. doi:10.1242/jcs.00408

Lozach, P.-Y., Burleigh, L., Staropoli, I., Navarro-Sanchez, E., Harriague, J., Virelizier, J.-L., Rey, F. A., et al. (2005). Dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN)-mediated enhancement of dengue virus infection is independent of DC-SIGN internalization signals. *The Journal of Biological Chemistry*, *280*(25), 23698-708. doi:10.1074/jbc.M504337200

Lu, L., Lu, H., & Skolnick, J. (2002). MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading. *Proteins*, *49*(3), 350-64. doi:10.1002/prot.10222

Luftig, M. A., Mattu, M., Di Giovine, P., Geleziunas, R., Hrin, R., Barbato, G., Bianchi, E., et al. (2006). Structural basis for HIV-1 neutralization by a gp41 fusion intermediate-directed antibody. *Nature Structural & Molecular Biology*, *13*(8), 740-7. Nature Publishing Group. doi:10.1038/nsmb1127

MacPherson, J. I., Dickerson, J. E., Pinney, J. W., & Robertson, D. L. (2010). Patterns of HIV-1 protein interaction identify perturbed host-cellular subsystems. *PLoS Computational Biology*, *6*(7), e1000863. doi:10.1371/journal.pcbi.1000863

Matarrese, P., & Malorni, W. (2005). Human immunodeficiency virus (HIV)-1 proteins and cytoskeleton: partners in viral life and host cell death. *Cell Death and Differentiation*, *12 Suppl 1*(5), 932-41. doi:10.1038/sj.cdd.4401582

Mattera, R., Ritter, B., Sidhu, S. S., McPherson, P. S., & Bonifacino, J. S. (2004). Definition of the consensus motif recognized by gamma-adaptin ear domains. *The Journal of Biological Chemistry*, *279*(9), 8018-28. doi:10.1074/jbc.M311873200

Maulik, U., Bhattacharyya, M., Mukhopadhyay, A., & Bandyopadhyay, S. (2011). Identifying the immunodeficiency gateway proteins in humans and their involvement in microRNA regulation. *Molecular BioSystems*, *7*(6), 1842-51. doi:10.1039/c1mb05026e

McCarthy, F. M., Wang, N., Magee, G. B., Nanduri, B., Lawrence, M. L., Camon, E. B., Barrell, D. G., et al. (2006). AgBase: a functional genomics resource for agriculture. *BMC Genomics*, *7*(1), 229. doi:10.1186/1471-2164-7-229

Mhawech, P. (2005). 14-3-3 proteins--an update. *Cell Research*, *15*(4), 228-36. doi:10.1038/sj.cr.7290291

Mishra, G. R., Suresh, M., Kumaran, K., Kannabiran, N., Suresh, S., Bala, P., Shivakumar, K., et al. (2006). Human protein reference database--2006 update. *Nucleic Acids Research*, *34*(Database issue), D411-4. doi:10.1093/nar/gkj141

Mohamed, M. R., Rahman, M. M., Lanchbury, J. S., Shattuck, D., Neff, C., Dufford, M., van Buuren, N., et al. (2009). Proteomic screening of variola virus reveals a unique NF-kappaB inhibitor that is highly conserved among pathogenic orthopoxviruses. *Proceedings of the National Academy of Sciences*, *106*(22), 9045-50. doi:10.1073/pnas.0900452106

Morris, T. L., Arnold, R. R., & Webster-Cyriaque, J. (2007). Signaling cascades triggered by bacterial metabolic end products during reactivation of Kaposi's sarcoma-associated herpesvirus. *Journal of Virology*, *81*(11), 6032-42. doi:10.1128/JVI.02504-06

Mukhopadhyay, A., Maulik, U., & Bandyopadhyay, S. (2012). A Novel Biclustering Approach to Association Rule Mining for Predicting HIV-1-Human Protein Interactions. *PloS ONE*, *7*(4), e32289. doi:10.1371/journal.pone.0032289

Muñoz-Jordan, J. L., Sánchez-Burgos, G. G., Laurent-Rolle, M., & García-Sastre, A. (2003). Inhibition of interferon signaling by dengue virus. *Proceedings of the National Academy of Sciences*, *100*(24), 14333-8. doi:10.1073/pnas.2335168100

NIAID. (2012). Biodefense and Emerging Infectious Diseases. NIAID Category A, B, and C Priority Pathogens. Retrieved November 5, 2012, from http://www.niaid.nih.gov/topics/BiodefenseRelated/Biodefense/Pages/CatA.aspx

Noisakran, S., Sengsai, S., Thongboonkerd, V., Kanlaya, R., Sinchaikul, S., Chen, S.-T., Puttikhunt, C., et al. (2008). Identification of human hnRNP C1/C2 as a dengue virus NS1-interacting protein. *Biochemical and Biophysical Research Communications*, *372*(1), 67-72. doi:10.1016/j.bbrc.2008.04.165

Nouretdinov, I., Gammerman, A., & Qi, Y. (2012). Determining confidence of predicted interactions between HIV-1 and human proteins using conformal method. *Pacific Symposium on Biocomputing*, 311-322.

Pattanakitsakul, S.-N., Rungrojcharoenkit, K., Kanlaya, R., Sinchaikul, S., Noisakran, S., Chen, S.-T., Malasit, P., et al. (2007). Proteomic analysis of host responses in

HepG2 cells during dengue virus infection. *Journal of Proteome Research*, *6*(12), 4592-600. doi:10.1021/pr070366b

Pellegrini, L., Yu, D. S., Lo, T., Anand, S., Lee, M., Blundell, T. L., & Venkitaraman, A. R. (2002). Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature*, *420*(6913), 287-93. doi:10.1038/nature01230

Perera, R., & Kuhn, R. J. (2008). Structural proteomics of dengue virus. *Current Opinion in Microbiology*, *11*(4), 369-77. doi:10.1016/j.mib.2008.06.004

Persidsky, Y., Heilman, D., Haorah, J., Zelivyanskaya, M., Persidsky, R., Weber, G. A., Shimokawa, H., et al. (2006). Rho-mediated regulation of tight junctions during monocyte migration across the blood-brain barrier in HIV-1 encephalitis (HIVE). *Blood*, *107*(12), 4770-80. doi:10.1182/blood-2005-11-4721

Platt, K. B., Linthicum, K. J., Myint, K. S., Innis, B. L., Lerdthusnee, K., & Vaughn, D. W. (1997). Impact of dengue virus infection on feeding behavior of Aedes aegypti. *The American Journal of Tropical Medicine and Hygiene*, *57*(2), 119-25.

Pornillos, O., Higginson, D. S., Stray, K. M., Fisher, R. D., Garrus, J. E., Payne, M., He, G.-P., et al. (2003). HIV Gag mimics the Tsg101-recruiting activity of the human Hrs protein. *The Journal of Cell Biology*, *162*(3), 425-34. doi:10.1083/jcb.200302138

Qi, Y., Klein-Seetharaman, J., & Bar-Joseph, Z. (2007). A mixture of feature experts approach for protein-protein interaction prediction. *BMC Bioinformatics*, *8*(Suppl 10), S6. doi:10.1186/1471-2105-8-S10-S6

Qi, Y., Tastan, O., Carbonell, J. G., Klein-Seetharaman, J., & Weston, J. (2010). Semi-supervised multi-task learning for predicting interactions between HIV-1 and human proteins. *Bioinformatics*, *26*(18), i645-52. doi:10.1093/bioinformatics/btq394

Raiborg, C., Bremnes, B., Mehlum, A., Gillooly, D. J., D'Arrigo, A., Stang, E., & Stenmark, H. (2001). FYVE and coiled-coil domains determine the specific localisation of Hrs to early endosomes. *Journal of Cell Science*, *114*(12), 2255-2263.

Ramanathan, M. P., Chambers, J. A., Pankhong, P., Chattergoon, M., Attatippaholkun, W., Dang, K., Shah, N., et al. (2006). Host cell killing by the West Nile Virus NS2B-NS3 proteolytic complex: NS3 alone is sufficient to recruit caspase-8-based apoptotic pathway. *Virology*, *345*(1), 56-72. doi:10.1016/j.virol.2005.08.043

Rawlinson, S. M., Pryor, M. J., Wright, P. J., & Jans, D. A. (2009). CRM1-mediated nuclear export of dengue virus RNA polymerase NS5 modulates interleukin-8 induction and virus production. *The Journal of Biological Chemistry*, *284*(23), 15589-97. doi:10.1074/jbc.M808271200

Reyes-Del Valle, J., Chávez-Salinas, S., Medina, F., & Del Angel, R. M. (2005). Heat shock protein 90 and heat shock protein 70 are components of dengue virus receptor complex in human cells. *Journal of Virology*, *79*(8), 4557-67. doi:10.1128/JVI.79.8.4557-4567.2005

Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R., & Shi, Y. (2005, April 14). Structure of the apoptotic protease-activating factor 1 bound to ADP. *Nature*. doi:10.1038/nature03465

Ritchie, M. E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A., & Smyth, G. K. (2007). A comparison of background correction methods for two-colour microarrays. *Bioinformatics*, *23*(20), 2700-7. doi:10.1093/bioinformatics/btm412

Salas-Benito, J., Valle, J. R.-D., Salas-Benito, M., Ceballos-Olvera, I., Mosso, C., & del Angel, R. M. (2007). Evidence that the 45-kD Glycoprotein, Part of a Putative Dengue Virus Receptor Complex in the Mosquito Cell Line C6/36, Is a Heat-Shock Related Protein. *The American Journal of Tropical Medicine and Hygiene*, *77*(2), 283-290.

Sanchez-Vargas, I., Travanty, E. A., Keene, K. M., Franz, A. W. E., Beaty, B. J., Blair, C. D., & Olson, K. E. (2004). RNA interference, arthropod-borne viruses, and mosquitoes. *Virus Research*, *102*(1), 65-74. doi:10.1016/j.virusres.2004.01.017

Saphire, A. C. S., Gallay, P. A., & Bark, S. J. (2006). Proteomic analysis of human immunodeficiency virus using liquid chromatography/tandem mass spectrometry effectively distinguishes specific incorporated host proteins. *Journal of Proteome Research*, *5*(3), 530-8. doi:10.1021/pr050276b

Scaria, V., Hariharan, M., Maiti, S., Pillai, B., & Brahmachari, S. K. (2006). Host-virus interaction: a new role for microRNAs. *Retrovirology*, *3*, 68. doi:10.1186/1742-4690-3-68

Schang, L. M. (2006). First demonstration of the effectiveness of inhibitors of cellular protein kinases in antiviral therapy. *Expert Review of Anti-infective Therapy*, *4*(6), 953-956.

Schleker, S., Sun, J., Raghavan, B., Srnec, M., Müller, N., Koepfinger, M., Murthy, L., et al. (2011). The current Salmonella-host interactome. *Proteomics. Clinical Applications*, 117-133. doi:10.1002/prca.201100083

Schütz, P., Karlberg, T., van den Berg, S., Collins, R., Lehtiö, L., Högbom, M., Holmberg-Schiavone, L., et al. (2010). Comparative structural analysis of human DEAD-box RNA helicases. *PloS ONE*, *5*(9), 11. doi:10.1371/journal.pone.0012791

Selenko, P., Sprangers, R., Stier, G., Bühler, D., Fischer, U., & Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. *Nature Structural Biology*, *8*(1), 27-31. doi:10.1038/83014

Sessions, O. M., Barrows, N. J., Souza-Neto, J. A., Robinson, T. J., Hershey, C. L., Rodgers, M. A., Ramirez, J. L., et al. (2009). Discovery of insect and human dengue virus host factors. *Nature*, *458*(7241), 1047-50. doi:10.1038/nature07967

Severson, D. W., & Behura, S. K. (2012). Mosquito genomics: progress and challenges. *Annual Review of Entomology*, *57*, 143-66. doi:10.1146/annurev-ento-120710-100651

Shafee, N. (2003). Dengue virus type 2 NS3 protease and NS2B-NS3 protease precursor induce apoptosis. *Journal of General Virology*, *84*(8), 2191-2195. doi:10.1099/vir.0.19022-0

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498-504. doi:10.1101/gr.1239303

Shapira, S. D., Gat-Viks, I., Shum, B. O. V., Dricot, A., de Grace, M. M., Wu, L., Gupta, P. B., et al. (2009). A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell*, *139*(7), 1255-67. Elsevier Ltd. doi:10.1016/j.cell.2009.12.018

Smyth, G. (2005). *Limma: linear models for microarray data*. *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (pp. 397-420). Retrieved from http://www.springerlink.com/index/g26110k024423738.pdf

Smyth, G. K., & Speed, T. (2003). Normalization of cDNA microarray data. *Methods*, *31*(4), 265-73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14597310

Souza-Neto, J. A., Sim, S., & Dimopoulos, G. (2009). An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. *Proceedings of the National Academy of Sciences*, *106*(42), 17841-6. doi:10.1073/pnas.0905006106

Srichaikul, T., & Nimmannitya, S. (2000). Haematology in dengue and dengue haemorrhagic fever. *Baillière's Best Practice & Research in Clinical Haematology*, *13*(2), 261-76. doi:10.1053/beha.2000.0073

Ta, H. X., & Holm, L. (2009). Evaluation of different domain-based methods in protein interaction prediction. *Biochemical and Biophysical Research Communications*, *390*(3), 357-62. doi:10.1016/j.bbrc.2009.09.130

Tafforeau, L., Chantier, T., Pradezynski, F., Pellet, J., Mangeot, P. E., Vidalain, P.-O., Andre, P., et al. (2011). Generation and comprehensive analysis of an influenza

virus polymerase cellular interaction network. *Journal of Virology*, *85*(24), 13010-8. doi:10.1128/JVI.02651-10

Tagari, M., Tate, J., Swaminathan, G. J., Newman, R., Naim, A., Vranken, W., Kapopoulou, A., et al. (2006). E-MSD: improving data deposition and structure quality. *Nucleic Acids Research*, *34*(Database issue), D287-90. doi:10.1093/nar/gkj163

Tan, S.-L., Ganji, G., Paeper, B., Proll, S., & Katze, M. G. (2007). Systems biology and the host response to viral infection. *Nature Biotechnology*, *25*(12), 1383-9. doi:10.1038/nbt1207-1383

Tang, C., Ndassa, Y., & Summers, M. F. (2002). Structure of the N-terminal 283-residue fragment of the immature HIV-1 Gag polyprotein. *Nature Structural Biology*, *9*(7), 537-43. doi:10.1038/nsb806

Tastan, O., Qi, Y., Carbonell, J. G., & Klein-Seetharaman, J. (2009). Prediction of interactions between HIV-1 and human proteins by information integration. *Pacific Symposium on Biocomputing*, 516-27.

Taylor, D. R. (1999). Inhibition of the Interferon- Inducible Protein Kinase PKR by HCV E2 Protein. *Science*, *285*(5424), 107-110. doi:10.1126/science.285.5424.107

Thepparit, C., & Smith, D. R. (2004). Serotype-specific entry of dengue virus into liver cells: identification of the 37-kilodalton/67-kilodalton high-affinity laminin receptor as a dengue virus serotype 1 receptor. *Journal of Virology*, *78*(22), 12647-56. doi:10.1128/JVI.78.22.12647-12656.2004

Uetz, P., Dong, Y.-A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S. V., et al. (2006). Herpesviral protein networks and their interaction with the human proteome. *Science*, *311*(5758), 239-42. doi:10.1126/science.1116804

Umareddy, I., Pluquet, O., Wang, Q. Y., Vasudevan, S. G., Chevet, E., & Gu, F. (2007). Dengue virus serotype infection specifies the activation of the unfolded protein response. *Virology Journal*, *4*, 91.

Umareddy, I., Tang, K. F., Vasudevan, S. G., Devi, S., Hibberd, M. L., & Gu, F. (2008). Dengue virus regulates type I interferon signalling in a strain-dependent manner in human cell lines. *The Journal of General Virology*, *89*(Pt 12), 3052-62. doi:10.1099/vir.0.2008/001594-0

Van Gorp, E. C. M., Minnema, M. C., Suharti, C., Mairuhu, A. T. A., Brandjes, D. P. M., ten Cate, H., Hack, C. E., et al. (2001). Activation of coagulation factor XI, without detectable contact activation in dengue haemorrhagic fever. *British Journal of Haematology*, *113*(1), 94-99. doi:10.1046/j.1365-2141.2001.02710.x

Velankar, S., McNeil, P., Mittard-Runte, V., Suarez, A., Barrell, D., Apweiler, R., & Henrick, K. (2005). E-MSD: an integrated data resource for bioinformatics. *Nucleic Acids Research*, *33*(Database issue), D262-5. doi:10.1093/nar/gki058

Wang, J., Li, C., Wang, E., & Wang, X. (2009). Uncovering the rules for protein-protein interactions from yeast genomic data. *Proceedings of the National Academy of Sciences*, *106*(10), 3752-7. doi:10.1073/pnas.0806427106

Wang, L., Zhang, H., Solski, P. A., Hart, M. J., Der, C. J., & Su, L. (2000). Modulation of HIV-1 Replication by a Novel RhoA Effector Activity. *Journal of Immunology*, *164*(10), 5369-5374.

Wilk, T., Gowen, B., & Fuller, S. D. (1999). Actin Associates with the Nucleocapsid Domain of the Human Immunodeficiency Virus Gag Polyprotein. *Journal of Virology*, *73*(3), 1931-1940.

Worthylake, D. K., Wang, H., Yoo, S., Sundquist, W. I., & Hill, C. P. (1999). Structures of the HIV-1 capsid protein dimerization domain at 2.6 Å resolution. *Acta Crystallographica Section D: Biological Crystallography*, *55*(1), 85-92. doi:10.1107/S0907444998007689

Xi, Z., Ramirez, J. L., & Dimopoulos, G. (2008). The Aedes aegypti toll pathway controls dengue virus infection. *PLoS Pathogens*, *4*(7), e1000098. doi:10.1371/journal.ppat.1000098

Xu, T., Sampath, A., Chao, A., Wen, D., Nanao, M., Chene, P., Vasudevan, S. G., et al. (2005). Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 A. *Journal of Virology*, *79*(16), 10278-88. doi:10.1128/JVI.79.16.10278-10288.2005

Yamanishi, Y., Vert, J.-P., & Kanehisa, M. (2004). Protein network inference from multiple genomic data: a supervised approach. *Bioinformatics*, *20 Suppl 1*, i363-70. doi:10.1093/bioinformatics/bth910

Yeung, M. L., Houzet, L., Yedavalli, V. S. R. K., & Jeang, K.-T. (2009). A genome-wide short hairpin RNA screening of jurkat T-cells for human proteins contributing to productive HIV-1 replication. *The Journal of Biological Chemistry*, *284*(29), 19463-73. doi:10.1074/jbc.M109.010033

Yocupicio-Monroy, M., Padmanabhan, R., Medina, F., & del Angel, R. M. (2007). Mosquito La protein binds to the 3' untranslated region of the positive and negative polarity dengue virus RNAs and relocates to the cytoplasm of infected cells. *Virology*, *357*(1), 29-40. doi:10.1016/j.virol.2006.07.042

Yoshida, H., Okada, T., Haze, K., Yanagi, H., Yura, T., Negishi, M., & Mori, K. (2001). Endoplasmic reticulum stress-induced formation of transcription factor complex

ERSF including NF-Y (CBF) and activating transcription factors 6α and 6β that activates the mammalian unfolded protein response. *Molecular and Cellular Biology*, *21*(4), 1239-48. doi:10.1128/MCB.21.4.1239-1248.2001

Yu, C.-Y., Hsu, Y.-W., Liao, C.-L., & Lin, Y.-L. (2006). Flavivirus infection activates the XBP1 pathway of the unfolded protein response to cope with endoplasmic reticulum stress. *Journal of Virology*, *80*(23), 11868-80. doi:10.1128/JVI.00879-06

Yu, J., Pacifico, S., Liu, G., & Finley, R. L. (2008). DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics*, *9*(1), 461. doi:10.1186/1471-2164-9-461

Yu, Y., Black, J. B., Goldsmith, C. S., Browning, P. J., Bhalla, K., & Offermann, M. K. (1999). Induction of human herpesvirus-8 DNA replication and transcription by butyrate and TPA in BCBL-1 cells. *The Journal of General Virology*, *80 (Pt 1)*, 83-90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9934688

Zhang, L. V., Wong, S. L., King, O. D., & Roth, F. P. (2004). Predicting co-complexed protein pairs using genomic and proteomic data integration. *BMC Bioinformatics*, *5*(38). doi:10.1186/1471-2105-5-38

Zhang, L., Villa, N., Rahman, M., Smallwood, S., Shattuck, D., Neff, C., Dufford, M., et al. (2009). Analysis of Vaccinia Virus−Host Protein−Protein Interactions: Validations of Yeast Two-Hybrid Screenings. *Journal of Proteome Research*, *8*, 4311-4318.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*(1), 40. doi:10.1186/1471-2105-9-40

Zhou, H., Xu, M., Huang, Q., Gates, A. T., Zhang, X. D., Castle, J. C., Stec, E., et al. (2008). Genome-scale RNAi screen for host factors required for HIV replication. *Cell Host & Microbe*, *4*(5), 495-504. doi:10.1016/j.chom.2008.10.004