

INVESTIGATION OF INFECTION MECHANISMS THROUGH PATHOGEN-HUMAN
PROTEIN-PROTEIN INTERACTIONS BY BIOINFORMATICS APPROACHES

by

Saliha Durmuş Tekir

B.S., Chemical Engineering, Boğaziçi University, 2004

M.S., Chemical Engineering, Boğaziçi University, 2007

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

Graduate Program in Chemical Engineering
Boğaziçi University
2013

To My Mother, Tülay Durmuş

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor, Prof. Kutlu Ülgen, for her everlasting support, encouragement and kind attitude throughout my PhD studies. Additionally, I sincerely thank Assist. Prof. Tunahan Çakır for his support and his contribution to all chapters of the dissertation. I am thankful to Assoc. Prof. Erdoğan Sevilgen and Assist. Prof. Arzucan Özgür for their considerable contributions regarding the novel software PHISTO presented in Chapter 3. I would like to acknowledge the other members of my thesis committee, Prof. Zeynep Petek Çakar, Prof. Türkan Haliloğlu and Prof. Özlem Keskin for their valuable comments.

This research was supported by Boğaziçi University Research Fund through project 5554D, TÜBİTAK through project 110M428, Ekin Kimya Tic. Ltd. Şti. and Elginkan Foundation. I gratefully acknowledge doctoral scholarship provided by TÜBİTAK-BİDEB and financial contributions regarding conferences provided by BÜVAK.

I would like to deeply thank to all contributors for PHISTO, especially Ali Semih Sayılırbaş, Emre Ardıç and İlknur Karadeniz. I also gratefully acknowledge Nezih Hekim, Gamze Artun, Bülent Üner and Murat Akman for their supports for PHISTO.

I am deeply indebted for never-ending support of my family. I gratefully thank my mother, Tülay for her unbelievable support throughout my education and her taking care of my daughter during my PhD studies is invaluable for me. In fact, this thesis would not be possible without her and therefore I dedicate it to her. My very special thanks are for my five-year-old princess Zuhale for enabling me to taste the feeling of being mother and to develop ability to perform many tasks at the same time. I also would like to express my very special thanks to men in my life; my father Osman, my brother Salih and my husband Ünsal for their invaluable support.

My final thanks are for all current and former members of our Biosystems Engineering Research Group, I spent time together in KB 440 during all my graduate studies, the last nine years, for providing me a nice study environment.

ABSTRACT

INVESTIGATION OF INFECTION MECHANISMS THROUGH PATHOGEN-HUMAN PROTEIN-PROTEIN INTERACTIONS BY BIOINFORMATICS APPROACHES

Since ancient times, even in today's modern world, infectious diseases cause lots of people to die. The interactions between the proteins of infectious microorganisms, pathogens and their human hosts allow the microorganisms to manipulate human cellular mechanisms to their own advantage, resulting in infection in the host organism. Improved understanding of pathogen-host molecular interactions will increase our knowledge of the mechanisms involved in infection, and allow novel therapeutic solutions to be devised. In the post-genomic era, following the advances in genomics, proteomics, and then interactomics, interspecies protein interaction data of pathogen-human systems can be produced in large-scale within very recent years. In this thesis, pathogen-host interaction search tool (PHISTO) has been developed as a Web-accessible platform providing relevant information about pathogen-host interactions (PHIs). It enables access to the most up-to-date PHI data for all pathogen types for which experimentally-verified protein interactions with human are available. The platform also offers integrated tools for visualization of PHI networks, graph theoretical analysis of human proteins targeted by pathogens and BLAST search. PHISTO aims to facilitate PHI studies that provide potential therapeutic targets for infectious diseases. Comparative interactomics analyses were performed to investigate infection mechanisms of bacteria versus viruses and then DNA viruses versus RNA viruses, using PHI data from PHISTO, constituting the first comprehensive attempts to get such comparisons. Crucial insights on common and special infection strategies used by these pathogen types were provided by bioinformatics analyses of PHI networks, focusing on the properties of the targeted human proteins.

ÖZET

PATOJEN-İNSAN PROTEİN-PROTEİN ETKİLEŞİMLERİNİN BİYOİNFORMATİK ANALİZİ İLE ENFEKSİYON MEKANİZMALARININ İNCELENMESİ

Enfeksiyöz hastalıklar eski çağlardan günümüz modern dünyasına değin insan ölümlerinin başlıca nedenleri arasında gelmektedir. Enfeksiyon yapıcı mikroorganizmalar, patojenler, insan vücuduna girdiklerinde insan proteinleri ile etkileşerek, insanın hücresel mekanizmalarını kendi yararlarına kullanarak enfeksiyona neden olurlar. Bu patojen-konak moleküler etkileşimlerinin anlaşılması, enfeksiyon mekanizmaları hakkındaki bilgimizi arttıracak ve yeni tedavi çözümlerinin bulunmasını sağlayacaktır. Genom, proteom ve sonrasında interaktom verilerindeki gelişmeleri takiben, büyük ölçekli patojen-insan arası protein etkileşimleri son yıllarda üretilebilmektedir. Bu tez kapsamında, patojen-konak etkileşimleri (PKE) hakkında detaylı bilgi sunan patojen-konak etkileşim arama aracı, PHISTO yeni bir Web-erişilebilir platform olarak geliştirilmiştir. PHISTO, insan ile arasında deneysel olarak bulunmuş protein etkileşim verisi olan bütün patojen tipleri için en güncel PKE verisine ulaşmayı sağlamaktadır. Bu platform kullanıcılara ayrıca PKE ağyapılarının görselleştirilmesi, patojenler tarafından hedef alınan insan proteinlerinin grafik teorisi ile analizi ve BLAST arama konularında hesaplamalı analiz araçları sunmaktadır. PHISTO, böylelikle enfeksiyöz hastalıklar için potansiyel ilaç hedefleri belirleyebilecek PKE çalışmalarını kolaylaştırmayı hedeflemektedir. PHISTO'daki PKE verileri kullanılarak, öncelikle bakteri ve virüs sonrasında da DNA virüs ve RNA virüs enfeksiyon mekanizmalarının aydınlatılması amacıyla, karşılaştırılmalı interaktom analizleri, bu çalışmayla literatürde ilk defa bu kadar kapsamlı olarak gerçekleştirilmiştir. PKE ağyapılarının biyoinformatik analizleri sonucu patojen tipleri tarafından ortak ve özel olarak kullanılan enfeksiyon stratejileri hakkında önemli bulgular elde edilmiştir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT	v
ÖZET	vi
LIST OF FIGURES	x
LIST OF TABLES	xiv
LIST OF ACRONYMS/ABBREVIATIONS	xvi
1. INTRODUCTION	1
2. LITERATURE SURVEY	4
2.1. PPI Networks Related to Human Pathogens	4
2.1.1. PPI Networks of Viral Pathogens	5
2.1.2. PPI Networks of Bacterial Pathogens	8
2.1.3. PPI Networks of Protozoan Pathogens	12
2.1.4. The Impacts of Pathogenic PPI Networks	12
2.2. Pathogen-Human Interaction Networks	13
2.2.1. Virus-Human Interaction Networks	14
2.2.2. Bacteria-Human Interaction Networks	17
2.2.3. Analysis of Comprehensive PHI Data	20
2.2.4. The Impacts of PHI Networks	21
3. PHISTO: PATHOGEN-HOST INTERACTION SEARCH TOOL	23
3.1. Data Extraction and Integration	24
3.1.1. Data Update System	26
3.1.2. Text Mining of Missing Experimental Methods in PHI Data	27
3.1.3. Data Content	28
3.2. User Interface	28
3.2.1. Data Query	28
3.2.2. Data Presentation	30
3.2.3. BLAST Search	35
3.2.4. Graph Theoretical Analysis of Targeted Human Proteins	36
3.2.5. Help Page	38

3.3. Discussion	38
4. INFECTION STRATEGIES OF BACTERIAL AND VIRAL PATHOGENS THROUGH PATHOGEN-HUMAN PROTEIN-PROTEIN INTERACTIONS ...	40
4.1. Protein Interaction Data	41
4.1.1. PHI Data	41
4.1.2. Human PPI Data	41
4.1.3. Human Protein Sets	44
4.2. Computational Methods	44
4.2.1. Degree and Betweenness Centrality Calculations	44
4.2.2. GO Enrichment Analysis	45
4.3. Results	45
4.3.1. Pathogen-Targeted Human Proteins	45
4.3.2. Degree and Betweenness Centrality Distributions	48
4.3.3. GO Enrichment Analysis	51
4.4. Discussion	54
4.4.1. Special Infection Strategies	54
4.4.2. Common Infection Strategies	58
5. COMPARATIVE INTERACTOMICS FOR VIRAL INFECTION STRATEGIES: DNA VIRUSES VERSUS RNA VIRUSES	60
5.1. Protein Interaction Data	62
5.1.1. Virus-Human PHI Data	62
5.1.2. Human Protein Sets	64
5.2. Computational Methods	64
5.2.1. All DNA Viruses versus All RNA Viruses	65
5.2.2. GO Enrichment Analysis	69
5.3. Virus Families with the Largest PHI Data	74
5.3.1. DNA Virus Families	74
5.3.2. RNA Virus Families	75
5.4. Virus Strains with the Largest PHI Data	77
5.4.1. DNA Virus Strains	77
5.4.1.1. Herpesviruses	78
5.4.1.2. Papillomaviruses	78
5.4.1.3. Poxviruses	79

5.4.2. RNA Virus Strains	80
5.4.2.1. Flaviviruses	80
5.4.2.2. Orthomyxoviruses	81
5.4.2.3. Paramyxoviruses	82
5.5. Graph Analysis of Viral PHI Networks	83
6. CONCLUSIONS AND RECOMMENDATIONS	85
6.1. Conclusions	85
6.2. Recommendations	86
6.2.1. Improvements on PHISTO	86
6.2.2. Future Works on PHI Analysis	87
REFERENCES	89

LIST OF FIGURES

Figure 2.1.	PPI map of KSHV obtained from IntAct. The figure was drawn in Cytoscape version 2.8.2.	7
Figure 2.2.	PHI map of EBV-human obtained from PHISTO. Dark grey nodes are EBV proteins and light grey nodes are human proteins. The figure was drawn in Cytoscape version 2.8.2.	15
Figure 2.3.	Number of publications including PHI-related terms in PubMed in the post genomic era.	21
Figure 3.1.	The architecture of PHISTO.	24
Figure 3.2.	User Interface of the off-line application for data update.	26
Figure 3.3.	Homepage of PHISTO user interface.	29
Figure 3.4.	Quick Search example.	30
Figure 3.5.	Advanced Search example.	31
Figure 3.6.	Browse example.	32
Figure 3.7.	Sample portion of a PHI data presentation table.	32
Figure 3.8.	Bipartite graph of human Papillomavirus 31-human PHI data.	33
Figure 3.9.	Results Statistics options.	34

Figure 3.10.	Distribution of the PHI results of Quick Search of "protease" via pie chart on pathogen types and bacterial families.	34
Figure 3.11.	Distribution of the PHI results of Quick Search of "protease" via bar chart on pathogen types and bacterial families.	35
Figure 3.12.	BLAST interface.	36
Figure 3.13.	Degree and Betweenness Centrality values tabulated for the human proteins targeted by bacterial family Chlamydiae.	37
Figure 3.14.	Degree distribution of the human proteins targeted by bacterial family Chlamydiae in comparison to all proteins in the human PPI network.	37
Figure 3.15.	Help page.	38
Figure 4.1.	The number of human proteins that are grouped based on their interactions with viruses, bacteria and fungi/protozoa (targeted by fungi and/or protozoa).	45
Figure 4.2.	The cumulative degree distributions of bacteria-targeted human proteins.	48
Figure 4.3.	The cumulative degree distributions of virus-targeted human proteins.	49
Figure 4.4.	The cumulative betweenness centrality distributions of bacteria-targeted human proteins.	49
Figure 4.5.	The cumulative betweenness centrality distributions of virus-targeted human proteins.	50

Figure 4.6.	The cumulative degree distributions of targeted human proteins excluding Yersinia, Bacillus and HIV data.	50
Figure 4.7.	The cumulative degree distributions of targeted human proteins excluding Yersinia, Bacillus and HIV data.	51
Figure 5.1.	The number of human proteins targeted by DNA viruses and/or RNA viruses.	65
Figure 5.2.	The number of DNA virus families-targeted human proteins that are grouped based on their interactions with families; Herpesviruses, Papillomaviruses and Poxviruses.	75
Figure 5.3.	The number of RNA virus families-targeted human proteins that are grouped based on their interactions with families; Flaviviruses, Orthomyxoviruses and Paramyxoviruses.	76
Figure 5.4.	The number of Herpesvirus strains-targeted human proteins that are grouped based on their interactions with strains; Human Herpesvirus 4 strain B95-8, Human Herpesvirus 1 strain 17 and Human Herpesvirus 4.	78
Figure 5.5.	The number of Papillomavirus strains-targeted human proteins that are grouped based on their interactions with strains; Human Papillomavirus 16, Human Papillomavirus 11 and Human Papillomavirus 18.	79
Figure 5.6.	The number of Poxvirus strains-targeted human proteins that are grouped based on their interactions with strains; Vaccinia virus strain WR, Orf virus strain SA00 and Vaccinia virus strain Copenhagen.	79

Figure 5.7.	The number of Flavivirus strains-targeted human proteins that are grouped based on their interactions with strains; Hepatitis C virus, Hepatitis C virus isolate Con1 and Hepatitis C virus isolate H.	81
Figure 5.8.	The number of Orthomyxovirus strains-targeted human proteins that are grouped based on their interactions with Influenza A virus strains; H1N1-1934, H5N1-1997 and H3N2-1972.	81
Figure 5.9.	The number of Paramyxovirus strains-targeted human proteins that are grouped based on their interactions with strains; Measles virus strain Schwarz, Human Respiratory Syncytial Virus strain A2 and Hendra virus isolate horse.	82

LIST OF TABLES

Table 2.1.	Examples of PPI networks of pathogens detailed in the literature.	11
Table 2.2.	Examples of PHI networks detailed in the literature.	19
Table 4.1.	Contents of bacteria-human PHI data.	42
Table 4.2.	Contents of fungi-human PHI data.	42
Table 4.3.	Contents of protozoa-human PHI data.	42
Table 4.4.	Contents of virus-human PHI data.	43
Table 4.5.	Highly targeted human proteins.	46
Table 4.6.	First 20 enriched GO Process terms in human proteins targeted by at least three bacterial groups (three bacteria-targeted set).	52
Table 4.7.	First 20 enriched GO Process terms in human proteins targeted by at least three viral groups (three viruses-targeted set).	52
Table 4.8.	First 20 enriched GO Process terms in human proteins targeted by both bacterial and viral groups (bacteria-virus-targeted set).	53
Table 5.1.	Contents of DNA viruses-human PHI data.	63
Table 5.2.	Contents of RNA viruses-human PHI data.	63
Table 5.3.	Comparison of viral PHI data: DNA viruses versus RNA viruses.	65

Table 5.4.	Human proteins targeted by at least three DNA virus families.	66
Table 5.5.	Human proteins targeted by at least five RNA virus families.	67
Table 5.6.	First 20 enriched GO Process terms in human proteins targeted by only DNA viruses.	70
Table 5.7.	First 20 enriched GO Process terms in human proteins targeted by only RNA viruses.	70
Table 5.8.	First 20 enriched GO Process terms in human proteins targeted by at least three DNA virus families.	71
Table 5.9.	First 20 enriched GO Process terms in human proteins targeted by at least five RNA virus families.	72
Table 5.10.	First 20 enriched GO Process terms in human proteins targeted by both DNA and RNA viruses.	73
Table 5.11.	Contents of PHI data belonging to DNA virus families with the largest data.	74
Table 5.12.	Contents of PHI data belonging to RNA virus families with the largest data.	75
Table 5.13.	Contents of PHI data belonging to DNA virus strains with the largest data.	77
Table 5.14.	Contents of PHI data belonging to RNA virus strains with the largest PHI data.	80
Table 5.15.	Proteins with the highest degree values for PHI data of DNA viruses.	84

LIST OF ACRONYMS/ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
BC	Betweenness centrality
BLAST	Basic local alignment search tool
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
GO	Gene ontology
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
KEGG	Kyoto encyclopedia of genes and genomes
KSHV	Kaposi sarcoma-associated herpesvirus
ORF	Open reading frame
PHI	Pathogen-host interaction
PHISTO	Pathogen-host interaction search tool
PPI	Protein-protein interaction
RNA	Ribonucleic acid
SARS	Severe acute respiratory syndrome
VZV	Varicella-zoster virus
WHO	World health organization

1. INTRODUCTION

Despite immense technological advances in medicine, pathogenic organisms remain to be the source of much human morbidity and mortality. World Health Organization (WHO) reported that more than 20% of total deaths in the world are due to infectious diseases. However that number, amount of annual infectious disease deaths, is projected to decline over the next 20 years, with new technologies that will be developed (WHO, 2012). HIV/AIDS, acute lower respiratory tract infections, hemorrhagic fever, diarrheal diseases, tuberculosis and malaria are particularly notorious for high mortality rates (Fauci, 2001; Forst, 2006; Taylor *et al.*, 2010). The continuous emergence of both new diseases and drug-resistant pathogens has heightened the global burden of infectious diseases in the 21st century (Fauci, 2001; Bünnige, 2011). To tackle with such biological threats, an improved understanding of pathogenic microorganisms and their interactions with host organisms is needed since pathogen-host molecular interactions have crucial roles in initiating, sustaining, or preventing infection. Pathogenic microorganisms communicate with human cells through interactions with human proteins both on the surface of the cell and within the interior of the cell. These interactions allow the microorganisms to enter the host cell and manipulate cellular mechanisms in order to use the host cell's capabilities to their own advantage, resulting in infection in the host organism. Detailed knowledge of pathogen-host protein interactions may enable us to comprehend the mechanisms of infection and to identify better strategies to prevent or cure infection (Münter *et al.*, 2006; Dyer *et al.*, 2008).

The very first genome sequencing was published in 1977 with the DNA sequence for the genome of a virus, bacteriophage phiX174 (Sanger *et al.*, 1977). Following the sequencing of the bacterial pathogen *Haemophilus influenzae* in 1995 (Fleischmann *et al.*, 1995) and the human genome in 2000 (Venter *et al.*, 2001), sequence data for prokaryotic and eukaryotic genomes have appeared at an accelerated rate. Today, genomic data for most of the pathogen and host organisms are available (Benson *et al.*, 2013). These data are used to study individual genes and corresponding proteins as well as to identify intra- and interspecies connections between proteins. In the light of these advances, the initial steps towards complete understanding of infection mechanisms through protein

interactions have been recently published, the review of which is given in the following chapter.

In the present study, it is aimed to provide a framework to investigate infection mechanisms through PHI networks. The second chapter, entitled as "Literature Survey", summarizes the efforts to systematic determination and analysis of protein interaction networks underlying infection pathogenesis and thus presents the current picture of the research on infectious diseases (Durmuş Tekir *et al.*, 2013a). The following three chapters, on the other hand, detail research performed in this thesis, that are "PHISTO: Pathogen-Host Interaction Search Tool", "Infection Strategies of Bacterial and Viral Pathogens through Pathogen-Human Protein-Protein Interactions" and "Comparative Interactomics for Viral Infection Strategies: DNA Viruses versus RNA viruses", sequentially.

The third chapter, introduces the novel software developed; PHISTO as a pathogen-host interaction (PHI) resource enriched with bioinformatics network analysis tools (Durmuş Tekir *et al.*, 2013b). It stores experimentally-verified protein-protein interactions (PPIs) between pathogenic microorganisms and their human hosts. The need in the related research field for such a tool is introduced at the beginning of the chapter. The architecture of the software are detailed in "Data Extraction and Integration" and "User Interface" sections. Mechanisms of data import from other PPI databases and data update system are given in the first section. The additional functionalities offered to users through user interface of PHISTO are presented in the second section.

The fourth chapter gives the details of bioinformatics analysis of the comprehensive PHI data with a special emphasis on the comparison of attacking mechanisms of bacterial and viral pathogens against human (Durmuş Tekir *et al.*, 2012). The analysis is based on the properties of human proteins that are targeted by these pathogens. The common infection mechanisms shared by both pathogen types and also the mechanisms that are specific to each pathogen type are described.

The fifth chapter focuses on the comparison of the infection strategies developed by different types of viral pathogens; DNA viruses and RNA viruses, employing a similar approach as in chapter 4. The effects of the differences in their genetic materials on the

mechanisms of the infections they cause, are discussed here. Additionally, the interaction characteristics of the PHI networks were investigated graph theoretically, and potential drug targets were identified.

The summary of the main results and main contributions to the research field as well as the recommendations for future work are given in the “Conclusions and Recommendations” chapter.

2. LITERATURE SURVEY

Complete genome sequences for a number of pathogenic microorganisms, as well as the human host, have led to the revelation of their PPI networks. In this post-genomic era, PHIs operating during infection can also be mapped. The identification of new drug and vaccine targets for infectious diseases is only possible when the molecular machinery within individual pathogenic and host organisms is understood, in addition to pathogen-host interaction systems. For instance, anti-infection therapeutics should target essential genes in the pathogens which have no homology with human genes (Barker, 2006). Therefore, detailed systematic analyses of PPI and PHI data together are required for a complete understanding of pathogenesis of infections. In this chapter, the striking results recently obtained during the construction and investigation of these networks are presented, in a chronological order. Emphasis is placed on studies producing large-scale interaction data by high-throughput experimental techniques (Durmuş Tekir *et al.*, 2013a).

2.1. PPI Networks Related to Human Pathogens

From a classical perspective, a protein is a functional unit that specifies a small, but discrete, part of the cellular physiology of an organism. In the post-genomic era, a protein is seen to function as an element within network of its interaction, and its role should be evaluated within this network together with its interacting partners (Eisenberg *et al.*, 2000). Advances in genomics and proteomics have been followed by the first large-scale efforts to identify functional networks of interacting proteins using the two-hybrid method (Fields and Song, 1989; Finley and Brent, 1994; Fromont-Racine *et al.*, 1997), pull-down assays (Gavin *et al.*, 2002; Ho *et al.*, 2002), and protein chips (Zhu *et al.*, 2001). To increase our understanding of the mechanisms of infection, PPI networks of pathogenic organisms should be determined in order to capture their functional and structural organizations. Pathogenic PPI maps reveal biological pathways and processes, allowing prediction of protein functions and discovery of new drug and vaccine targets.

The first genome-wide protein interaction networks were determined for viruses (Bartel *et al.*, 1996; Flajolet *et al.*, 2000; McCraith *et al.*, 2000). The first large-scale

bacterial networks (Rain *et al.*, 2001; Butland *et al.*, 2005; Arifuzzaman *et al.*, 2006) followed successes in eukaryote mapping (Fromont-Racine *et al.*, 1997; Ito *et al.*, 2000; Uetz *et al.*, 2000; Walhout *et al.*, 2000). Today, the genome-wide PPI maps for a large number of pathogens and hosts are available in public databases: BIND (Alfarano *et al.*, 2005), BioGrid (Stark *et al.*, 2011), DIP (Salwinski *et al.*, 2004), HPRD (Prasad *et al.*, 2009), IntAct (Kerrien *et al.*, 2012), MINT (Licata *et al.*, 2012), MIPS (Guldener *et al.*, 2006), Reactome (Croft *et al.*, 2011) and STRING (Szklarczyk *et al.*, 2011).

2.1.1. PPI Networks of Viral Pathogens

Primarily due to their small genome size, whole genome PPI maps were first constructed for viruses. The first interaction map of whole proteome was determined for *Escherichia coli* bacteriophage T7, mapping 25 interactions among viral proteins (Bartel *et al.*, 1996). Subsequently, genome-wide analyses of important human pathogens, Hepatitis C virus (Flajolet *et al.*, 2000; Dimitrova *et al.*, 2003), Vaccinia virus (McCraith *et al.*, 2000), Herpesviruses (Uetz *et al.*, 2006; Calderwood *et al.*, 2007), SARS coronavirus (von Brunn *et al.*, 2007; Pan *et al.*, 2008), and Influenza A virus (Shapira *et al.*, 2009) were performed through intraviral PPI maps.

Hepatitis C virus (HCV), a Flaviviridae family member causing severe liver disease, is a positive-sense single-stranded RNA virus. It encodes only a single polyprotein which is co- or post-translationally processed into at least 10 viral proteins (Houghton *et al.*, 1994). A controlled two-hybrid strategy based on a random genomic HCV library screen was used by Flajolet *et al.* (2000), resulting in the identification of known and novel PPIs. Interactions among structural and non-structural proteins were revealed in the study, leading to the conclusion that almost all of the viral proteins encoded by the genome function in the HCV life-cycle, as in the cases of other members of the Flaviviridae (Knipe and Howley, 2007). The roles of these functional interactions were partially enlightened within the framework of the constructed genome-wide interaction map. Interacting domains of the viral polyprotein were also identified to shed light on the development of anti-viral agents (Flajolet *et al.*, 2000). Another genome-wide PPI map of HCV was then generated for the viral non-structural proteins (Dimitrova *et al.*, 2003).

Vaccinia virus, well-known as a smallpox vaccine and also the source of potential recombinant vaccines against cancer and infectious diseases, is a member of Poxviridae family. It is a large, double-stranded DNA virus. Poxviruses replicate themselves in the cytoplasm of the host cells without depending on the host's transcriptional machinery. The large genome of Vaccinia virus can potentially express more than 200 proteins (Goebel *et al.*, 1990; Moss, 1996). McCraith *et al.* (2000) performed a comprehensive two-hybrid analysis of full-length Vaccinia virus proteins and detected 37 PPIs (including 28 novel interactions) among both characterized and uncharacterized proteins. Many of the PPIs mapped involved one partner which was known to function in a specific process, coupled with another of unknown function, allowing functions to be assigned to previously unannotated proteins within DNA replication, transcription, virion structure, or host evasion.

Another double-stranded DNA virus family is Herpesviridae whose members encode 70-170 proteins. Herpesviruses cause human diseases such as Kaposi sarcoma, B-cell lymphomas, chickenpox, shingles, and nasopharyngeal carcinoma (Davison and Scott, 1986; Chang *et al.*, 1994; Dolan *et al.*, 2006). The genome-wide intraviral protein interaction maps for three members of this family, Kaposi sarcoma-associated Herpesvirus (KSHV), Varicella-zoster virus (VZV), and Epstein-Barr virus (EBV) were generated by two-hybrid and analyzed comprehensively to reveal viral network properties (Uetz *et al.*, 2006; Calderwood *et al.*, 2007). In the work of Uetz *et al.* (2006), 123 PPIs for KSHV and 173 PPIs for VZV were identified, the largest dataset published to date, allowing the construction of the first viral networks. Topological network analyses of these interactome maps indicated that the viral networks appear as a single, highly coupled module (Figure 2.1) with relatively many hubs and few peripheral nodes (Uetz *et al.*, 2006) in contrast to scale-free cellular networks with well-separated functional modules (Barabasi and Albert, 1999; Maslov and Sneppen, 2002). Just after this study was published, Calderwood *et al.* (2007) reported the detection of 43 PPIs among EBV proteins. The construction of a PPI map for EBV by merging these interactions with already published ones resulted in a network of 52 proteins with 60 interactions. This large-scale network allowed the prediction of functions of uncharacterized proteins, further defining viral mechanisms. In these consecutive studies (Uetz *et al.*, 2006; Calderwood *et al.*, 2007), core proteins common to all Herpesviruses and noncore ones specific to each strain were revealed.

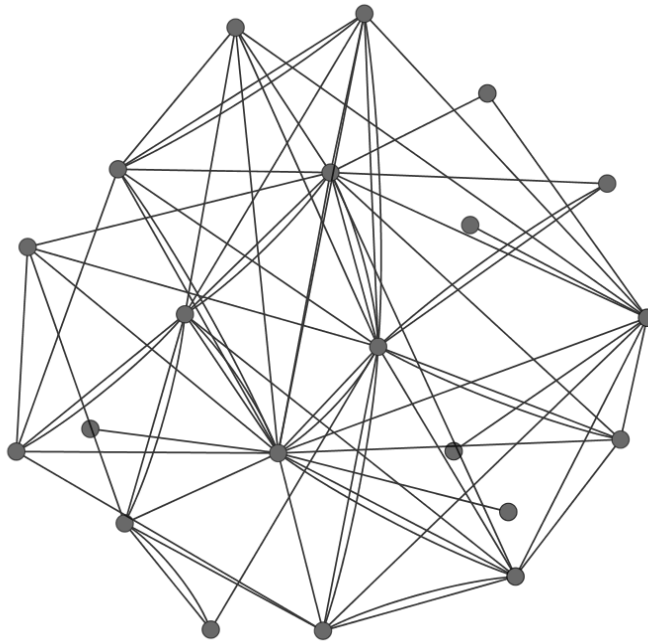


Figure 2.1. PPI map of KSHV obtained from IntAct. The figure was drawn in Cytoscape version 2.8.2.

The severe acute respiratory syndrome coronavirus (SARS-CoV) is a positive-sense single-stranded RNA virus belonging to the family of the largest RNA viruses known, Coronaviridae. Its genome encodes 14 open reading frames expressing up to 30 structural and non-structural proteins that have roles in viral replication, assembly, and other functions for viral amplification in host cells (Snijder *et al.*, 2003; von Brunn *et al.*, 2007; Pan *et al.*, 2008). For a genome-wide analysis of PPIs of SARS-CoV, interactions between all SARS-CoV proteins were determined (von Brunn *et al.*, 2007; Pan *et al.*, 2008) by two-hybrid producing 65 and 40 interactions, respectively. Intraviral PPIs were analyzed to elucidate the functions of the proteins as well as to identify the essential proteins in viral replication. von Brunn *et al.* (2007) compared the intraviral network topology of SARS-CoV with a previously defined viral network (Uetz *et al.*, 2006) and cellular networks (Rual *et al.*, 2005; Stelzl *et al.*, 2005; Lehner and Fraser, 2004), concluding SARS-CoV network contained similarities to the KSHV network (Uetz *et al.*, 2006).

Influenza A is a member of negative-sense single-stranded RNA viruses of Orthomyxoviridae family. It is the sources of all flu pandemics infecting multiple species. For H1N1 A/PR/8/34 strain of influenza virus, 31 intraviral PPIs among 10 viral proteins

were detected by two-hybrid (Shapira *et al.*, 2009). The topology of the constructed intraviral network revealed a highly interconnected nature, as observed previously for other viral networks (Uetz *et al.*, 2006; Bailer and Haas, 2009).

Insights gained into molecular mechanisms and topological network properties provided by the genome-wide analyses of intraviral PPI maps (Table 2.1) may be used as a basis for further characterization of the functions and mechanisms of viral proteins, especially for other members of the same virus families.

2.1.2. PPI Networks of Bacterial Pathogens

Having successfully built genome-wide PPI maps for viruses, similar two-hybrid methodology was applied to construct PPI networks for the larger, more complex genomes of pathogenic bacteria. The first prokaryotic PPI map was built for *Helicobacter pylori* (Rain *et al.*, 2001). Other large-scale prokaryotic networks eventually emerged for *Campylobacter jejuni* (Parrish *et al.*, 2007), *Treponema pallidum* (Titz *et al.*, 2008), *Mycobacterium tuberculosis* (Wang *et al.*, 2010), and *Bacillus subtilis* (Marchadier *et al.*, 2011). Genome-scale analysis of interacting proteins that assemble into protein complexes were performed for *Escherichia coli* (Butland *et al.*, 2005; Arifuzzaman *et al.*, 2006) and *Mycoplasma pneumoniae* (Kühner *et al.*, 2009).

The first large-scale intrabacterial PPI map was constructed for the human gastric pathogen, and gram-negative bacterium *H. pylori*, identifying 1280 interactions between 46.6% of all 261 bacterial proteins using the two-hybrid method (Rain *et al.*, 2001). The comparison of these *H. pylori* PPIs with previously described interactions between orthologous *E. coli* proteins resulted in prediction of protein functions within biological pathways such as chemotaxis and urease activity, essential for *H. pylori* pathogenicity. In that study, the interacting domains of *H. pylori* proteins were also identified and used in protein function predictions. Interacting domains may serve in mapping new functional domains, providing crucial information for antibacterial drug design studies.

Gram-negative bacterium *E. coli*, the main cause of urinary tract infections and a model bacterial system, is one of the best characterized and early studied organisms (Buck

et al., 1985; Datsenko and Wanner, 2000; Gully, 2003). However, any large-scale analysis of protein complexes in *E. coli* was not performed until the studies of Butland *et al.* (2005) and Arifuzzaman *et al.* (2006). First, 716 binary interactions involving 83 essential and 152 nonessential proteins, were identified by pull-down assay using tandem affinity purification-mass spectrometry, targeting 1000 ORFs (about one-quarter of the *E. coli* genome) (Butland *et al.*, 2005). A small number (15 %) of these interactions were already available in DIP, BIND, and STRING. Ten newly described *E. coli* PPIs were found as orthologous to the interactions reported for *H. pylori* (Rain *et al.*, 2001). The novel interactions were analyzed for functional annotations of uncharacterized proteins, allocating them within ribosome function, RNA processing, RNA binding, and so on. The graph theoretical analysis of the PPI map of *E. coli* revealed scale-free behavior and a high correlation between connectivity and the degree of conservation. The genome-wide PPI map of *E. coli* K-12 strain with 11,511 interactions among 2667 proteins was then constructed by a similar method (Arifuzzaman *et al.*, 2006). The comprehensive analysis of this large-scale network also validated the scale-free nature and the connectivity-conservation correlation found previously (Butland *et al.*, 2005). Arifuzzaman *et al.* (2006) identified 107 functional units which have roles in metabolic pathways, transcriptional and translational machinery, recombination and flagella assembly.

The gram-negative, food-borne pathogen *C. jejuni* is the major cause of gastroenteritis. The proteome-level analysis revealed 11,687 interactions involving 80% of 1654 *C. jejuni* proteins (Parrish *et al.*, 2007), the most comprehensive bacterial PPI map determined by two-hybrid method. A scale-free network was obtained, removing low confidence-scored interactions. This PPI map of *C. jejuni* was used to identify evolutionarily conserved subnetworks through comparison with protein networks of *H. pylori* (Rain *et al.*, 2001), *E. coli* (Butland *et al.*, 2005) and *Saccharomyces cerevisiae* in DIP. Further analyses of the identified conserved sub-networks allowed the prediction of new *C. jejuni* interactions using orthologous interactions. This comparative analysis also enabled the identification of essential *C. jejuni* genes based on their orthology to essential genes in other organisms. This comprehensive interactome data were next used to predict protein roles and to map functional pathways such as chemotaxis.

The causative agent of syphilis, *T. pallidum*, has one of the smallest genomes known in extracellular bacteria, encoding 1039 proteins (Fraser, 1998). The global PPI network of *T. pallidum*, involving 3649 interactions connecting 726 bacterial proteins, was identified by two-hybrid method (Titz *et al.*, 2008). The high-confidence subset connects 576 proteins by 991 interactions. In that study, an integrated network of DNA-metabolism related processes was constructed and 18 proteins were functionally annotated within this network. Additionally, various orthologous interactions were predicted for completely sequenced genomes, allowing the description of phylogenetically conserved interaction patterns.

Atypical pneumonia causing human pathogen, *M. pneumoniae* also has one of the smallest genomes in self-replicating organisms with 689 protein-encoding genes, making it a good model organism to study proteome organization in prokaryotes (Himmelreich *et al.*, 1996). A proteome-wide analysis was performed by tandem affinity purification mass spectrometry, identifying 62 homo-multimeric and 116 hetero-multimeric protein complexes (Kühner *et al.*, 2009). About a third of the found hetero-multimeric complexes were observed to interact with proteins forming 35 larger, multiprotein complexes implying higher level of proteome organization and protein multifunctionality, allowing functional annotations of assemblies as well as prediction of biological roles of individual proteins within the complexes.

M. tuberculosis causes millions of deaths each year with tuberculosis infection (WHO, 2009). After computational efforts to construct large-scale PPI maps of *M. tuberculosis* (Raman and Chandra, 2008; Cui *et al.*, 2009), its genome-wide network was identified experimentally by two-hybrid (Wang *et al.*, 2010). This global network is composed of 8042 interactions among 2907 proteins which represent 74.1% of the whole proteome. The topological properties of the undirected network of these interactions were calculated and compared with those of the previously defined prokaryotic PPI networks (Rain *et al.*, 2001; Butland *et al.*, 2005; Arifuzzaman *et al.*, 2006; Parrish *et al.*, 2007; Titz *et al.*, 2008). Similar scale-free behavior following a power-law distribution was observed. In fact, the networks obtained by pull-down assay (Butland *et al.*, 2005; Arifuzzaman *et al.*, 2006) differ in values of clustering coefficient from the networks obtained by two-hybrid analysis (Rain *et al.*, 2001; Parrish *et al.*, 2007; Titz *et al.*, 2008). Moreover, Wang

et al. (2010) performed a cross-species network comparison analysis of *M. tuberculosis* interactions with the available large-scale PPI data (Rain *et al.*, 2001; Butland *et al.*, 2005; Arifuzzaman *et al.*, 2006; Parrish *et al.*, 2007; Titz *et al.*, 2008) and identified conserved sub-networks. Additionally, the highly connected critical proteins and mechanisms of the protein secretion pathways which have roles in *M. tuberculosis* pathogenesis were revealed.

Table 2.1. Examples of PPI networks of pathogens detailed in the literature.

Pathogen name	Pathogen type	# of PPIs	References
<i>E. coli</i> bacteriophage T7	DNA virus	25	Bartel <i>et al.</i> , 1996
HCV	RNA virus	NA	Flajolet <i>et al.</i> , 2000
Vaccinia virus	DNA virus	37	McCraith <i>et al.</i> , 2000
<i>H. Pylori</i>	Gram- bacteria	1280	Rain <i>et al.</i> , 2001
<i>E. coli</i>	Gram- bacteria	716	Butland <i>et al.</i> , 2005
<i>P. falciparum</i>	Protozoa	2823	LaCount <i>et al.</i> , 2005
KSHV	DNA virus	123	Uetz <i>et al.</i> , 2006
VZV	DNA virus	173	Uetz <i>et al.</i> , 2006
<i>E. coli</i>	Gram- bacteria	11511	Arifuzzaman <i>et al.</i> , 2006
EBV	DNA virus	60	Calderwood <i>et al.</i> , 2007
SARS-CoV	RNA virus	65	von Brunn <i>et al.</i> , 2007
<i>C. jejuni</i>	Gram- bacteria	11687	Parrish <i>et al.</i> , 2007
<i>T. pallidum</i>	Gram- bacteria	3649	Titz <i>et al.</i> , 2008
SARS-CoV	RNA virus	40	Pan <i>et al.</i> , 2008
Influenza A virus (H1N1 A/PR/8/34)	RNA virus	31	Shapira <i>et al.</i> , 2009
<i>M. pneumoniae</i>	Bacteria without cell wall	178	Kühner <i>et al.</i> , 2009
<i>M. tuberculosis</i>	Bacteria without cell wall	8042	Wang <i>et al.</i> , 2010
<i>B. subtilis</i>	Gram+ bacteria	793	Marchadier <i>et al.</i> , 2011

A large-scale PPI network was recently constructed for the gram-positive bacterium *B. subtilis* (which is rarely pathogenic) by two-hybrid (Marchadier *et al.*, 2011). This network of 793 interactions involves 287 bacterial proteins. Due to its role as a model organism, many studies were performed to characterize the biological functions of its PPIs in cellular processes (Noirot-Gros *et al.*, 2002; Duigou *et al.*, 2005; Soufo *et al.*, 2008). However, many processes remained uncharacterized. Hence Marchadier *et al.* (2011) performed a comprehensive analysis with the integration of transcriptomic data focusing on cell division, cell responses to stresses, the bacterial cytoskeleton, DNA replication and chromosome maintenance.

These sequential efforts on construction of large-scale PPI networks for prokaryotes (Table 2.1) constitute the first comprehensive description of the intraspecies mechanisms of the bacterial pathogens.

2.1.3. PPI Networks of Protozoan Pathogens

The protozoan pathogen *Plasmodium falciparum* causes malaria which results in deaths of nearly a million of people each year (Guiguemde *et al.*, 2010). A comprehensive protein interaction map of this pathogen was generated by two-hybrid method, identifying a highly interconnected, scale-free network of 2823 interactions within 1267 proteins (~25% of the predicted *P. falciparum* proteins) (LaCount *et al.*, 2005). In this network, 33% of the interactions are between two uncharacterized proteins whereas 49% of the interactions include one such protein. Bioinformatics analysis of this network yielded functional annotations of the proteins within the processes; chromatin modification, transcription, messenger RNA stability, ubiquitination, and invasion of host cells. More detailed studies of PPIs within *P. falciparum* are required in order to unravel its pathogenesis mechanisms thoroughly.

2.1.4. The Impacts of Pathogenic PPI Maps

Despite the increasing rate in the identification of genome-wide PPI networks, they remain unconstructed for most pathogens. In the light of accelerating advances in genomics, proteomics, and interactomics, large-scale maps for many more organisms are

expected to be built in the near future. Increasing numbers of PPI networks will allow the comparison of networks across diverse organisms, resulting in generalized conclusions about pathogenic molecular mechanisms. The first examples of such comparative studies have been highlighted in Sections 2.1.1 and 2.1.2 above. Integration of several high-throughput interaction datasets to generate more detailed networks is also possible, as indicated by examples for the *E. coli* system (Yellabonia *et al.*, 2007; Hu *et al.*, 2009). The frequency of such integrated networks is expected to increase, owing to the large number of diverse data sets. These will be invaluable in defining whole proteomic maps of the pathogens.

One of the most striking results of bioinformatics analyses on the constructed PPI maps is the identification of essential proteins functioning within pathogens. These proteins should be examined thoroughly to test their potential as novel therapeutic targets. The exploration of genome-wide PPI maps of the pathogens permits the assignment of unannotated proteins to biological pathways with function prediction. The proteins annotated to the host invasion processes may provide a launching point for pathogen-host interaction studies.

2.2. Pathogen-Human Interaction Networks

Biochemical interactions of pathogens with their hosts are necessary to invade the host organism. These connections between pathogens and hosts include interactions between proteins, nucleotide sequences, and small ligands (Lengeling *et al.*, 2001; Stebbins, 2005). However, the protein interactions of pathogen-host systems have been identified as the most important and therefore the most studied type of pathogen-host interactions (PHIs) (Stebbins, 2005; Korkin *et al.*, 2011). Since these interspecies crosstalks determine the pathogenesis, focusing on the whole PHI system, instead of investigating a pathogen or host individually, may allow us to capture critical mechanisms (i.e. strategies used by pathogens and host immune responses) during infection that cannot be provided by traditional methods.

Due to a lack of sufficient experimental PHI data until recent years, many computational PHI prediction methods have been developed (Davis *et al.*, 2007; Dyer *et*

et al., 2007; Krishnadev and Srinivasan, 2008; Lee *et al.*, 2008; Tastan *et al.*, 2009; Evans *et al.*, 2009; Doolittle and Gomez, 2010). These studies focused mainly on interactions of *P. falciparum* and human immunodeficiency virus (HIV), as these are some of the most threatening pathogens to humans. Very recently experiments have been carried out to determine the first large-scale molecular interactions between human and viruses (Calderwood *et al.*, 2007; de Chassey *et al.*, 2008; Shapira *et al.*, 2009) and bacteria (Dyer *et al.*, 2010; Yang *et al.*, 2011). As a result of an increase in data available for pathogen-host systems, PHI-specific databases have been introduced such as PHI-base (Winnenburg *et al.*, 2008), PHIDIAS (Xiang *et al.*, 2007), VirusMINT (Chatr-aryamontri *et al.*, 2009), VirHostNet (Navratil *et al.*, 2009), HPIDB (Kumar and Nanduri, 2010), and PATRIC (Gillespie *et al.*, 2011). Although these advances in data archiving are promising, most data relevant to PHI are still buried in the biomedical literature. Some rare efforts have been performed to obtain hidden PHIs from the literature by text mining (Yin *et al.*, 2010; Anthony *et al.*, 2010; Thieu *et al.*, 2012).

2.2.1. Virus-Human Interaction Networks

As in the case of intraspecies pathogen PPIs, large-scale PHI data were generated for viral systems before bacterial systems (Table 2.2). The first examples are for commonly observed human pathogens, EBV (Calderwood *et al.*, 2007), HCV (de Chassey *et al.*, 2008) and Influenza A virus (H1N1 and H3N2) (Shapira *et al.*, 2009) and then recently for HIV (Jager *et al.*, 2012).

In Calderwood *et al.* (2007), protein interactions between Herpesvirus, EBV and human were mapped by two-hybrid in conjunction with EBV intraviral PPI mapping, providing 173 PHIs between 40 EBV proteins and 112 human proteins. A systematic analysis of these interactome maps of PPIs and PHIs enabled hypotheses of the roles of EBV proteins in pathogenesis to be generated. Furthermore, intraspecies protein interaction data for human were integrated from databases (BIND, DIP, HPRD, MIPS) and from the literature (Rual *et al.*, 2005; Stelzl *et al.*, 2005) to analyze the organization of the human proteins targeted by EBV within human molecular machinery. It was found that EBV proteins tend to target human proteins which are highly connected (hubs) and central to many paths (bottlenecks) in the human PPI network. On the other hand, the degree

distribution of the EBV-human protein interaction network could not be fitted to any model because of its incompleteness (Figure 2.2). Attempts to analyze incomplete maps of PPIs and PHIs are still able to supply a partial understanding of mechanisms underlying infection. A similar thorough analysis was earlier performed with Herpesviral protein networks of KSHV and VZV and their interaction with the human proteome (Uetz *et al.*, 2006). In that study, protein interactions between Herpesviruses and human were predicted using the interacting orthologs of both proteins in other organisms (Lehner and Fraser, 2004). Combined virus-human networks were constructed by starting with the viral networks, adding their human protein targets, and then adding the cellular interactions among the targeted human proteins. The topological analyses of the combined Herpesviruses-human networks revealed distinct properties from both viral and human interactomes providing insights into the impact of the two organisms on each other (Uetz *et al.*, 2006).

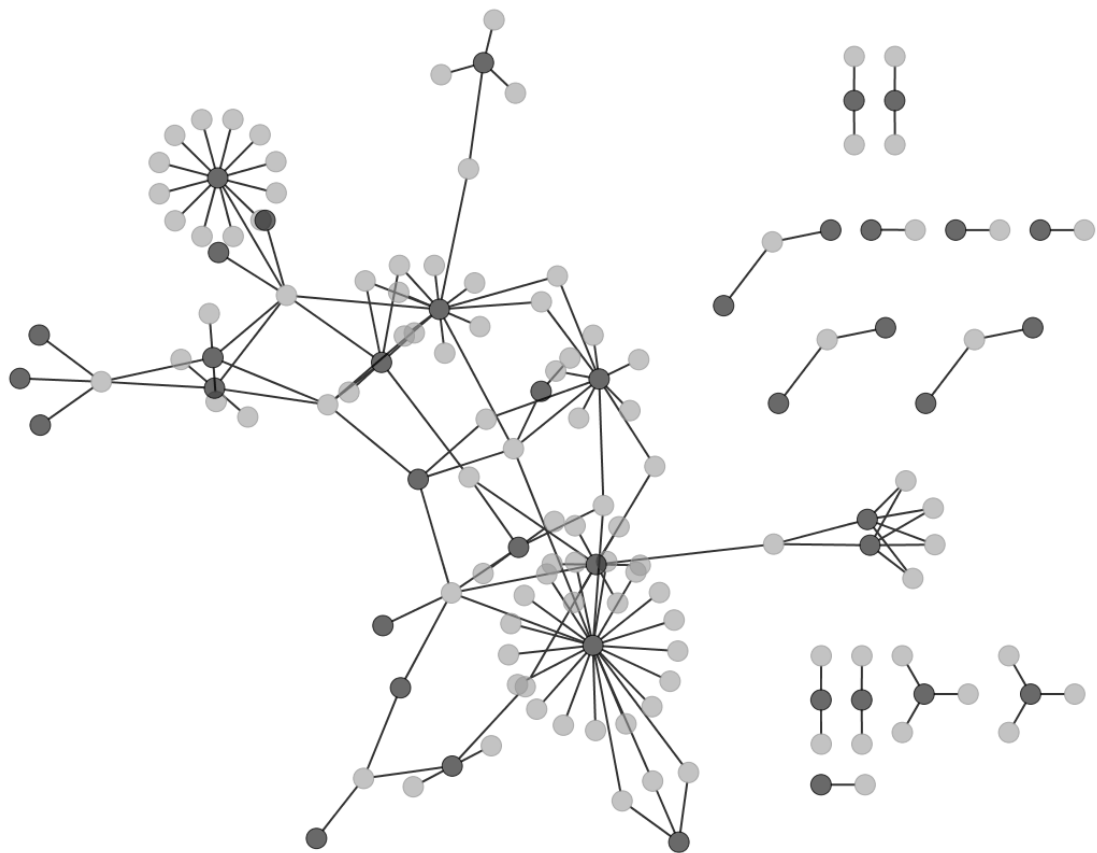


Figure 2.2. PHI map of EBV-human obtained from PHISTO. Dark grey nodes are EBV proteins and light grey nodes are human proteins. The figure was drawn in Cytoscape version 2.8.2.

A proteome-wide PHI map for the Flavivirus HCV was mapped by two-hybrid method and then by literature mining of previously found interactions between HCV and human (de Chassey *et al.*, 2008). A map of 481 interactions between 11 HCV proteins and 421 human proteins was generated (314 PHIs by two-hybrid). 65% of this PHI network included novel interactions. The integrated human network of 44,223 PPIs among 9520 proteins (Gandhi *et al.*, 2006) was used to evaluate the interplay between HCV and human. Very similar behavior to EBV (Calderwood *et al.*, 2007) was observed for HCV in terms of attacking hub and bottleneck proteins in the human network. To assess the human pathways targeted by HCV, KEGG functional annotation pathways (Aoki-Kinoshita and Kanehisa, 2007) were used. Four pathways were detected to be enriched in HCV-targeted human proteins. Three of them were associated already with HCV clinical syndromes as insulin, TGF- β and Jak/STAT pathways. The last enriched pathway, focal adhesion, is a novel observation as a human pathway affected during HCV infection (de Chassey *et al.*, 2008).

For H1N1 A/PR/8/34 strain of Influenza virus, 135 PHIs between 10 viral and 87 human proteins, most of which are expressed in primary human bronchial cells, were detected by two-hybrid (Shapira *et al.*, 2009). Some of the PHIs constructed had been published previously (Hale *et al.*, 2008). In the analysis of the influenza A-human interaction network, important properties about connectivity of proteins were observed. First, viral proteins interact with significant number of human proteins, reflecting the multifunctionality of the small number proteins encoded in RNA viruses. Second, each of 24 human proteins connects with two or more viral proteins forming virus-human multiprotein complexes. Additionally, it was observed that viral proteins generally target human proteins which are highly connected within their own network, as it was the case in Herpesviruses-human system (Calderwood *et al.*, 2007). In Shapira *et al.* (2009) another PHI network was identified for strain of Influenza virus, H3N2 A/Udorn/72 by the same experimental approach. This PHI network consists of 81 interactions between 10 viral and 66 human proteins, reflecting a similar nature to the network for H1N1 strain-human system. This confirms the conserved functions of Influenza virus proteins through strains. Besides direct physical interactions between viral and human proteins, host responses in bronchial cells to influenza infection was identified by expression profiling, generating a regulatory map of interactions between influenza proteins and their human targets.

Comprehensive analysis of the physical and regulatory maps of the PHI system elucidated human mechanisms involved in infection. For example, NF- κ B, mitogen-activated protein kinase, apoptosis, and Wnt signaling pathways are regulated through transcriptional and/or physical interactions during influenza A infection.

One of the most dangerous human pathogens, HIV, belongs to positive-sense single-stranded RNA virus family retroviridae. Acquired immunodeficiency syndrome-causing HIV has been extensively studied since its first observation near the end of the 20th century (Brennan and Durack, 1981; Weiss, 1993; Stumptner-Cuvelette *et al.*, 2001; Klase *et al.*, 2009). Similar to other RNA viruses, HIV has a small genome and depends largely on human cellular machinery to be replicated. Identifying the physical contacts between HIV and human proteins during HIV replication is critically important for a full understanding of HIV infection. Being one of the most studied pathogens, there are many PHI data for HIV-1. However, the current PHI data have been produced mainly by small-scale experiments (Nisole *et al.*, 2002; Yedavalli *et al.*, 2005; He *et al.*, 2006). Very recently, a global PHI network was generated for HIV-human protein complexes by affinity tagging and purification mass spectrometry, producing 497 PHIs between 16 HIV-1 proteins and 435 human proteins (Jager *et al.*, 2012). It was observed that HIV-targeted human proteins are highly conserved across primates. The novel interactions identified in that study requires further work to detail their biological significance in terms of HIV infection. Besides whole proteins, domains of the interacting proteins were investigated and the enriched domain types in targeted human proteins were indicated for facilitating future structural modeling studies regarding HIV-human system.

The first large-scale interaction networks between viruses and human (Table 2.2) provide crucial clues about the viral infections, verifying the critical importance of PHI analyses in infection researches.

2.2.2. Bacteria-Human Interaction Networks

Until very recent years, the PHI data were scarce for bacterial systems because of lack of any large-scale experiments. The first extensive bacterial PHI networks were identified for important human pathogens, *Bacillus anthracis*, *Francisella tularensis*, and

Yersinia pestis (Dyer *et al.*, 2010), then another high-throughput experimental study generating PHI data of *Y. pestis* was reported (Yang *et al.*, 2011).

Gram-positive bacteria *B. anthracis* and *Y. pestis* and gram-negative bacterium *F. tularensis* are respiratory pathogens causing anthrax, bubonic plague, and acute pneumonic disease, respectively. Using a two-hybrid assay, large-scale interaction data were generated between these bacteria and human providing 3073 PHIs between 943 *B. anthracis* proteins and 1748 human proteins, 4059 PHIs between 1218 *Y. pestis* proteins and 2108 human proteins, and 1383 PHIs between 349 *F. tularensis* proteins and 999 human proteins (Dyer *et al.*, 2010). The first conclusion of computational analyses of these comprehensive bacteria-human networks, in combination with the integrated human PPI network from databases BIND, DIP, HPRD, IntAct, MINT, MIPS, and Reactome, was that bacterial proteins tend to target hubs and bottlenecks in the human network. Secondly, the roles of human proteins targeted by these bacteria were investigated using their gene ontology annotations (Ashburner *et al.*, 2000). The tendency of all three pathogens to target human proteins involved in immune responses was observed as previously reported (Lai *et al.*, 2001; Park *et al.*, 2002; Zhang *et al.*, 2005). Besides being effectors of immune signaling, the bacteria-targeted human proteins also have crucial roles in apoptosis (Dyer *et al.*, 2010). Thirdly, the conserved protein interaction modules of the three PHI networks were computed (Flannick *et al.*, 2006; Sharan *et al.*, 2005) for a more systematic comparative analysis. Conserved modules revealed common attacks by the bacterial pathogens to same human pathways.

Subsequently, another PHI map was generated for plague causing *Y. pestis* by a different two-hybrid strategy by choosing only potential virulence factors as bait proteins (Yang *et al.*, 2011). 204 PHIs were yielded between 66 *Y. pestis* proteins and 109 human proteins and then 23 previously reported PHIs were integrated to construct a comprehensive network between *Y. pestis* and human. A graph theoretic analysis confirms that *Y. pestis* preferentially targets hub and bottleneck proteins in the human intranetwork as concluded previously for viruses (Calderwood *et al.*, 2007; Shapira *et al.*, 2009) and bacteria (Dyer *et al.*, 2010). Signaling pathways, crucial for human immune system, were found to be enriched in human proteins targeted by *Y. pestis*. These pathways include mitogen-activated protein kinase signaling and Toll-like receptor signaling and also

pathways functioning in focal adhesion, regulation of cytoskeleton, and leukocyte transendothelial migration. Finally, *Y. pestis*-targeted human proteins were compared with those targeted by viruses whose PHI networks were identified previously. 16 of 109 *Y. pestis*-targeted human proteins are included in PHI networks of EBV (Calderwood *et al.*, 2007) and HCV (de Chassey *et al.*, 2008) indicating the common infection strategies of both viruses and bacteria.

The recent detected first large-scale PHI networks of bacteria-human systems (Table 2.2) contribute largely to the understanding of bacterial infection mechanisms with immune evasion.

Table 2.2. Examples of PHI networks detailed in the literature.

Pathogen name	Pathogen type	# of PHIs	# of pathogen proteins	# of human proteins	References
EBV	DNA virus	173	40	112	Calderwood <i>et al.</i> , 2007
HCV	RNA virus	481	11	421	de Chassey <i>et al.</i> , 2008
Influenza A virus (H1N1 A/PR/8/34)	RNA virus	135	10	87	Shapira <i>et al.</i> , 2009
Influenza A virus (H3N2 A/Udorn/72)	RNA virus	81	10	66	Shapira <i>et al.</i> , 2009
<i>B. anthracis</i>	Gram+ bacteria	3073	943	1748	Dyer <i>et al.</i> , 2010
<i>Y. pestis</i>	Gram+ bacteria	4059	1218	2108	Dyer <i>et al.</i> , 2010
<i>F. Tularensis</i>	Gram- bacteria	1383	349	999	Dyer <i>et al.</i> , 2010
<i>Y. pestis</i>	Gram+ bacteria	204	66	109	Yang <i>et al.</i> , 2011
HIV	RNA virus	497	16	435	Jager <i>et al.</i> , 2012

2.2.3. Analysis of Comprehensive PHI Data

As PHI data available for various pathogens increase, a need to analyze comprehensive PHI data for all pathogen types together arises in order to draw a generalized picture. Although infection mechanisms of individual pathogens have been studied through intraspecies pathogenic PPI maps and interspecies PHI maps, a general overview of infection mechanisms was missing until analysis of PHI data from different infection agents was attempted (Dyer *et al.*, 2008).

In the absence of large-scale PHI networks for bacterial, protozoan and fungal systems, Dyer *et al.* (2008) performed the first global analysis of 10,477 protein interactions between 190 pathogen strains of viruses, bacteria, protozoa, and human through properties of targeted 1233 human proteins. Diversity of the available PHI data was not rich, 98.3% of 10,477 PHIs belonged to the virus-human systems with 77.9% of the interaction data drawn from HIV-human interaction systems. The importance of the pathogen-targeted proteins was evaluated within the intraspecies human PPI network of 75,457 interactions. These PHI and PPI data were integrated from public databases; MINT, IntAct, DIP, HPRD, Reactome, BIND and MIPS. Firstly, targeting hub and bottleneck proteins was concluded to be global behavior for all pathogens, as reported for individual pathogen strains previously (Calderwood *et al.*, 2007; Shapira *et al.*, 2009; Dyer *et al.*, 2010; Yang *et al.*, 2011). Gene ontology (Ashburner *et al.*, 2000) functions enriched in the targeted human proteins by different pathogens revealed common infection mechanisms. Attack of human transcription factors and key proteins that control the cell cycle and regulate apoptosis and transport of genetic material across the nuclear membrane were found to be among the common viral strategies. Despite its scarcity (174 interactions in the dataset), bacterial PHI data allowed identification of specific human proteins that function in the host immune response (via Toll-like receptors and I- κ B kinase/NF- κ B signaling cascade) as a target of bacterial infection strategy (Dyer *et al.*, 2008).

Global analysis of PHI data provides insights into the strategies adapted by bacteria and viruses to subvert human cellular processes and immune system for the infection. However, large-scale PHI networks for pathogens other than bacteria and viruses are still undetermined, leaving their pathogenesis mechanisms to be relatively uncharacterized.

2.2.4. The Impacts of PHI Networks

Research on infectious diseases through PHIs has accelerated within the post-genomic era (Figure 2.3). However, large-scale PHI networks have been infrequently studied. Efforts to identify and analyze large-scale PHIs for diverse pathogen types would be expected to parallel the acceleration of biotechnology and bioinformatics research. Increasing amounts of data available will allow more complete data sets to be compiled, resulting in characterization of topological properties of PHI networks. The first attempt to fit the degree distribution of EBV-human interaction network failed due to scarcity of data (Calderwood *et al.*, 2007). On the other hand, bioinformatics analyses of the pathogen-targeted human proteins succeeded in unraveling some infection strategies such as targeting human hubs and bottlenecks, manipulating cellular processes for the usage of pathogens' own advantages and evasion of immune defenses (Calderwood *et al.*, 2007; Dyer *et al.*, 2008; de Chassey *et al.*, 2008; Dyer *et al.*, 2010; Yang *et al.*, 2011). The huge amount of data expected to be generated for PHI systems will enable us to capture all details of infection processes, potentially leading to the development of new and more efficient therapeutics.

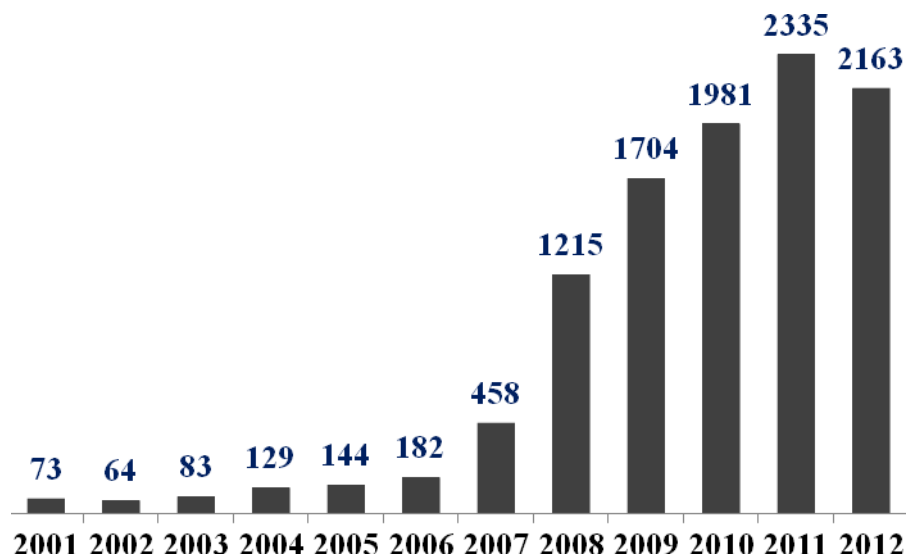


Figure 2.3. Number of publications including PHI-related terms in PubMed in the post genomic era.

Conventional treatments for infectious diseases often aim to kill pathogens by targeting their essential proteins. This approach unfortunately forces the pathogens to evolve for survival and consequently selects resistant strains (especially in the case of RNA viruses with a high mutation rate). To fight drug-resistant pathogens, novel alternative therapeutics are emerging which target host proteins required by pathogens to replicate and persist within the host organism. If these host factors are indispensable for pathogens, but not essential for host cells, their silencing may inactivate pathogenic activity, allowing them to serve as therapeutic targets (Bünnige, 2001; Murali *et al.*, 2011). In the light of PHI studies, some human factors required by viral and bacterial pathogens have been determined for HIV (Brass *et al.*, 2008; König *et al.*, 2008; Zhou *et al.*, 2008; Bushman *et al.*, 2009; Murali *et al.*, 2011), HCV (Li *et al.*, 2009), West Nile virus (Krishnan *et al.*, 2008), Influenza virus (Karlsson *et al.*, 2010; König *et al.*, 2010), and *M. tuberculosis* (Kumar *et al.*, 2010) in recent years. The identified host factors do not yet serve as targets for any anti-infection drug currently, requiring further experimental studies.

3. PHISTO: PATHOGEN-HOST INTERACTION SEARCH TOOL

The availability of the sequenced human genome as well as the complete sequences of a number of microorganisms has significantly contributed to the understanding of their proteomes, enabling the advances in high-throughput protein interaction detection methods. This, in turn, has led to the production of large-scale interspecies protein interaction data of pathogen-human systems. The next crucial step is to make these PHI data publicly available from one single source for systems pharmacology research to contribute to efforts focusing on improving the global understanding of drug targets, suggesting new targets, providing a deeper understanding of the effects of drugs, and developing new therapeutics for infectious diseases. However, public resources for studying pathogen-host systems do not offer sufficient platforms to access updated PHI data comprehensively. Since the produced PHI data are scattered throughout various protein interaction databases, research in the field of interaction-based infection strategies of pathogens requires a careful and tedious data integration work.

The current PHI databases do not provide comprehensive and updated data; they are mostly specific to some pathogens. VirHostNet (Navratil *et al.*, 2009) and VirusMINT (Chatr-aryamontri *et al.*, 2009) present data only about virus-host interaction systems. On the other hand, PATRIC (Gillespie *et al.*, 2011) presents data only about bacterial systems. There are other informative databases for PHI systems such as Phi-base (Winnenburg *et al.*, 2008) and PHIDIAS (Xiang *et al.*, 2007) which store information about pathogenicity, virulence and effector genes, genome sequences of pathogens, conserved domains of proteins and gene expression data without a focus on pathogen-host protein interactions. The only available resource to access pathogen-human protein interaction data for all pathogen types in a single database is HPIDB (Kumar and Nanduri, 2010) which does not offer any additional functionality to analyze PHI networks. At this point, PHISTO (Durmuş Tekir *et al.*, 2013b) has been developed in the framework of this thesis to serve as an up-to-date and functionally-enhanced source of comprehensive PHI data through a user-friendly interface (www.phisto.org).

PHIs in PHISTO are imported from several PPI databases using the PSICQUIC tool (Aranda *et al.* 2011). Text mining is used to label PHIs extracted without any information on interaction detection method. Tools for visualization of small PHI networks and graph-theoretical analysis of targeted human proteins may enable users to gain crucial insights on infection mechanisms. The BLAST interface offers to search for orthologous PHIs for pathogens lacking experimental data.

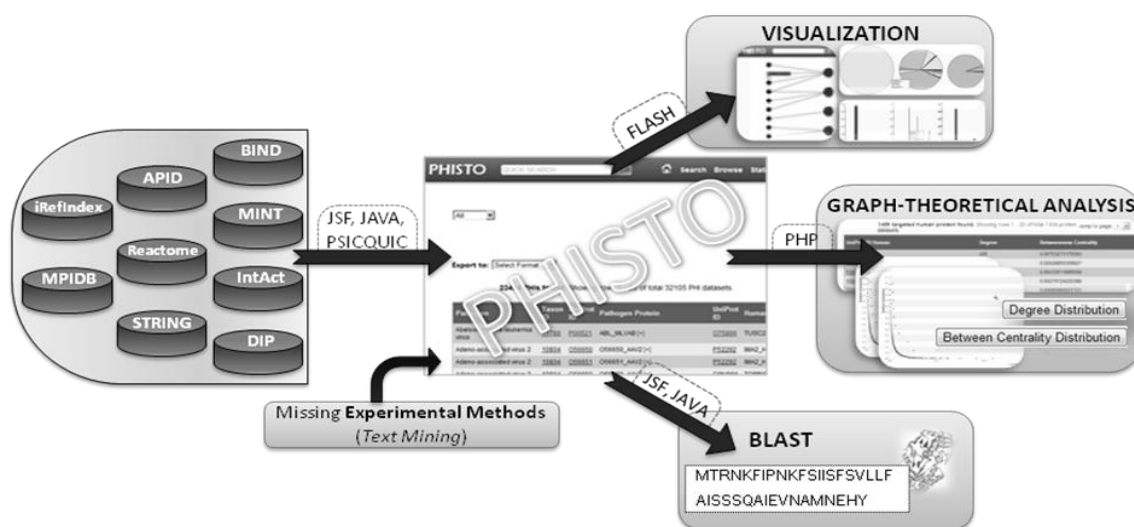


Figure 3.1. The architecture of PHISTO.

PHISTO is designed as a Web-accessible platform with two-tier architecture. The back tier is a MySQL-based database. The front tier is a PHP- and Javascript-based user interface that runs on an Apache Web server. Figure 3.1 illustrates the architecture of PHISTO.

3.1. Data Extraction and Integration

The data stored in PHISTO are extracted and integrated from nine publicly available databases: APID (Prieto *et al.*, 2006), BIND (Alfarano *et al.*, 2005), DIP (Salwinski *et al.*, 2004), IntAct (Kerrien *et al.*, 2012), iRefIndex (Razick *et al.*, 2008), MINT (Licata *et al.*, 2012), MPIDB (Goll *et al.*, 2008), Reactome (Croft *et al.*, 2011) and STRING (Szklarczyk *et al.*, 2011). The PSICQUIC tool (Aranda *et al.*, 2011) is used, which is a service to standardize and automate the access to data in molecular interaction databases. Data are extracted in MITAB format, a standardized format for molecular interaction data (Kerrien

et al., 2007). PHISTO database contains two types of data: The user-accessible PHI data (interspecies protein interaction data between pathogen proteins and human proteins), and human PPI data (intraspecies protein interaction data) that cannot be directly accessed by the users, but are used by the Graph Analysis tool.

To present protein interaction data in a consistent format, Uniprot IDs (Apweiler *et al.*, 2004) and names of interacting proteins, taxonomy IDs and names of pathogens, experimental methods, and PubMed IDs of literature references are collected. The data are stored in separate database tables with one PHI table containing the core data (UniProt IDs, Taxonomy IDs, Experimental Methods, PubMed IDs) and the others being mostly ID-related tables. Thanks to the implementation of ID-based core data, standardized data formats, and relational data tables, the PHISTO database can easily be maintained and enhanced. It is automatically updated by the Java-based off-line application on a monthly basis.

To determine the pathogen list in PHISTO, first PPIs between human and organisms other than human were extracted. In other words, all available interspecies protein interactions for human were collected. Then, the organisms whose proteins interact with human proteins were investigated manually one by one to determine whether they are pathogenic or not. Pathogenic microorganisms were labeled as "pathogen" and the remaining organisms were labeled as "non-pathogen". Therefore, two lists of organisms are maintained, for which experimentally-verified protein interactions with human are available. In each PHI update, all interspecies protein interactions of human are collected from PPI databases. After the data collection step, the off-line application compares the newly collected protein interaction data between human and the organisms in the pathogen list and the novel ones are added to PHISTO as new PHIs. Then, the organisms in the newly collected data that are not included in both of the lists of pathogens and non-pathogens are manually investigated for their pathogenicity one by one. The pathogenic ones are added to the pathogen list and their protein interactions are added to PHISTO as new PHIs, increasing the number of pathogenic organisms in PHISTO. The non-pathogenic ones are added to the non-pathogen list so that there is no need to investigate their pathogenicity in the following updates.

3.1.1. Data Update System

The off-line system (developed in July 2012) updates PHISTO database periodically in an automatic fashion or manually. Here is a step-by-step explanation of how the PHISTO update system works:

- (i) The update system automatically accesses the databases via the PSICQUIC protocol. It extracts the related data based on taxonomy IDs of organisms, and downloads them.
- (ii) It processes the downloaded data by removing duplicate entries. To assure uniqueness in protein names, UniProt IDs (Apweiler *et al.*, 2004) of proteins are used.
- (iii) It uploads the processed data into the PHISTO database.

The functional user interface (Figure 3.2) for the off-line update system allows the administrator to control the data update process via following features:

- (i) Start automatic update
- (ii) Start or stop update process manually
- (iii) Schedule periodic automatic updates
- (iv) Determine the list of databases to be used in updates
- (v) Perform independent update of only PHI data or only human PPI data.

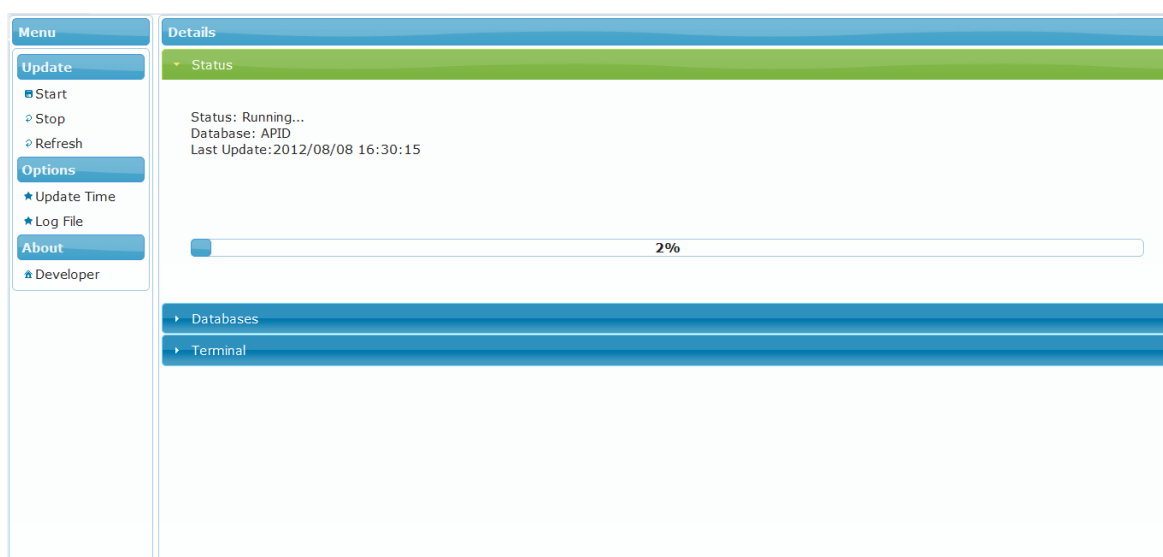


Figure 3.2. User Interface of the off-line application for data update.

3.1.2. Text Mining of Missing Experimental Methods in PHI Data

While collecting PHI data from PPI databases, the extracted information for each data: Uniprot IDs of interacting two proteins, taxonomy ID of pathogenic organism, experimental method(s) used to detect protein interaction, and PubMed ID of literature reference(s). PHIs without a PubMed ID and PHIs in which a protein does not have a Uniprot ID are not collected. In other words, PHI data which lack any of these five identifiers (two Uniprot IDs of interacting pathogen protein and human protein, taxonomy ID of pathogen, experimental method, PubMed ID) are not collected to be able to store consistent full data. Then, protein names are extracted from Uniprot database and pathogen names are extracted from NCBI taxonomy database, using Uniprot IDs and taxonomy IDs, respectively. These information are stored in ID-related tables of PHISTO (Uniprot ID vs. protein names and taxonomy ID vs. pathogen name). An ID-related table for experimental methods was also constructed as PSI-MI ID (Kerrien *et al.*, 2007) versus experimental method name to be able to present consistent method names avoiding redundancy. Only general names listed for each PSI-MI ID are used in PHISTO (e.g. "yeast two hybrid", "two hybrid", "2 hybrid", "Y2H", etc. for the same method can be used together in a resource, however in PHISTO only one name is used for an experimental method such as using only "two hybrid" for PSI-MI ID: 18). Despite having ID information also for experimental methods, in the search results, only the method names are presented without IDs for a clear presentation. Finally, eight identifiers (taxonomy ID and name of pathogen, Uniprot ID and name of pathogen protein, Uniprot ID and name of human protein, experimental method and PubMed ID in order) are presented for each PHI data in the search results.

The analysis of contents of the PPI databases from which PHI data are extracted, revealed that they include lots of PPI data which lack some of the mentioned identifiers. PHI data which lack only "experimental method" identifier were also collected in June, 2012. 12,751 PHI data that were not labeled with the experimental methods used to detect these interactions were extracted. Besides automatic data extraction system, manual curation of the PPI databases were performed to find PHI data without experimental method information. Then, text mining was used to annotate these data with "experimental

method names" by going to known articles (with PubMed ID information) for each such interaction data to be able to involve more comprehensive full PHI data.

A dictionary of interaction detection methods, which includes 115 different general method names with 159 synonyms, was compiled from the PSI-MI Ontology version 2.5 (Kerrien *et al.*, 2007). An experimental method may have more than one name, although one of them is used more often than the others. The widely used name is the general name, whereas the others are called its synonyms. The abstracts of the articles that contain PHIs without experimental method information were obtained from PubMed (www.ncbi.nlm.nih.gov/pubmed), using PubMed IDs. 2952 experimental method names were assigned to 2109 unique PHIs and then stored in PHISTO. These PHIs are displayed with an asterisk following the experimental method name in the search results, indicating that their experimental methods were obtained using text mining. The text mining of missing experimental methods in PHI data will be performed once in a year.

3.1.3. Data Content

Currently, PHISTO stores 23,661 PHIs between human and 300 pathogen strains (247 viral, 45 bacterial, 3 fungal, 5 protozoan) belonging to 26 viral, 18 bacterial, 2 fungal, and 1 protozoan families (see Statistics page of PHISTO for the up-to-date data content in details). Within the current PHIs, there are 14,510 virus-human, 9131 bacteria-human, 15 protozoa-human, 5 fungi-human binary interspecies protein interactions. 5083 human proteins are interacting with 2653 proteins of these pathogens.

3.2. User Interface

The functional and easy-to-use Web-based user interface (Figure 3.3) provides various search, browse, and data analysis options.

3.2.1. Data Query

The "Quick Search" option is for performing a query without a specified identifier, whereas the "Advanced Search" option can be used to search based on any selected subset

of identifiers (i.e., taxonomy ID and name of pathogen, Uniprot ID and name of pathogen protein, Uniprot ID and name of human protein, experimental method and literature reference). The “Browse” option provides users with easy access to the entire PHI data of any specified pathogen within the taxonomical classification. Thanks to the “Browse” functionality, users can query the database for a selected pathogen type (virus, bacteria, etc.), family, species or strain.

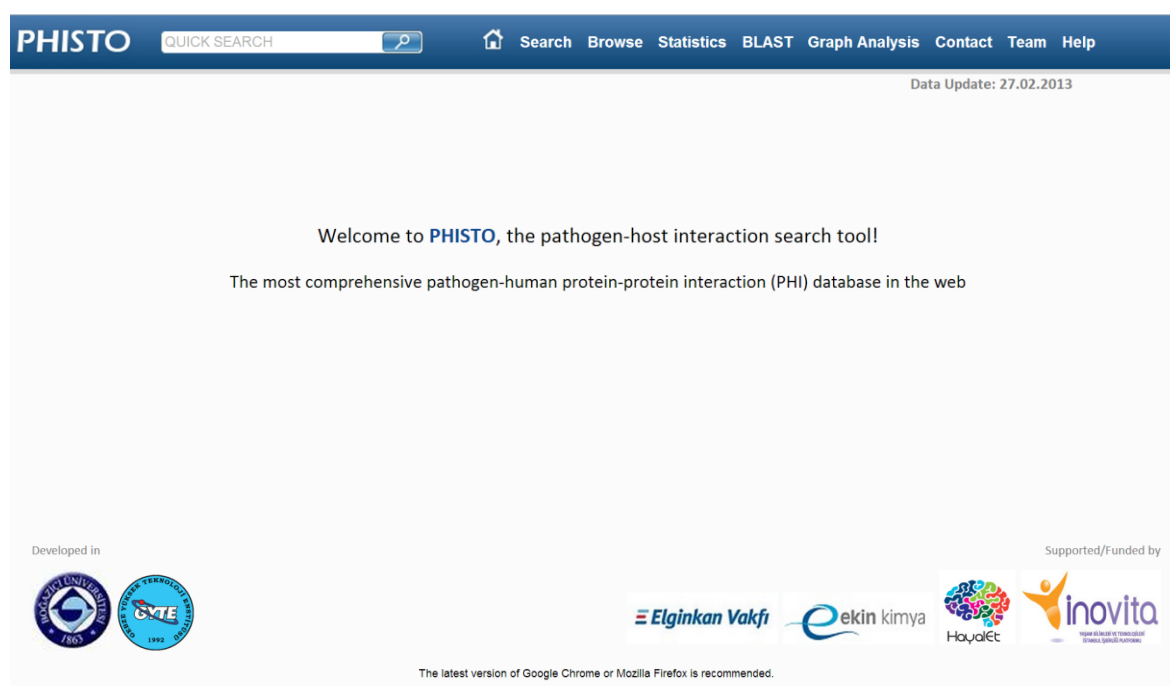
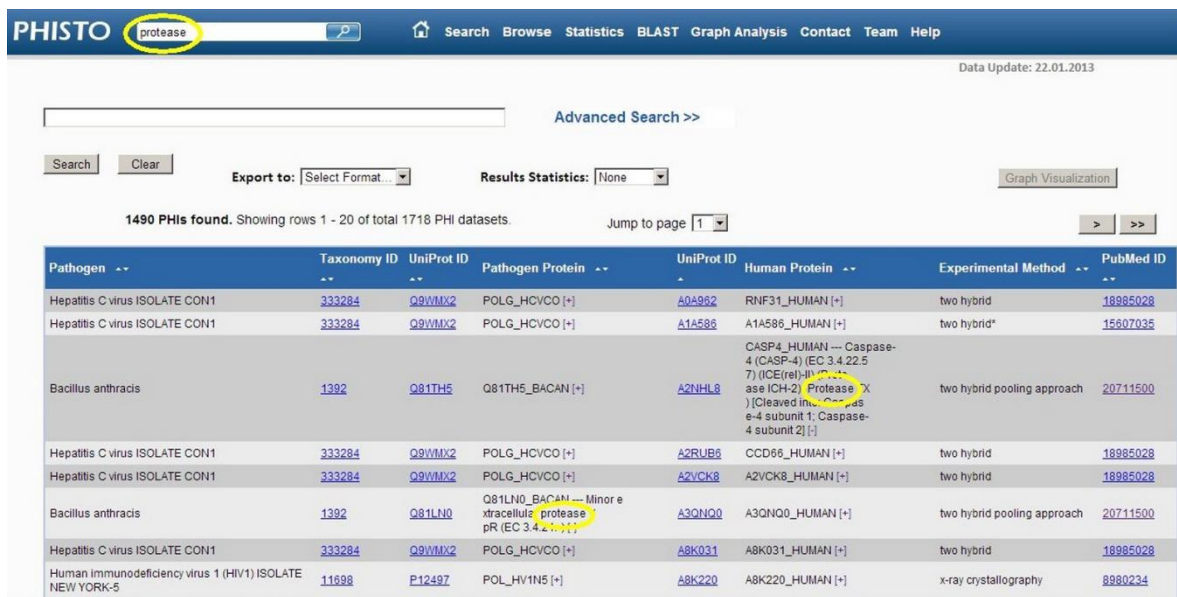


Figure 3.3. Homepage of PHISTO user interface.

For a quick search, user can enter whatever s/he wants into "Quick Search" box in the top menu or into "Simple Search" box in the "Search" page. The term will be searched in any of eight identifiers of PHI data. In the example below, the word "protease" was found in the long names of pathogen/human proteins (Figure 3.4).

For an advanced search, user can select a specific identifier (eg. pathogen name), to make search only on this identifier (eg. *Escherichia coli*). Then user can combine multiple identifiers into one search (eg. PHIs for *Escherichia coli*, but only those found by pull down or two hybrid as experimental method). If expressed in terms of AND and OR: pathogen name: *Escherichia coli* AND (method: pull down OR method: two hybrid). User can

arrange the parantheses in this search by using the number selections. A new parantheses means a new level (and hence marked by two in the search). If there was another inner bracket, it would be marked by three. Above is the search results of this example (Figure 3.5).



Pathogen	Taxonomy ID	UniProt ID	Pathogen Protein	UniProt ID	Human Protein	Experimental Method	PubMed ID
Hepatitis C virus ISOLATE CON1	333284	Q9WMX2	POLG_HCVCO [+]	A0A952	RNF31_HUMAN [+]	two hybrid	18985028
Hepatitis C virus ISOLATE CON1	333284	Q9WMX2	POLG_HCVCO [+]	A1A586	A1A586_HUMAN [+]	two hybrid*	15607035
Bacillus anthracis	1392	Q81TH5	Q81TH5_BACAN [+]	A2NHL8	CASP4_HUMAN -- Caspase-4 (CASP-4) (EC 3.4.22.57) (ICE(rel)-like) (Caspase-4 subunit 1, Caspase-4 subunit 2) [+]	two hybrid pooling approach	20711500
Hepatitis C virus ISOLATE CON1	333284	Q9WMX2	POLG_HCVCO [+]	A2RUB6	CCD66_HUMAN [+]	two hybrid	18985028
Hepatitis C virus ISOLATE CON1	333284	Q9WMX2	POLG_HCVCO [+]	A2VCK8	A2VCK8_HUMAN [+]	two hybrid	18985028
Bacillus anthracis	1392	Q81LN0	Q81LN0_BACAN -- Minor extracellular protease pR (EC 3.4.21.1) [+]	A3QNG0	A3QNG0_HUMAN [+]	two hybrid pooling approach	20711500
Hepatitis C virus ISOLATE CON1	333284	Q9WMX2	POLG_HCVCO [+]	A8K031	A8K031_HUMAN [+]	two hybrid	18985028
Human immunodeficiency virus 1 (HIV1) ISOLATE NEW YORK-5	11698	P12497	POL_HV1N5 [+]	A8K220	A8K220_HUMAN [+]	x-ray crystallography	8980234

Figure 3.4. Quick Search example.

"Browse" functionality allows user to select PHIs of a pathogen type, family, species or strain. For instance, user may want to obtain PHIs for type: virus, family: retroviridae, species: HIV2, strain: all (Figure 3.6). For such a browse request, user selects virus as pathogen type from the first drop-down menu, then select retroviridae as virus family from the second drop-down menu since HIV2 is a retrovirus. Then s/he selects HIV2 as retroviridae species from the third drop-down menu, finally selects "All" from the strains menu to obtain PHI data belonging to all strains of HIV2 species.

3.2.2. Data Presentation

The PHI search results are presented in a clear and consistent ID-based formatted table, which includes information about the eight identifiers mentioned previously. Sorting of the PHI data rows in the results is available with respect to each title, which may be very useful in investigating the results. Proteins, pathogens and publications listed in the results are linked to external databases; UniProt, NCBI Taxonomy, and PubMed respectively

offering users quick navigation in these informative databases. The search results can be exported to a variety of file formats (txt, tab, csv, xls). The long names of the interacting pathogen/human proteins can be visualized by clicking on "[+]" sign just after the short protein names as shown in the first row of the example below (Figure 3.7).

PHISTO QUICK SEARCH

[Home](#) [Search](#) [Browse](#) [Statistics](#) [BLAST](#)

[Simple Search >>](#)

1 AND Escherichia coli in Pathogen

2 AND pull down in Observation Method

2 OR two hybrid in Observation Method

[Add another field >>](#)

Export to: Select Format... Results Statistics: None

14 PHIs found. Showing rows 1 - 15 of total 15 PHI datasets.

Pathogen	Taxonomy ID	UniProt ID	Pathogen Protein	UniProt ID	Human Protein	Experimental Method	PubMed ID
Escherichia coli	562	Q5WMC9	Q5WMC9_ECOLX [+]	P18333	NCK1_HUMAN [+]	pull down	11533668
Escherichia coli	562	P71213	P71213_ECOLX [+]	P01009	A1AT_HUMAN [+]	pull down	15271889
Escherichia coli	562	P71213	P71213_ECOLX [+]	E9KL23	E9KL23_HUMAN [+]	pull down	15271889
Escherichia coli	562	Q9R8E4	Q9R8E4_ECOLX [+]	P60953	CDC42_HUMAN [+]	pull down	19620963
Escherichia coli K-12	83333	P00274	THIO_ECO57 [+]	P24941	CDK2_HUMAN [+]	two hybrid	8606778
Escherichia coli K-12	83333	P39171	ISCS_ECO57 [+]	P16070	CD44_HUMAN [+]	two hybrid	11983079
Escherichia coli K-12	83333	P04867	ISCS_ECOLI [+]	P16070	CD44_HUMAN [+]	two hybrid	11983079
Escherichia coli K-12	83333	P04A25	THIO_ECOLI [+]	P24941	CDK2_HUMAN [+]	two hybrid	8606778
Escherichia coli K-12	83333	Q9R8E4	Q9R8E4_ECOLX [+]	P60953	CDC42_HUMAN [+]	pull down	19620963
Escherichia coli O127:H6 STRAIN E2348 / 69	574521	B7UM88	B7UM88_ECO27 [+]	Q9Y5X1	SNX9_HUMAN [+]	two hybrid	16585770
Escherichia coli O127:H6 STRAIN E2348 / 69	574521	B7UM99	TIR_ECO27 [+]	Q14247	SRC8_HUMAN [+]	pull down	19419567
Escherichia coli O157:H7	83334	P62816	ISPE_ECO57 [+]	Q8TEQ6	GEMI5_HUMAN [+]	pull down	11714716
Escherichia coli O157:H7	83334	Q7DB74	Q7DB74_ECO57 [+]	Q15085	ARHGB_HUMAN [+]	pull down	20300064
Escherichia coli O157:H7	83334	Q7DB74	Q7DB74_ECO57 [+]	Q92888	ARHG1_HUMAN [+]	pull down	20300064
Escherichia coli O157:H7	83334	Q7DB74	Q7DB74_ECO57 [+]	Q9NZN5	ARHGC_HUMAN [+]	pull down	20300064

Figure 3.5. Advanced Search example.

The "Graph Visualization" option allows the resulting PHI network to be represented as a bipartite graph. After making a search or browse action, the use of "Graph Visualization" button on the results page provides the corresponding bipartite graph in a new window (Figure 3.8). Both the nodes (proteins) and edges (interactions) in the graphs are clickable with corresponding links to UniProt database for protein information and to PubMed database for publication information, respectively.

PHISTO QUICK SEARCH

Pathogen Type: Virus

Family: Retroviridae

Species: Human immunodeficiency virus 2 (HIV2)

Strain: **Select**

- Select
- ALL
- Human immunodeficiency virus 2 (HIV2)
- Human immunodeficiency virus 2 (HIV2) ISOLATE BEN
- Human immunodeficiency virus 2 (HIV2) ISOLATE CAM2
- Human immunodeficiency virus 2 (HIV2) ISOLATE D194
- Human immunodeficiency virus 2 (HIV2) ISOLATE D205
- Human immunodeficiency virus 2 (HIV2) ISOLATE GHANA-1
- Human immunodeficiency virus 2 (HIV2) ISOLATE KR
- Human immunodeficiency virus 2 (HIV2) ISOLATE NIH-Z
- Human immunodeficiency virus 2 (HIV2) ISOLATE ROD
- Human immunodeficiency virus 2 (HIV2) ISOLATE SBLISY
- Human immunodeficiency virus 2 (HIV2) ISOLATE ST

Figure 3.6. Browse example.

Export to: **Select Format...** (TXT, TAB, CSV, XLS) | Results Statistics: None | Graph Visualization

Showing rows 1 - 20 of total 42 PHI datasets. | Jump to page 1

Pathogen	Taxonomy ID	UniProt ID	Pathogen Protein	UniProt ID	Human Protein	Experimental Method	PubMed ID
Shigella flexneri	623	P18011	IPAB_SHIFL --- Invasin I paB (62 kDa antigen) [-]	Q9UJ95	MD2L2_HUMAN --- Mitotic spindle assembly checkpoint protein MAD2B (Mitotic arrest deficient 2-like protein 2) (MAD2-like protein 2) (REV7 homolog) (hREV7) [-]	anti tag coimmunoprecipitation	17719540
Shigella flexneri	623	P18012	IPAC_SHIFL [+]	P35222	CTNB1_HUMAN [+]	two hybrid	12927518
Shigella flexneri	623	P18011	IPAB_SHIFL [+]	Q9UJ95	MD2L2_HUMAN [+]	two hybrid	17719540
Shigella flexneri	623	P18012	IPAC_SHIFL [+]	P35222	CTNB1_HUMAN [+]	coimmunoprecipitation	12927518
Shigella flexneri	623	P18011	IPAB_SHIFL [+]	P16070	CD44_HUMAN [+]	coimmunoprecipitation	11207560
Shigella flexneri	623	Q7BCK4	ICSA_SHIFL [+]	Q00401	WASL_HUMAN [+]	coimmunoprecipitation	9582270
Shigella flexneri	623	P18011	IPAB_SHIFL [+]	P16070	CD44_HUMAN [+]	filter binding	11207560
Shigella flexneri	623	Q6XVZ2	Q6XVZ2_SHIFL [+]	P18206	VINC_HUMAN [+]	filter binding	17289036
Shigella flexneri	623	Q7BCK4	ICSA_SHIFL [+]	Q00401	WASL_HUMAN [+]	filter binding	9582270

Figure 3.7. Sample portion of a PHI data presentation table.

Statistics of search results can be visualized via pie or bar charts, describing the distribution of PHIs over the types, families, species and strains of pathogens. For example, the distribution of the results of the quick search with keyword "protease" (Figure 3.4) are displayed in Figures 3.10 and 3.11. If user is interested to see the percentage of the searched data belonging to each pathogen type, family, species or strain, s/he can use "Results Statistics" options (Figure 3.9). A new window is opened with a pie-chart

showing the distribution of data on pathogen types. With mouse-over functionality, user can see the percentage value of each portion with corresponding pathogen types. Each portion in the chart is clickable. For example, clicking on bacteria (green) part, a new chart appears to show distribution of bacterial data in terms of its families (Figure 3.10). Sequentially, one can click each portion of this second chart to see the distribution of the corresponding bacterial family data on its species via a third chart. Similarly, clicking each portion of the third chart gives the distribution of the corresponding species on its strains via a fourth chart.

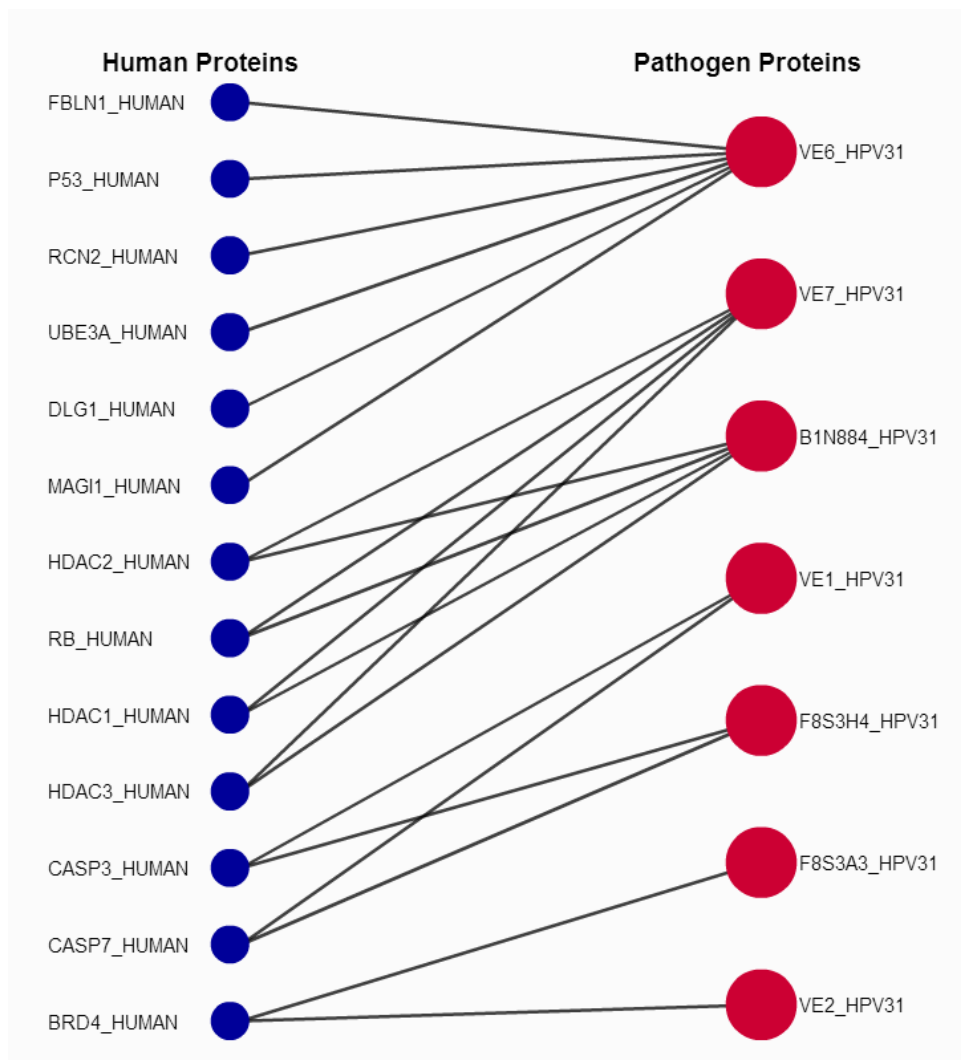


Figure 3.8. Bipartite graph of human papillomavirus 31-human PHI data.

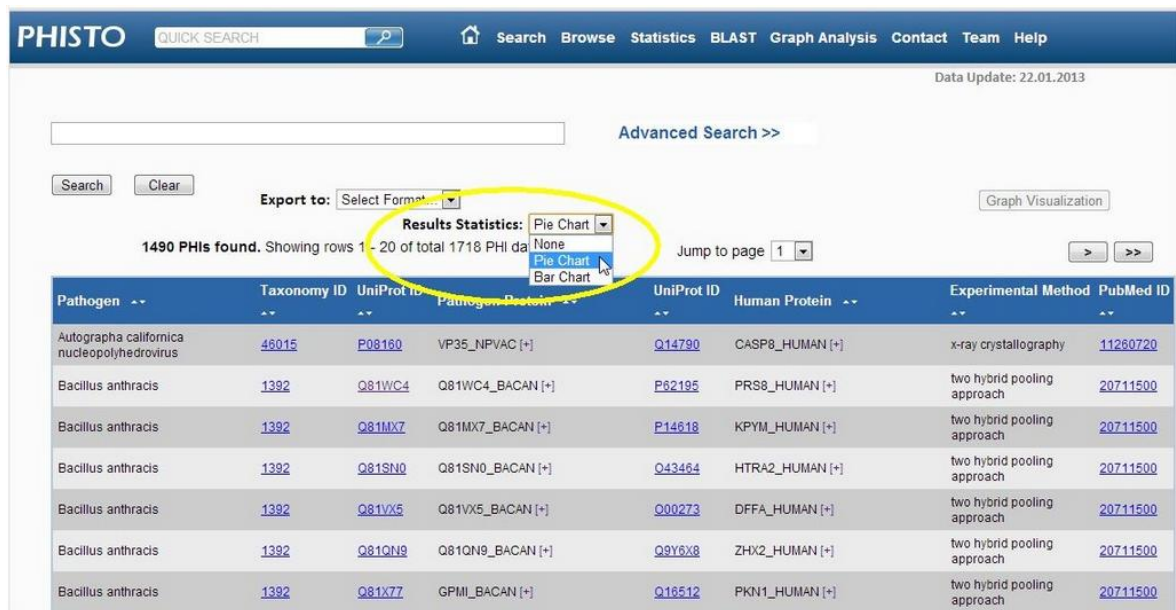


Figure 3.9. Results Statistics options.

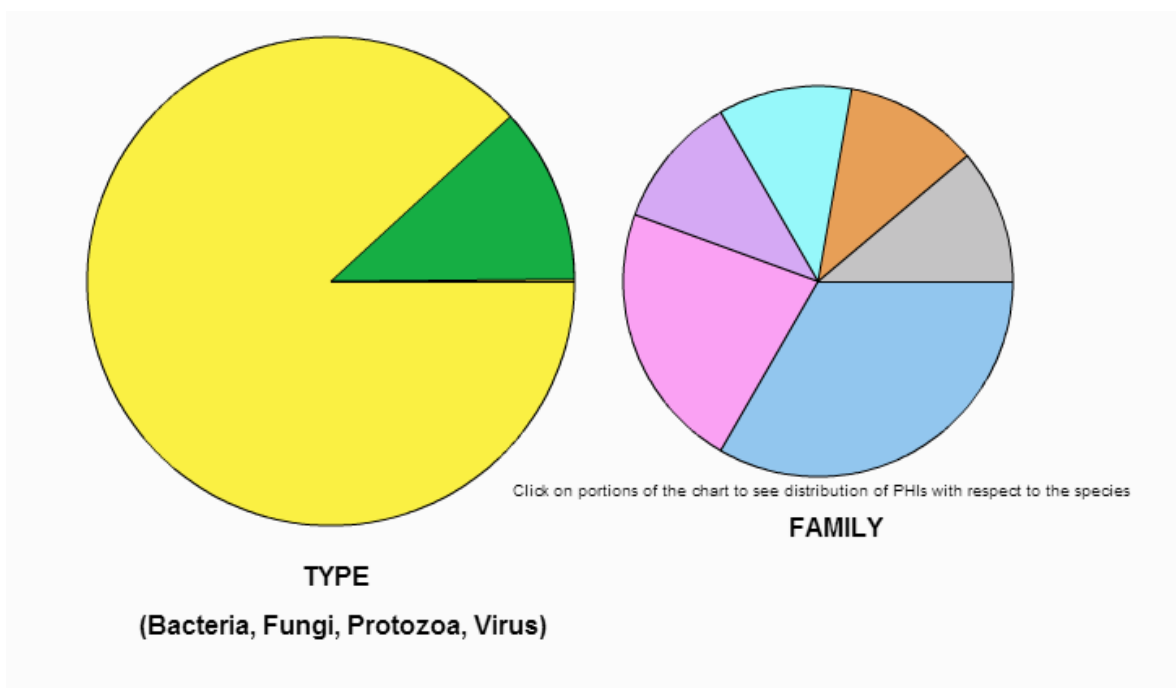


Figure 3.10. Distribution of the PHI results of Quick Search of "protease" via pie chart on pathogen types and bacterial families.

Another alternative is available for visualization of the distribution of PHIs with respect to pathogen types, families, species and strains via bar charts (Figure 3.11). Very similar explanations are valid for bar charts to observe the data statistics visually. Using

"Results Statistics" functionality via pie/bar chart allows the visualization of the data distribution for all possible cases.

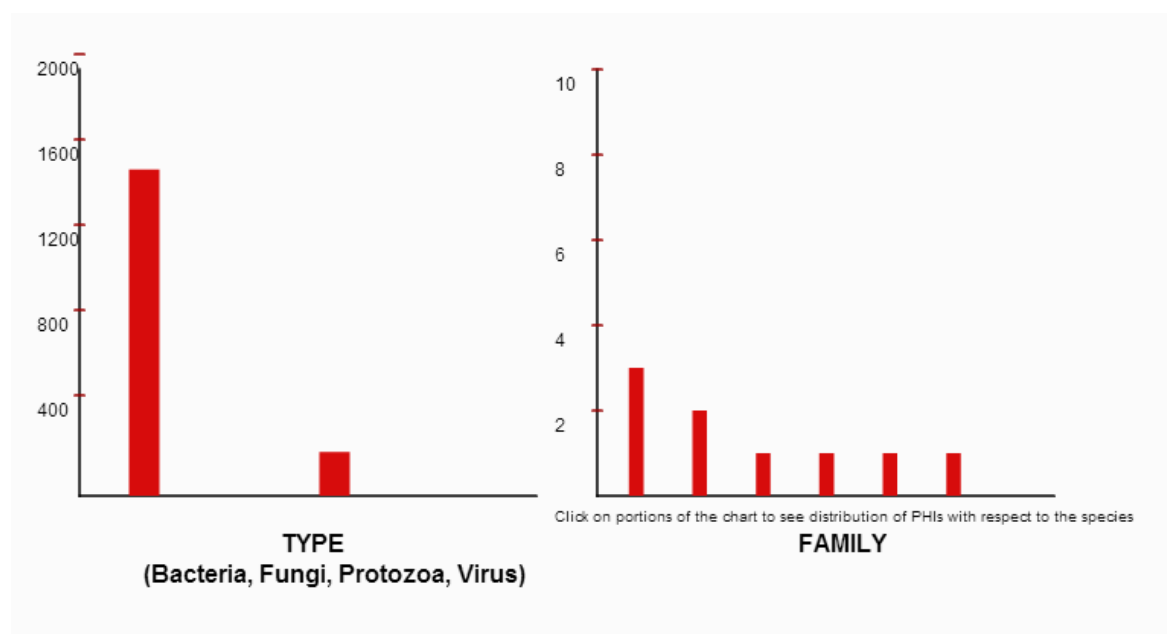
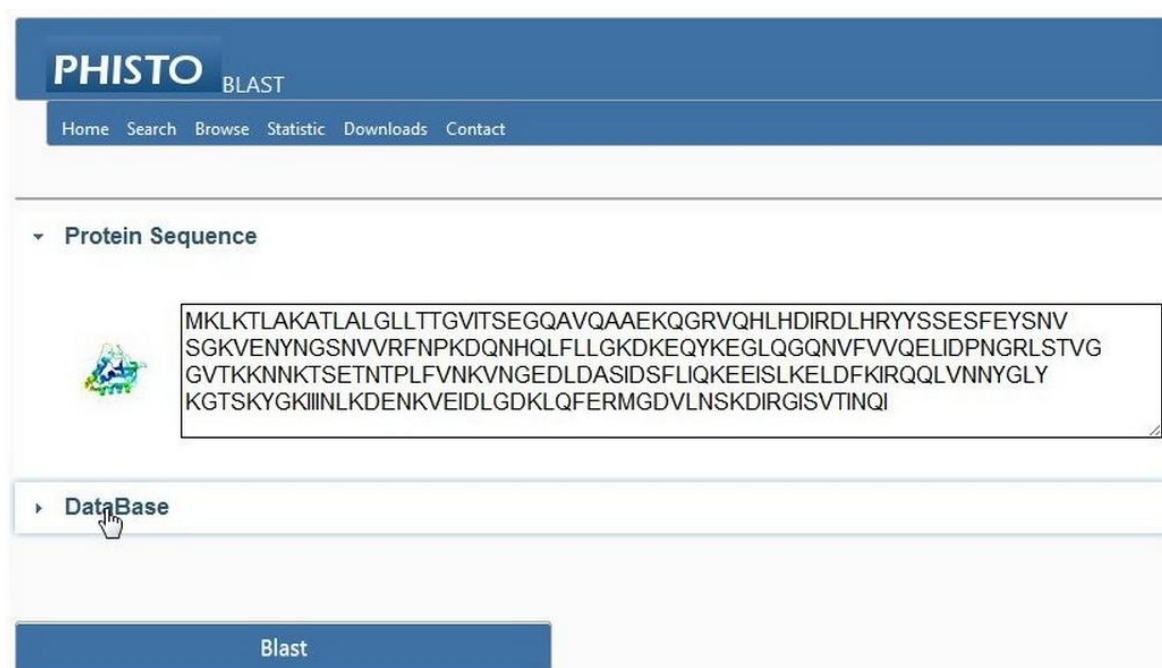


Figure 3.11. Distribution of the PHI results of Quick Search of "protease" via bar chart on pathogen types and bacterial families.

3.2.3. BLAST Search

For a protein of interest, one can search for homolog pathogen proteins in PHISTO using the "BLAST" interface. The protein sequence entered by the user is queried in the NCBI-BLAST database (<http://blast.ncbi.nlm.nih.gov>), and hits (proteins) not belonging to the pathogens covered by PHISTO are discarded. The result is presented to the user if there is any significant match with a protein belonging to pathogens available in the PHISTO database.

For a BLAST search, make the pathogen protein sequence ready in FASTA format. Paste it into the box in BLAST page (Figure 3.12). It may take some minutes to do BLAST search. The results give matched pathogen proteins (E-score <0.05) with links to UniProt page of the proteins.



PHISTO BLAST

Home Search Browse Statistic Downloads Contact

▼ Protein Sequence



```

MKLKTAKATLALGLTTGVITSEGQAVQAAEKQGRVQHLHDIRDHLHRYSSSESFEYSNV
SGKVENYNGSNVVRFPKQDQNHQLFLLGKDKEQYKEGLQGQNVFVQELIDPNGRLSTVG
GVTKNNKTSETNTPLFVNKVNAGEDLDASIDSFLIQKEEISLKELDFKIRQQLVNNYGLY
KGTSKYGKIINLKDENKVEIDLGDKLQFERMGDVLNSKDIRGISVTINQI

```

► DataBase

Blast

Figure 3.12. BLAST interface.

3.2.4. Graph Theoretical Analysis of Targeted Human Proteins

Finally, the “Graph Analysis” interface provides the degree and betweenness centrality distributions of the targeted human proteins compared to all proteins within the human intra-network. The “Graph Analysis” feature of PHISTO allows users to plot degree and betweenness centrality distributions of human proteins targeted by the proteins of a selected pathogen type, family, species or strain in comparison to all interacting human proteins within the human PPI network.

The human PPI information stored in the PHISTO database are converted into a graph where the nodes denote proteins, and edges denote the interactions between proteins. The constructed graph of the human PPI network is used to calculate the degree values. The degree of a protein in a graph is equal to its number of connections in the graph. Betweenness centrality of a protein in a graph is defined as the number of shortest paths between any two nodes in the graph passing through that protein. Degree and betweenness centrality values are calculated for all human proteins, and the values with corresponding Uniprot IDs of proteins are stored as a table in the PHISTO database. These calculations are updated every time the interaction data of PHISTO is updated.

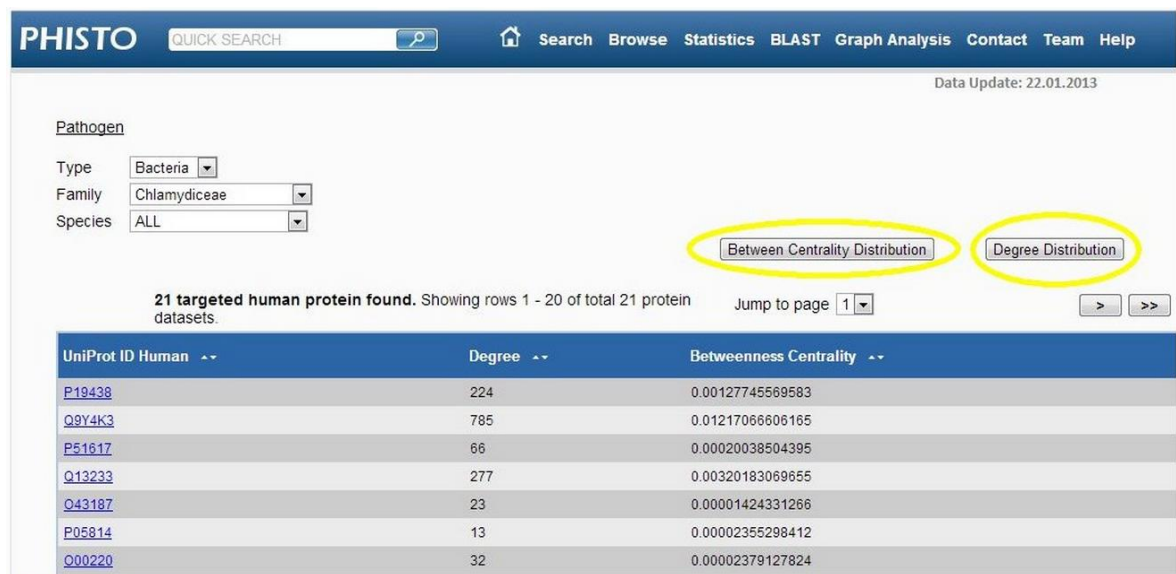


Figure 3.13. Degree and Betweenness Centrality values tabulated for the human proteins targeted by bacterial family Chlamydiae.

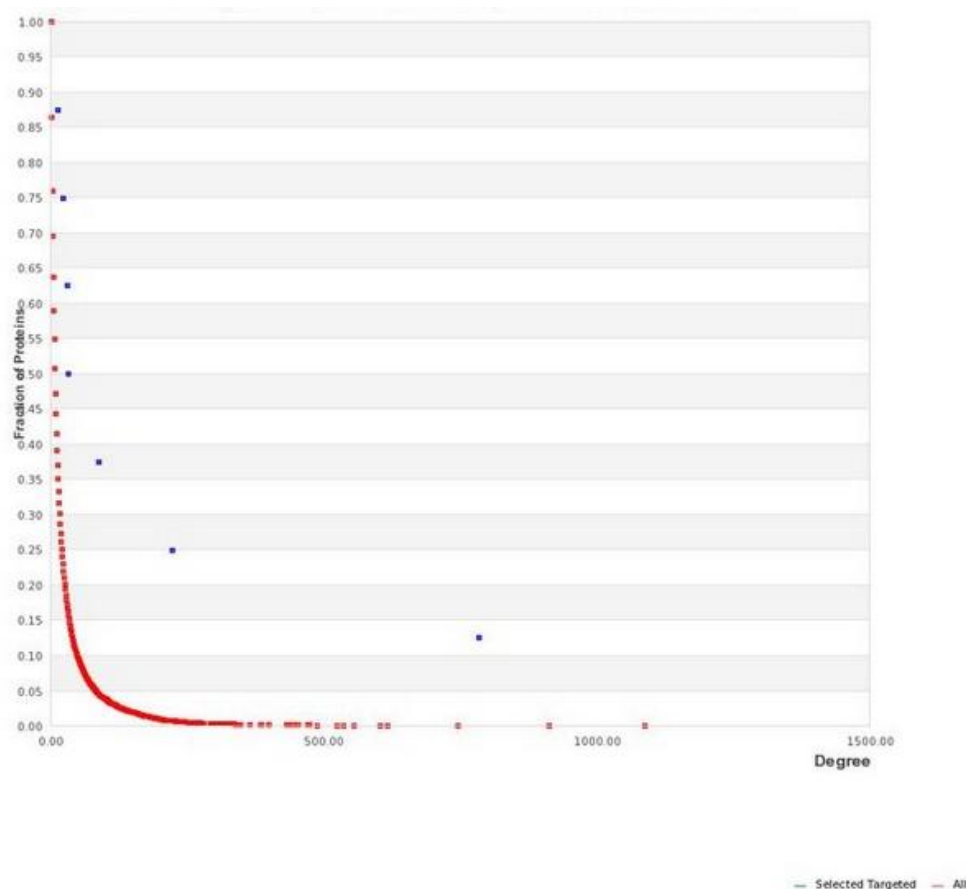


Figure 3.14. Degree distribution of the human proteins targeted by bacterial family Chlamydiae in comparison to all proteins in the human PPI network.

In “Graph Analysis” interface, the user browses the pathogen type, family, species or strain of interest from a drop-down menu first. The targeted human proteins with degree and centrality values are tabulated. The figure of degree and betweenness centrality distributions of the selected targeted human proteins in comparison to all, is displayed in a new window, by pressing "Degree Distribution" or "Betweenness Centrality Distribution" buttons.

3.2.5. Help Page

Users can find detailed explanations for the usage of all functionalities of PHISTO with examples in the "Help" page (Figure 3.15). Further technical details are available in the Supplementary Material of Durmuş Tekir *et al.* (2013b).

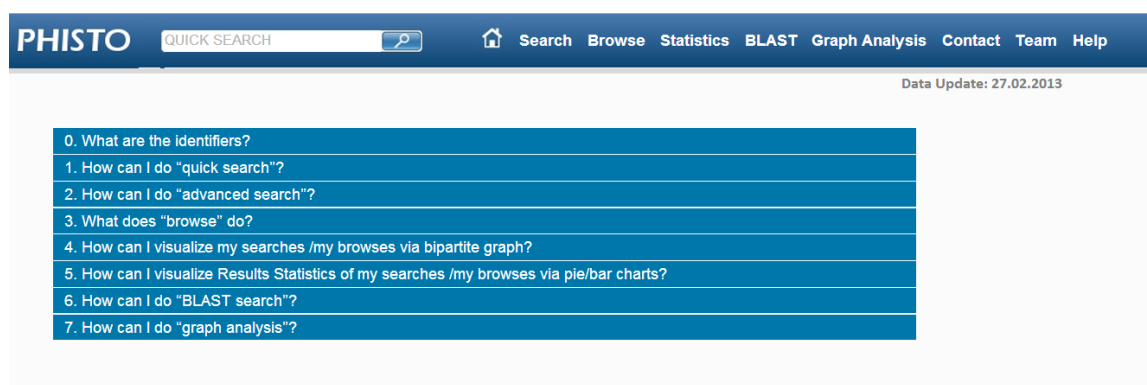


Figure 3.15. Help page.

3.3. Discussion

With regular monthly data updates and complete coverage of all data available for each pathogen type, PHISTO will always provide unified access to up-to-date PHI data. Otherwise, gathering such information would require careful and tedious data integration. PHISTO is the first PHI tool that uses text mining to identify the interaction detection methods of PHIs that have not been originally annotated with such information. Additionally, PHI results visualized as a bipartite graph allow users to capture pathogen-host mechanisms (i.e. most connected pathogen/host proteins can be easily observed at a glance). The sequence search option, BLAST, makes it possible to retrieve homologous PHIs. After homolog pathogen proteins in PHISTO are found, interacting human protein

partners of the homolog proteins may be assigned as putative interactors for the protein under investigation. Finally, the distributions of degree and betweenness centrality values in the graph analysis tool give crucial insights on attacking strategies of pathogens during infections, retrieving the properties of targeted human proteins (Dyer *et al.*, 2008; Durmuş Tekir *et al.*, 2012).

The goal with PHISTO is to provide a centralized and up-to-date platform for studying pathogen-host protein interaction systems with future curation of PHIs from literature by text mining and additional advanced analysis tools for PHI networks.

4. INFECTION STRATEGIES OF BACTERIAL AND VIRAL PATHOGENS THROUGH PATHOGEN-HUMAN PROTEIN-PROTEIN INTERACTIONS

Infection strategies have been studied through intraspecies protein interactions of various pathogens (Flajolet *et al.*, 2000; McCraith *et al.*, 2000; Rain *et al.*, 2001; Butland *et al.*, 2005; LaCount *et al.*, 2005; Uetz *et al.*, 2006; Arifuzzaman *et al.*, 2006; Calderwood *et al.*, 2007; von Brunn *et al.*, 2007; Parrish *et al.*, 2007; Titz *et al.*, 2008; Pan *et al.*, 2008; Kühner *et al.*, 2009; Wang *et al.*, 2010) as well as through interspecies protein interactions between pathogens and human (Filippova *et al.*, 2004; Mogensen *et al.*, 2006; Uetz *et al.*, 2006; Calderwood *et al.*, 2007; de Chassey *et al.*, 2008; Dyer *et al.*, 2008; Shapira *et al.*, 2009; Dyer *et al.*, 2010; Yang *et al.*, 2011; Jager *et al.*, 2012). Notwithstanding these, a general overview of infection mechanisms of different types of pathogens is missing since there has been a lack of interspecies interactome data until very recent years.

The systemic analysis of PHI data has so far focused mainly on virus-based infections due to the scarcity of data for other types of pathogenic organisms (Uetz *et al.*, 2006; Calderwood *et al.*, 2007; Dyer *et al.*, 2008; Shapira *et al.*, 2009). Large-scale bacterial PHIs (Dyer *et al.*, 2010) to get statistically meaningful results are available, providing a good opportunity to get a systemic picture of the pathogenesis of bacterial infections. Here, up-to-date PHI data reported in PHISTO with a specific focus on comparison between bacterial and viral infections of human were studied, constituting the first comprehensive analysis of PHI data together for both viral and bacterial pathogens (Durmuş Tekir *et al.*, 2012).

Various sets of human proteins targeted by bacteria, fungi, protozoa, and viruses were constructed to pick out specific infection strategies of different pathogen types. On the other hand, the set of human proteins targeted by both bacteria and viruses was used to obtain common infection strategies of these pathogens. Degree (connectivity) and betweenness centrality distributions of the human protein sets targeted by bacteria and viruses were computed to observe the network properties of targeted human proteins.

Additionally, gene ontology (GO; Ashburner *et al.*, 2000) terms were enriched in each of the above-mentioned protein set to find out attacked mechanisms in human. GO enrichment analysis was also performed for sets of human proteins targeted by each pathogen group included in the PHI data to decipher the pathogen group(s) manipulating specific processes in human.

4.1. Protein Interaction Data

4.1.1. PHI Data

The pathogen-human PHIs were downloaded from PHISTO in October 2011. The data cover 23,435 physical interactions occurring between 5210 human proteins and 3419 proteins of 257 pathogen strains of 72 pathogen groups including 24 bacterial (Table 4.1), 3 fungal (Table 4.2), 2 protozoan (Table 4.3), and 43 viral (Table 4.4) groups. Pathogen strains were grouped to present PHI results together for related strains. Bacterial groups are sets of strains of the same genus as the names of the groups are the names of the genres. For viral groups, there are two definitions. Some viral groups are sets of strains of the same family as the names of the groups are the names coming from the families of the strains (e.g. Papillomaviruses, Herpesviruses, Polyomaviruses). Some viral strains are grouped based on the related infections caused by them, as the names are generally coming from the diseases (e.g. HIV, Hepatitis viruses, Anemia viruses).

4.1.2. Human PPI Data

To obtain degree and betweenness centrality (BC) values of pathogen-targeted proteins, the human PPI network was constructed by integrating 194,006 interactions between 13,015 human proteins from BioGrid (Stark *et al.*, 2011), DIP (Salwinski *et al.*, 2004), IntAct (Kerrien *et al.*, 2012), Mint (Ceol *et al.*, 2010), and Reactome (Croft *et al.*, 2011) in April 2011.

Table 4.1. Contents of bacteria-human PHI data.

Bacterial group	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
<i>Aeromonas</i>	1	2	1	2
<i>Bacillus</i>	2	3021	940	1736
<i>Campylobacter</i>	1	3	1	3
<i>Chlamydia</i>	2	21	3	21
<i>Citrobacter</i>	1	1	1	1
<i>Clostridium</i>	3	47	9	10
<i>Corynebacterium</i>	1	1	1	1
<i>Escherichia</i>	4	30	14	27
<i>Finegoldia</i>	1	1	1	1
<i>Francisella</i>	1	1338	346	986
<i>Helicobacter</i>	2	3	3	2
<i>Klebsiella</i>	1	1	1	1
<i>Legionella</i>	1	1	1	1
<i>Listeria</i>	1	4	4	3
<i>Moraxella</i>	1	1	1	1
<i>Mycoplasma</i>	1	2	1	2
<i>Neisseria</i>	1	17	1	17
<i>Pseudomonas</i>	1	12	4	10
<i>Salmonella</i>	1	5	5	5
<i>Shigella</i>	1	11	9	8
<i>Staphylococcus</i>	3	12	10	10
<i>Streptococcus</i>	5	15	12	9
<i>Vibrio</i>	1	1	1	1
<i>Yersinia</i>	4	3999	1221	2120
TOTAL	41	8549	2591	3589

Table 4.2. Contents of fungi-human PHI data.

Fungal group	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
<i>Candida</i>	1	1	1	1
<i>Pneumocystis</i>	1	1	1	1
<i>Radiomyces</i>	1	2	1	2

Table 4.3. Contents of protozoa-human PHI data.

Protozoan group	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
<i>Plasmodium</i>	3	8	4	8
<i>Toxoplasma</i>	1	1	1	1

Table 4.4. Contents of virus-human PHI data.

Viral group	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
Adenovirus	13	121	36	80
Anemia virus	6	8	6	4
ASFV	1	1	1	1
Bacteriophage	6	6	6	5
Coxsackie virus	1	1	1	1
Dengue virus	3	3	3	2
Ebola virus	1	1	1	1
Echo virus	2	3	3	1
Ectromelia virus	1	2	2	2
Encephalitis virus	1	2	1	2
Foamy virus	1	1	1	1
Hantaan virus	1	6	1	6
Hendra virus	1	1	1	1
Hepatitis virus	21	1573	179	399
Herpesvirus	28	666	141	388
HIV	49	11435	279	1601
Influenza virus	9	523	27	182
Leukemia virus	3	10	3	10
Measles virus	3	10	4	4
Molluscum virus	1	1	1	1
MPMV	1	1	1	1
Nipah virus	1	1	1	1
Nucleopolyhedrovirus	1	1	1	1
Orf virus	2	2	2	1
Papillomavirus	14	290	51	128
Parainfluenza virus	1	2	1	2
Parvo virus	1	1	1	1
Polio virus	2	3	2	2
Polyoma virus	4	64	10	45
Puumala virus	1	3	1	3
Rabies virus	1	1	1	1
Rhino virus	1	1	1	1
Rota virus	4	8	5	6
Sarcoma virus	5	15	6	11
SARS	1	4	3	4
Sendai virus	1	1	1	1
Seoul virus	1	4	1	4
SIV	2	3	2	3
Stomatitis virus	3	7	3	6
T-lymphotrophic virus	3	38	7	35
Tula virus	1	2	1	2
Vaccinia virus	5	46	20	33
West Nile virus	1	1	1	1
TOTAL	209	14873	820	2398

4.1.3. Human Protein Sets

A total of 10 sets of human proteins interacting with pathogens were constructed from PHI data to analyze the properties of targeted human proteins as follows: The sets targeted by bacterial pathogens (bacteria-targeted set), fungal pathogens (fungi-targeted set), protozoan pathogens (protozoa-targeted set), and viral pathogens (virus-targeted set) were analyzed for specific infection strategies of these different pathogen types.

For a deeper comparison between bacterial and viral infections, human proteins interacting with at least two bacterial groups (two bacteria-targeted set) and two viral groups (two viruses-targeted set) and also human proteins interacting with at least three bacterial groups (three bacteria-targeted set) and three viral groups (three viruses-targeted set) were used. To obtain common infection strategies of pathogens, sets of human proteins targeted by all types of pathogens (pathogen-targeted set) and by both bacteria and viruses (bacteria-virus-targeted set) were also analyzed.

4.2. Computational Methods

4.2.1. Degree and Betweenness Centrality Calculations

Degree of a protein within a network is defined as its number of connections. Betweenness centrality of a protein is equal to the number of shortest paths between any pairs passing through that protein. The degree and centrality values of proteins in interaction networks provide valuable information about the role of corresponding proteins in the network's functional organization using the topology of the interconnections. For instance, hubs (highly connected proteins) and bottlenecks (central proteins to many paths in the network) are critical players in the intraspecies protein networks for information flow (Barabasi and Oltvai, 2004; Yu *et al.*, 2007).

The undirected human PPI network was represented as an adjacency matrix, and the degree and centrality values of each protein in the network were calculated in MATLAB environment. Betweenness centrality calculations were performed by freely available MATLAB BGL package developed by David Gleich (<https://launchpad.net/matlab-bgl>).

The results were normalized by $(n-1)(n-2)$, where n is the number of all proteins in the PPI network. Self-interactions were not taken into account in these calculations.

4.2.2. GO Enrichment Analysis

GO (Ashburner *et al.*, 2000) enrichments of all human protein sets were performed using BiNGO plugin (ver. 2.44) of Cytoscape (ver. 2.8.1) (Maere *et al.*, 2005). Significance level was set to 0.05 meaning that only terms enriched with a p-value of at most 0.05 were considered.

4.3. Results

4.3.1. Pathogen-Targeted Human Proteins

The distribution of 5210 human proteins on their targeting pathogens are shown in the Venn diagram (Figure 4.1).

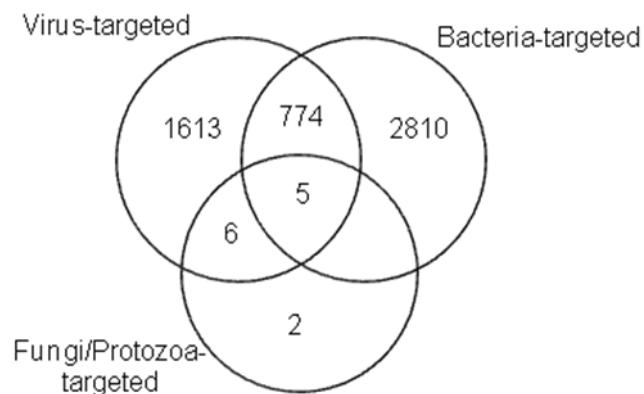


Figure 4.1. The number of human proteins that are grouped based on their interactions with viruses, bacteria and fungi/protozoa (targeted by fungi and/or protozoa).

The most targeted human proteins are listed in Table 4.5. The top of this list, P53 (Tumor suppressor), DRA (HLA class II histocompatibility antigen, DR alpha chain), SUMO1 (Small ubiquitin-related modifier 1), JUN (Transcription factor AP-1), NPM (Nucleophosmin), ROA1 (Heterogeneous nuclear ribonucleoprotein A1), and UBC9 (SUMO-conjugating enzyme) and the following proteins have potential to give important insights about infections.

Table 4.5. Highly targeted human proteins.

Protein	Degree	BC	Targeting bacterial groups	Targeting viral groups
P53	347	0.01547	<i>Bacillus, Escherichia, Francisella, Yersinia</i>	Adenovirus, Hepatitis virus, HIV, Papillomavirus, Polyomavirus, SIV
DRA	52	0.00003	<i>Bacillus, Francisella, Mycoplasma, Staphylococcus, Yersinia</i>	Herpesvirus, HIV, Influenza virus
SUMO1	103	0.00366	<i>Bacillus</i>	Herpesvirus, HIV, Papillomavirus, Puumala virus, SARS, Tula virus, Vaccinia virus
JUN	122	0.00335	<i>Bacillus, Francisella, Yersinia</i>	Hepatitis virus, HIV, Papillomavirus, Vaccinia virus
NPM	137	0.00166	<i>Bacillus, Francisella, Yersinia</i>	Adenovirus, Hepatitis virus, Herpesvirus, HIV
ROA1	246	0.00262	<i>Bacillus, Francisella, Yersinia</i>	Herpesvirus, HIV, Influenza virus, SARS
UBC9	134	0.00410	<i>Yersinia</i>	Hantaan virus, Herpesvirus, HIV, Influenza virus, Papillomavirus, Seoul virus
IGHG1	57	0.00219	<i>Bacillus, Francisella, Staphylococcus, Streptococcus, Yersinia</i>	Herpesvirus
RAC1	239	0.00279	<i>Bacillus, Clostridium, Pseudomonas, Salmonella, Yersinia</i>	HIV
CDC42	232	0.00405	<i>Bacillus, Francisella, Salmonella, Yersinia</i>	HIV, T-lymphotropic virus
DRB5	-	-	<i>Bacillus, Francisella, Streptococcus, Yersinia</i>	Herpesvirus, HIV
LCK	147	0.00202	<i>Bacillus, Francisella, Yersinia</i>	Hepatitis virus, Herpesvirus, HIV
XRCC6	131	0.00445	<i>Bacillus, Francisella, Yersinia</i>	Herpesvirus, HIV, Polyomavirus

Table 4.5. Highly targeted human proteins (cont.)

Protein	Degree	BC	Targeting bacterial groups	Targeting viral groups
KPYM	76	0.00041	<i>Bacillus, Francisella, Yersinia</i>	Hepatitis virus, Herpesvirus, Papillomavirus
ROA2	189	0.00069	<i>Bacillus, Francisella, Yersinia</i>	Herpesvirus, Influenza virus, Vaccinia virus
P85A	402	0.00914	<i>Bacillus, Francisella, Yersinia</i>	Anemia virus, HIV, Influenza virus
STAT3	77	0.00133	<i>Bacillus, Francisella, Yersinia</i>	Hepatitis virus, Herpesvirus, HIV
STAT1	71	0.00104	<i>Bacillus, Francisella, Yersinia</i>	Adenovirus, Herpesvirus, HIV
GBLP	93	0.00265	<i>Bacillus, Francisella, Yersinia</i>	Adenovirus, Herpesvirus, HIV
PARP4	1	0.00000	<i>Bacillus, Francisella, Yersinia</i>	Hepatitis virus, Herpesvirus, HIV
RB	149	0.00282	<i>Yersinia</i>	Adenovirus, Herpesvirus, HIV, Papillomavirus, Polyomavirus
SP1	103	0.00268	<i>Yersinia</i>	Adenovirus, Herpesvirus, HIV, Polyomavirus, T-lymphotropic virus
TAF1	58	0.00025	<i>Bacillus</i>	Adenovirus, Hepatitis virus, HIV, Papillomavirus, Polyomavirus
CDK2	151	0.00220	<i>Shigella</i>	Herpesvirus, HIV, Papillomavirus, Polyomavirus, T-lymphotropic virus
TF2B	69	0.00020	<i>Bacillus</i>	Hepatitis virus, Herpesvirus, HIV, Papillomavirus, Polyomaviru
EP300	123	0.00245	<i>Bacillus</i>	Adenovirus, Hepatitis virus, HIV, Papillomavirus, Polyomavirus
CBP	147	0.00304	<i>Yersinia</i>	Adenovirus, Hepatitis virus, HIV, Papillomavirus, Polyomavirus
TBP	147	0.00241	-	Adenovirus, Hepatitis virus, Herpesvirus, HIV, Papillomavirus, Polyomavirus

4.3.2. Degree and Betweenness Centrality Distributions

Figures 4.2 and 4.3 display the comparison between the degree distributions of non-targeted proteins in the human PPI network and bacteria- and virus-targeted sets. For both cases of bacteria and virus targeted sets, it is observed that pathogen-targeted human proteins have generally higher degree values than non-targeted ones. However a difference is observed in trends of degree distributions of multibacteria and multiviruses targeted sets. For bacteria-targeted cases (Figure 4.2), the increase in degree values of human proteins with increasing number of targeting pathogen groups is not as clear as those of virus-targeted cases (Figure 4.3). Very similar trends are obtained for centrality distributions of human proteins (Figure 4.4 and 4.5). In order to justify these global trends, the same analyses was then repeated with human protein sets excluding the overrepresented pathogens, i.e. *Bacillus*, *Yersinia* and HIV which target the largest number of human proteins (Tables 4.1 and 4.4). Similar results are still obtained when major pathogen groups are eliminated (Figure 4.6 and 4.7).

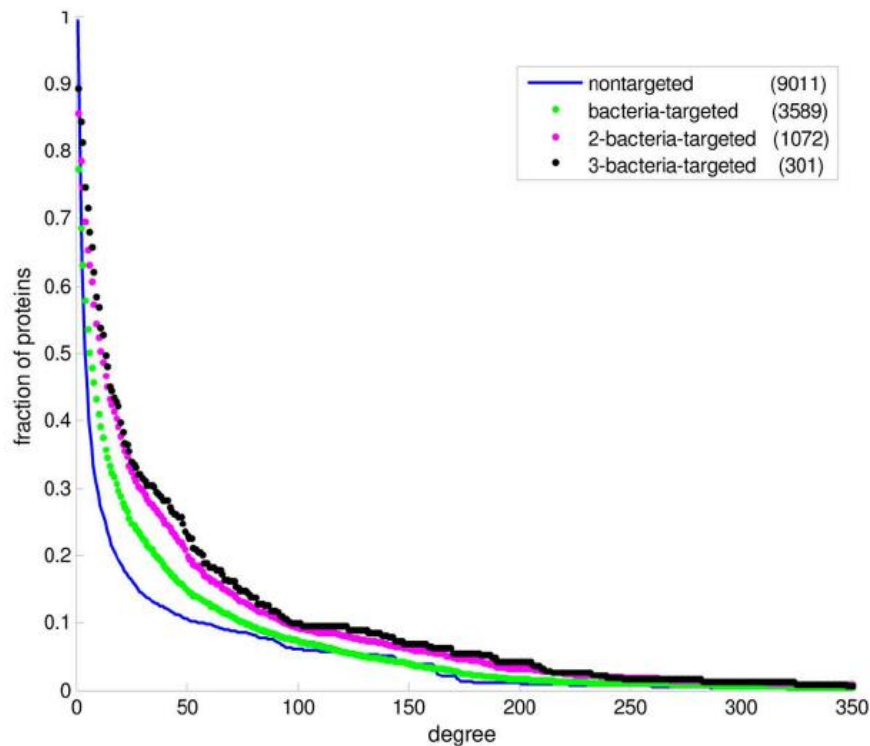


Figure 4.2. The cumulative degree distributions of bacteria-targeted human proteins.

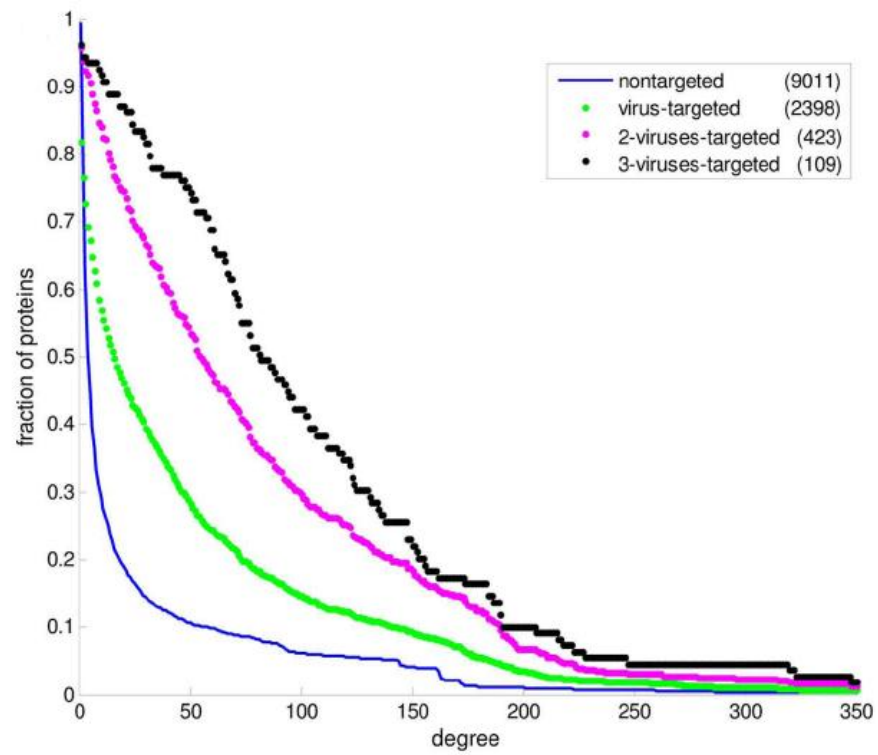


Figure 4.3. The cumulative degree distributions of virus-targeted human proteins.

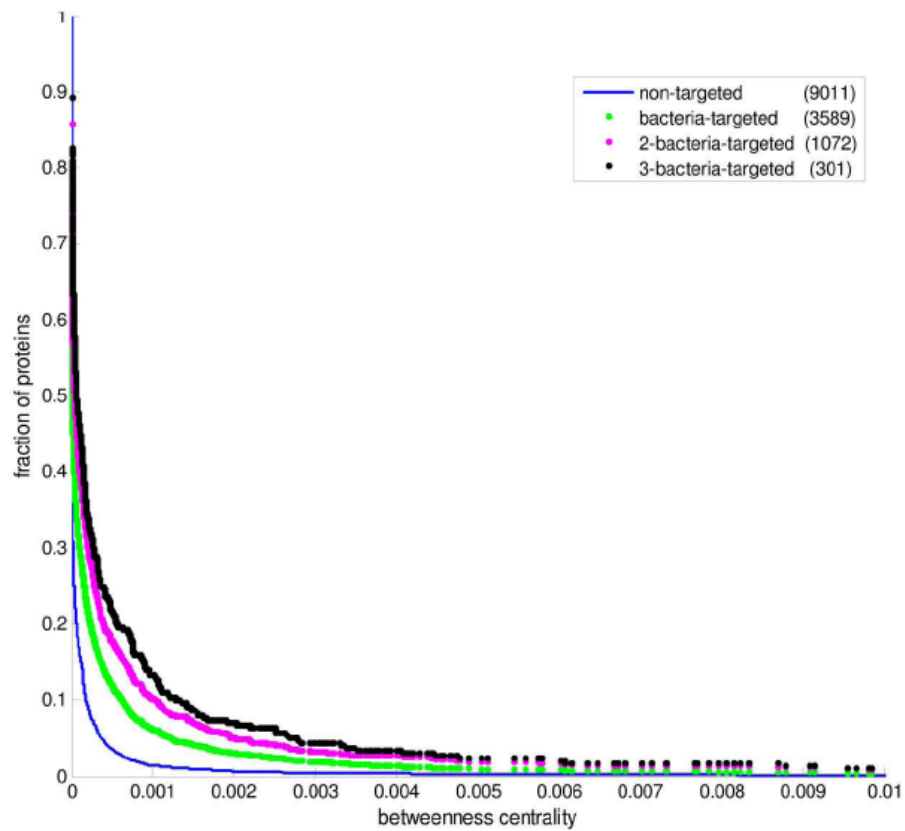


Figure 4.4. The cumulative BC distributions of bacteria-targeted human proteins.

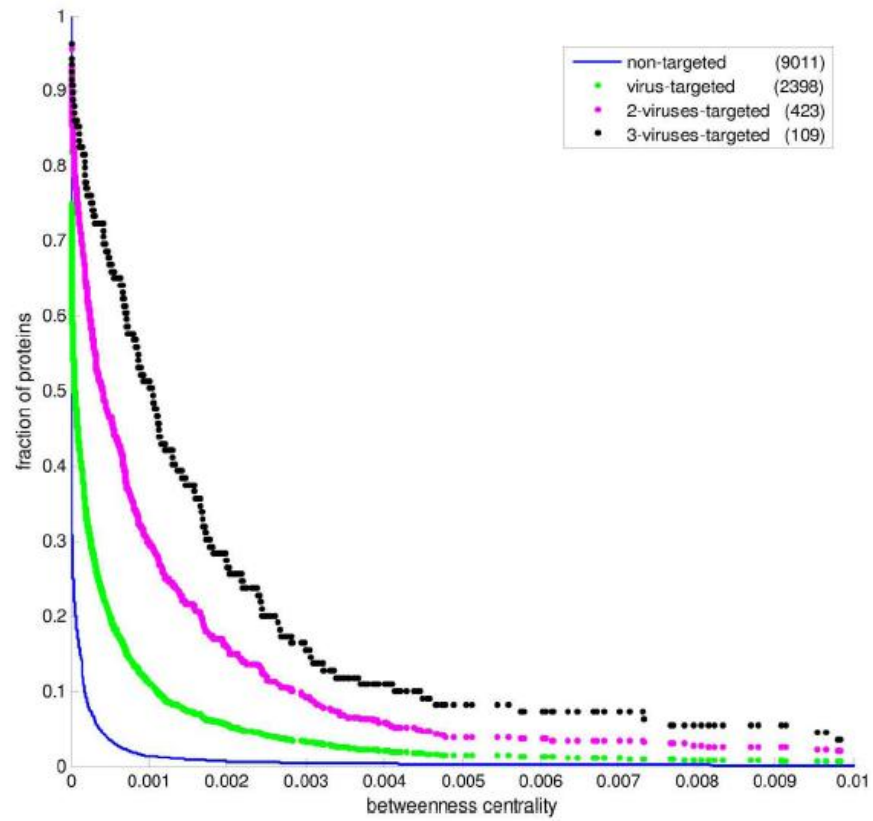


Figure 4.5. The cumulative BC distributions of virus-targeted human proteins.

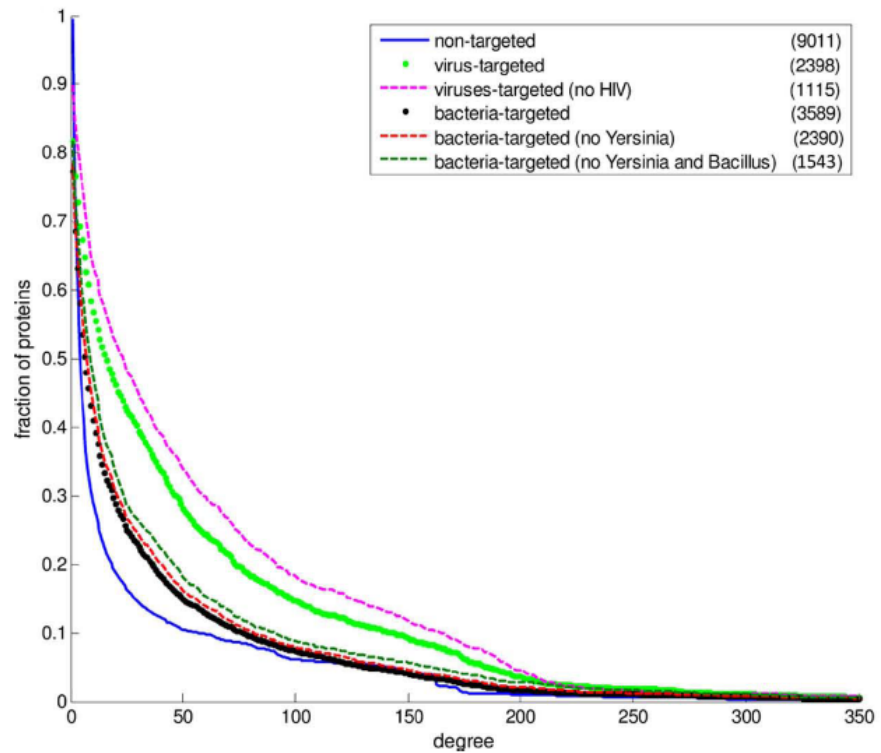


Figure 4.6. The cumulative degree distributions of targeted human proteins excluding Yersinia, Bacillus and HIV data.

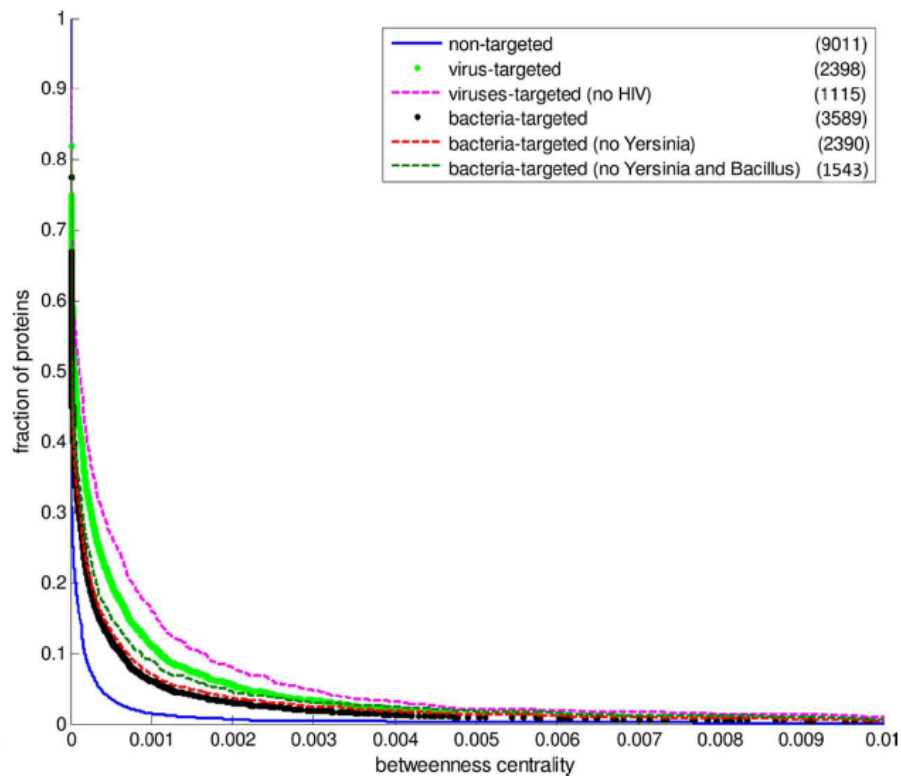


Figure 4.7. The cumulative BC distributions of targeted human proteins excluding Yersinia, Bacillus and HIV data.

4.3.3. GO Enrichment Analysis

Special attention should be paid to the results of sets of human proteins interacting with three and more bacterial groups and three and more viral groups for a comparison between their infection strategies. The human proteins targeted by more pathogen groups reflect more specificity to infection mechanism of the corresponding pathogen (bacteria or virus). The enriched GO terms in human proteins interacting with both bacterial and viral pathogens are also important to highlight common infection mechanisms. The first 20 enriched GO process terms for three bacteria-targeted-set, three viruses-targeted-set, and bacteria-virus-targeted set are listed in Tables 4.6-4.8 to point out the human processes that are attacked by pathogens.

Table 4.6. First 20 enriched GO Process terms in human proteins targeted by at least three bacterial groups (three bacteria-targeted set).

GO Process Term	p-value
I-kappaB kinase/NF-kappaB cascade	9.64E-13
regulation of biological process	9.69E-10
biological regulation	2.59E-09
negative regulation of biological process	4.89E-09
positive regulation by organism of immune response of other organism involved in symbiotic interaction	6.64E-09
modulation by organism of immune response of other organism involved in symbiotic interaction	6.64E-09
modulation by symbiont of host immune response	6.64E-09
positive regulation by symbiont of host immune response	6.64E-09
response to immune response of other organism involved in symbiotic interaction	6.64E-09
response to host immune response	6.64E-09
positive regulation by organism of defense response of other organism involved in symbiotic interaction	6.64E-09
positive regulation by symbiont of host defense response	6.64E-09
positive regulation by organism of innate immunity in other organism involved in symbiotic interaction	6.64E-09
modulation by organism of innate immunity in other organism involved in symbiotic interaction	6.64E-09
pathogen-associated molecular pattern dependent modulation by organism of innate immunity in other organism involved in symbiotic interaction	6.64E-09
modulation by organism of defense response of other organism involved in symbiotic interaction	6.64E-09
pathogen-associated molecular pattern dependent induction by organism of innate immunity of other organism involved in symbiotic interaction	6.64E-09
modulation by symbiont of host defense response	6.64E-09
pathogen-associated molecular pattern dependent induction by symbiont of host innate immunity	6.64E-09
modulation by symbiont of host innate immunity	6.64E-09

Table 4.7. First 20 enriched GO Process terms in human proteins targeted by at least three viral groups (three viruses-targeted set).

GO Process Term	p-value
interspecies interaction between organisms	1.89E-40
multi-organism process	1.19E-27
positive regulation of cellular process	1.12E-17
positive regulation of biological process	1.06E-16
cellular macromolecule metabolic process	1.12E-15

Table 4.7. First 20 enriched GO Process terms in human proteins targeted by at least three viral groups (cont.)

GO Process Term	p-value
nucleic acid metabolic process	4.49E-14
positive regulation of macromolecule metabolic process	4.60E-14
cell cycle process	6.72E-14
positive regulation of gene expression	1.49E-13
cell cycle	2.06E-13
positive regulation of metabolic process	3.79E-13
positive regulation of transcription	4.37E-13
macromolecule metabolic process	8.51E-13
positive regulation of cellular metabolic process	3.89E-12
cellular response to stimulus	6.61E-12
positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	7.26E-12
positive regulation of macromolecule biosynthetic process	1.32E-11
positive regulation of nitrogen compound metabolic process	1.41E-11
positive regulation of transcription, DNA-dependent	1.44E-11
regulation of cell cycle	1.47E-11

Table 4.8. First 20 enriched GO Process terms in human proteins targeted by both bacterial and viral groups (bacteria-virus-targeted set).

GO Process Term	p-value
interspecies interaction between organisms	1.64E-52
multi-organism process	1.01E-47
positive regulation of biological process	5.62E-47
regulation of biological process	1.32E-42
biological regulation	2.66E-40
positive regulation of cellular process	4.59E-40
negative regulation of biological process	6.32E-37
regulation of cellular process	2.00E-36
negative regulation of cellular process	4.84E-32
regulation of protein metabolic process	7.68E-30
regulation of macromolecule metabolic process	3.21E-29
regulation of cellular protein metabolic process	1.39E-28
regulation of cell death	1.74E-28
positive regulation of macromolecule metabolic process	1.91E-28
positive regulation of cellular metabolic process	7.04E-28
regulation of programmed cell death	1.13E-27
cellular macromolecule metabolic process	1.94E-27
positive regulation of metabolic process	2.92E-27
negative regulation of macromolecule metabolic process	5.31E-27
regulation of apoptosis	6.54E-27

4.4. Discussion

In this study, the aim was to provide a general overview of infection strategies used by different pathogens based on the comprehensive PHI data in PHISTO. Although large-scale pathogen-human protein interaction data have been identified in the last few years, the data for fungal and protozoan systems are still scarce (Tables 4.1-4.4) to extract significant conclusions about their infection mechanisms. On the other hand, interspecies protein interaction networks for bacterial and viral pathogens with human have been identified, enough to provide some insights about their strategies to subvert human cellular processes during infection.

With increasing PHI data of bacterial and viral pathogens, studies have been performed to enlighten specific bacteria-human (Mogensen *et al.*, 2006; Dyer *et al.*, 2010) and virus-human (Filippova *et al.*, 2004; Uetz *et al.*, 2006; Calderwood *et al.*, 2007; König *et al.*, 2008) interaction systems. Although some studies provided global views of infection strategies of viruses (Dyer *et al.*, 2008) and bacteria (Dyer *et al.*, 2010) separately, they do not provide a direct comparison between bacterial and viral infections. In fact, only <2% of the PHI data of Dyer *et al.* (2008) are for bacteria-human interactions whereas it is more than 36% in the database of PHISTO. Hence, the present study constitutes the first extensive comparison between bacteria-human and virus-human interspecies protein interaction networks to retrieve information about infection strategies specific to each system and then common to both systems. The findings should be interpreted with caution since the protein interaction networks between pathogens and human are not complete yet.

4.4.1. Special Infection Strategies

In recent studies it has been suggested that viral proteins (Calderwood *et al.*, 2007; Dyer *et al.*, 2008; Itzhaki, 2011) and bacterial proteins (Dyer *et al.*, 2010) have evolved to preferentially interact with hubs and bottlenecks in the human PPI network. The degree and betweenness centrality distributions of the bacteria-targeted and virus-targeted human protein sets are displayed in comparison with non-targeted proteins in the human PPI network in Figures 4.2-4.5. It is observed that the degree and centrality values of human proteins increase with increasing number of targeting bacterial and viral groups,

confirming the previous results with the most comprehensive PHI data. A novel finding by the comparative analysis is that the increase in degree and centrality values with increasing number of pathogen groups is much more pronounced in virus-targeted cases (Figures 4.3 and 4.5) than bacteria-targeted cases (Figures 4.2 and 4.4). Therefore it can be concluded that attacking to hub and bottleneck proteins in the human interaction network is more specific to viral infections.

These findings certainly make sense. During infection, relatively few pathogen-derived proteins face a huge number of host proteins. In many cases, pathogens have significantly smaller coding capacities than their respective host species, for instance, numerous viruses express less than 15 different proteins and several bacteria translocate a limited number of effector proteins into the host cell. Pathogens seem to overcome the discrepancy between the restricted number of own proteins and the huge host proteome by targeting central elements of the human PPI network. A comparison of bacteria and viruses revealed that this trend is even more pronounced in the case of viruses. Potential explanations might reside in the obligatory intracellular replication or the smaller genome size of viruses (Schleker and Trilling, 2013).

In the PHI data, some pathogen groups are overrepresented with their larger number of reported interactions with human (Tables 4.1 and 4.4). As most of these large-scale data have been produced with high-throughput detection methods, which are prone to experimental biases and errors, it was necessary to check whether the distributions of degree and centrality values of the pathogen-targeted human proteins would be same without the groups with large number of interacting partners of human proteins (i.e. *Bacillus*, *Yersinia* and HIV). Hence, the above-mentioned analyses were performed again with human protein sets excluding these major pathogen groups. 1199 human proteins targeted by only *Yersinia* strains were excluded from the bacteria-targeted set, and 1283 human proteins targeted by only HIV strains were excluded from the virus-targeted set to obtain the degree and betweenness centrality values of the remaining human proteins. The human proteins targeted by bacteria other than *Bacillus* and *Yersinia* were also analyzed to exclude the effect of large-scale data of the two groups. 1199 only-*Yersinia*-targeted and 847 only-*Bacillus*-targeted human proteins were excluded from the bacteria-targeted set. The behaviour of the remaining human proteins can be observed in Figures 4.6 and 4.7

resulting in similar trends with the global case. Additionally, a direct comparison of the degree and centrality between bacteria and virus targeted interaction partners with respect to non-targeted human proteins is also given in Figures 4.6 and 4.7. The difference in the behaviour of the bacteria- and virus-targeted sets are clear especially in degree distributions (Figure 4.6). The degree values of bacteria-targeted human proteins with or without *Bacillus* and *Yersinia* are nearly same. On the other hand, attack of viruses to more connected human proteins is more clear when HIV data are excluded.

From the enriched GO process terms in human proteins targeted by at least 3 bacterial groups (Table 4.6), it can be concluded that bacteria may have adapted to attack proteins involved generally in human immunity pathways. Therefore, the most specific bacterial infection strategy is through evading or suppressing human immune responses as also concluded previously (Lai *et al.*, 2001; Park *et al.*, 2002; Zhang *et al.*, 2005; Dyer *et al.*, 2010). The human immune system is manipulated by bacterial proteins attacking human proteins functioning in innate and adaptive immunity (i.e. TLR4 and TLR7), inflammation (i.e. NF- κ B and BCL6), activation of T cells (i.e. CXCR4 and LCK) (Zhang and Ghosh, 2000; Alonso *et al.*, 2004; Oda and Kitano, 2006; Dyer *et al.*, 2010). In the PHI data it is observed that *Yersinia* bacteria attack all these human defence mechanisms targeting all mentioned human proteins. Proteins of *Bacillus* and *Francisella* interact with NF- κ B and LCK (Dyer *et al.*, 2010) aiming to disrupt the mechanisms of inflammation and T cell responses. On the other hand, proteins of *Chlamydia*, *Escherichia*, and *Neisseria* interact with crucial players of innate and adaptive immunity, toll-like receptors (TLR4 and TLR7) (Croft *et al.*, 2011) to collapse the human immune system. There are several other bacteria-targeted human proteins involved in the immune system. Their interactions with bacterial proteins should be investigated carefully for a complete understanding of bacterial strategies targeting human defense mechanism during infection.

Viruses attack human cellular processes (Table 4.7) enabling themselves to proliferate in human during infection. All viruses use this mechanism since they need host's transcriptional machinery for viral genetic material transcription. Even the human proteins targeted by only one viral group are enriched in GO process terms relevant to regulation of cellular mechanisms. Viruses manipulate human cellular mechanisms by interacting with various proteins functioning in cell cycle (i.e. DLG1, PTMA, and EP300),

with human transcription factors to promote their own genetic material transcription (i.e. E2F1 and TAF1), with key proteins controlling apoptosis (i.e. P53), and with nuclear membrane proteins for transporting their genetic material across the nuclear membrane (i.e. RAN, and SUMO1) (Lechner and Laimins, 1994; Thompson *et al.*, 1997; Carrillo *et al.*, 2004; Thomas *et al.*, 2005; Dyer *et al.*, 2008). In the PHI data, Adenoviruses, HIV, Papillomaviruses, and Polyomaviruses are observed to target one or more proteins in each of four groups; cell cycle proteins, transcription factors, apoptosis regulator, and nuclear membrane proteins. Proteins of Hepatitis viruses interact with PTMA, EP300, TAF1, and p53 while proteins of Herpesviruses interact with PTMA and SUMO1. On the other hand, viral groups of Influenza, Puumala, Tula, SARS, and Vaccinia are observed to target nuclear membrane proteins. The other virus-targeted human proteins involved in cellular mechanisms should be investigated comprehensively for a complete understanding of viral strategies targeting human cellular mechanism.

It can be concluded that the main infection strategies of bacteria and viruses are through attacking human immune system and cellular processes, respectively. However, there are some exceptions such that some bacterial groups target human proteins functioning in cellular mechanisms whereas some viral groups target human proteins functioning in defense mechanisms. In the case of bacteria, the difference might arise from the life-style, e.g., intracellular bacteria like Chlamydia, Listeria and Mycoplasma are able to grow and reproduce only within the host cells just like viruses (Kaufmann, 1993). Therefore, human protein sets targeted by these intracellular bacterial groups are enriched in GO process terms related to the cellular mechanisms (e.g., regulation of cellular processes, regulation of transcription) in addition to the immune system. On the other hand, viruses like herpes and pox (ectromelia, molluscum, orf, vaccinia) viruses as well as HIV have the ability to evade human immune system (Alcami and Koszinowski, 2000) as observed in the results.

For more specific infection strategies of pathogen groups, the results of GO enrichment analysis for the human protein sets targeted by each of the 72 groups in the PHI data can be used. Additionally, intranetworks of pathogenic proteins in each pathogen group should be analyzed for drug target identification after a thorough understanding of pathogenesis via interspecies protein interactions.

4.4.2. Common Infection Strategies

In spite of the difference in the trends of distributions of degree and centrality values of human proteins in bacteria-targeted and virus-targeted sets, the tendency to attack human proteins that are highly connected (hubs) and central to shortest paths (bottlenecks) is common to all types of pathogens. It is observed in the PHI data that the degree and centrality values of pathogen-targeted human proteins are generally greater than non-targeted ones (Figures 4.2-4.7). This infection strategy of pathogens, attacking more connected and central nodes in the human PPI network, is probably due to enabling themselves to control and disrupt essential complexes and pathways more easily. With scale-free nature, the human PPI network is robust to attacks on random nodes. However, the selective attacks to even a small number of nodes of high degree can dramatically change the topology and functionality of the network (Albert *et al.*, 2000; Li *et al.*, 2006).

Although bacteria and viruses have a tendency to interact with different human proteins (Figure 4.1), they together target those (779 human proteins) enriched highly in the regulation of metabolic processes in addition to cellular processes (Table 4.8). Both pathogen types target human proteins involved in regulation of metabolic processes in order to use human metabolism for their own advantage. Destroying the regulation of metabolic processes may enable pathogenic microorganisms easily make use of human metabolism as well as nutrition and energy resources. Furthermore, unlike bacteria viruses do not have own metabolic pathways depending largely on host's metabolism. Therefore metabolic processes-related GO terms enriched in viruses-targeted proteins can be observed in Table 4.7 whereas there are no such terms enriched in bacteria-targeted proteins in Table 4.6, since bacteria do not need human metabolism as much as in the viral case. However, the common human targets of bacteria and viruses are observed to have roles in regulation of human metabolism. For instance, a pyruvate kinase isozyme, KP YM, functions in glycolysis catalyzing the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) to ADP, generating ATP. This metabolic human protein is targeted by three bacterial (*Bacillus*, *Francisella*, and *Yersinia*) and three viral groups (Hepatitis, Herpesviruses, and Papillomaviruses) in the PHI data. Alpha-enolase is another bacteria-virus-targeted enzyme which functions in glycolysis, just before KP YM enzyme, converting 2-phospho-glycerate to PEP. A metabolic step operating again around lower

glycolysis is the production of lactate from pyruvate. Both isoenzymes (LDHA, LDHB) are found to be a target for bacterial and viral groups. In addition to lower glycolysis, some enzymes functioning in lipid metabolism (ACSA, ACOT9, CPT1A) were identified as common targets of bacteria and viruses. Interestingly, two enzymes functioning for protection against oxidative-stress are in our common-target list: catalase (CATA) and glutathione peroxidase 3 (GPX3). These enzymes remove H_2O_2 , which is a reactive oxygen species (ROS) harmful for the cell.

The human proteins targeted by both bacteria and viruses have not been investigated in any previous study. Through the analyses using large scale PHI data it can be concluded that both bacteria and viruses attack to the proteins functioning in human metabolic processes as a common infection strategy. All bacteria-viruses-targeted human proteins involved in metabolic processes should be investigated carefully for a complete picture of commonalities in bacterial and viral infections.

5. COMPARATIVE INTERACTOMICS FOR VIRAL INFECTION

STRATEGIES: DNA VIRUSES VERSUS RNA VIRUSES

Viral families are grouped based on their type of genetic material, DNA or RNA (Baltimore, 1971). DNA viruses have DNA as their genetic material and replicate using DNA-dependent DNA polymerase. This nucleic acid is usually double-stranded (dsDNA) but may also be single-stranded (ssDNA). RNA viruses have RNA as their genetic material. In this case, the nucleic acid is usually ssRNA, but may be dsRNA. Baltimore, (1971) introduced a classification system for viruses. This system classifies viral families into seven groups depending on a combination of their genetic material, strandedness, sense and method of replication, as shown below:

- (i) Group I: dsDNA viruses
- (ii) Group II: ssDNA viruses (positive-sense)
- (iii) Group III: dsRNA viruses
- (iv) Group IV: (+) ssRNA viruses (positive-sense)
- (v) Group V: (-) ssRNA viruses (negative-sense)
- (vi) Group VI: ssRNA-RT viruses (positive-sense RNA with DNA intermediate)
- (vii) Group VII: dsDNA-RT viruses (with ssRNA intermediate)

According to Baltimore classification, DNA viruses are composed of Groups I and II whereas RNA viruses are composed of Groups III, IV and V. Groups VI and VII are not considered as DNA or RNA viruses. Although Group VI viruses contain RNA genomes, they are not considered RNA viruses, they are called Retroviruses since they replicate through DNA intermediates. The term for RNA viruses that excludes Retroviruses is Riboviruses (Groups III, IV, V). On the other hand, viruses in Group VII contain DNA genomes, but they are considered as reverse transcribing viruses, not DNA viruses, because they replicate through RNA intermediates.

Because of the restricted size of their genomes, viruses must enter host cells to replicate, assemble and propagate. They have evolved to manipulate host cell processes to promote their own life cycles and also to escape host immune defence mechanisms (Goff, 2007; Watanabe *et al.*, 2010; Pichlmair *et al.*, 2012). Compared to DNA viruses, RNA viruses exhibit smaller genomes that usually encode for only a dozen of proteins (Eigen,

1993), but still can establish complex networks of interactions with host cells. On the other hand, DNA viruses, along their evolution, have captured and integrated large DNA sequences from the host organisms in their genome. Consequently, they can encode complex functional domains of eukaryotic origin which enables DNA viruses to finely tune the metabolism of infected cells in order to control their own replication. In contrast, RNA virus proteins only have distant or no homologies with their eukaryotic counterparts. This suggests that RNA viruses have evolved a different strategy, and interact with the host by using “homemade” protein-binding motifs. Therefore, infection strategies developed by DNA and RNA viruses to exploit the host machinery and control the antiviral responses are very different. It is known that, in general, DNA viruses are responsible for chronic infections whereas RNA viruses are responsible for acute infections (Vidalain and Tangy, 2010).

With information accumulated over the last decades, detailed models of virus structures, replication machineries, and pathophysiologies are available. However, to reach a system view of virus replication cycle in host cells, more functional analyses are necessary, especially in the field of virus-host interactions (Vidalain and Tangy, 2010). Thanks to recent efforts (Calderwood *et al.*, 2007; de Chassey *et al.*, 2008; Shapira *et al.*, 2009; Pichlmair *et al.*, 2012), the field of virus-human protein interactions is currently just a step further beyond its infancy.

The most comprehensive comparative analysis performed on virus-human protein interactions compared infection strategies used by DNA and RNA viruses in terms of their common and unique mechanisms (Pichlmair *et al.*, 2012). In that study, 1681 PHI data between 70 viral proteins and 579 human proteins were found by tandem affinity purification and then analyzed. In this thesis, the current viral PHI data belonging to DNA and RNA viruses in PHISTO (approximately 5000 PHI data) were analyzed comparatively to give the details of the most comprehensive virus-human interaction data. The found common and special infection strategies of DNA and RNA viruses will be valuable for future design of broad and specific antiviral therapies. However the results drawn from this analysis should be interpreted with caution since PHI data for lots of virus families are still scarce.

5.1. Protein Interaction Data

5.1.1. Virus-Human PHI Data

In the first release of PHISTO, pathogen strains were presented in groups, where each pathogen group is made up of taxonomically related strains. The group-based PHI data were analyzed in the previous chapter of this thesis. However, with increasing data content, grouping new pathogen strains became a difficult task, since there is no unique rule for grouping. On the other hand, in some groups such as Hepatitis, there are three species; Hepatitis B virus, Hepatitis C virus and Hepatitis E virus which belong to three different viral families; Hepadnaviridae, Flaviviridae and Hepeviridae, respectively. Hepadnaviridae is a DNA virus family whereas the others are RNA virus families. These different characteristics of the members of a group, make the analysis of a pathogen group in observing the behaviour of similar pathogen strains very difficult. Furthermore, to obtain PHI data belonging to strains of a species is not possible with Pathogen type (Bacteria, Fungi, Protozoa, Virus) → Group → Strain data presentation format for "Browse" function of PHISTO. Strains are wild-type, mutated or genetically modified versions of a species. Therefore, the taxonomical presentation of pathogens has been converted into a more systematic structure. In the current version of PHISTO, the pathogens, for which experimental PHI data are available, are stored and presented in this new classification for "Browse" function: Pathogen type → Family → Species → Strain. Therefore, users can display PHI data of the strains, which belong to the same pathogen species, by browsing the corresponding species.

The family-based virus-human PHIs were downloaded from PHISTO in February 2013. 26 viral families were covered by the downloaded data. 12 families carry DNA and 14 families carry RNA as their genetic materials. Hepadnaviruses were excluded from DNA viruses since they belong to Group VII in the Baltimore classification. Retroviruses were excluded from RNA families since they belong to Group VI. Consequently, PHI data belonging to 11 DNA virus families and 13 RNA virus families (Tables 5.1 and 5.2) were analyzed comparatively in this chapter.

Table 5.1. Contents of DNA viruses-human PHI data.

DNA virus family (Group #)	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
Adenoviruses (I)	12	125	33	77
Asfarviruses (I)	1	3	3	2
Baculoviruses (I)	1	1	1	1
Circoviruses (II)	2	4	2	3
Herpesviruses (I)	31	922	149	579
Myoviruses (I)	3	4	3	3
Papillomaviruses (I)	15	330	57	133
Parvoviruses (II)	1	4	2	2
Polyomaviruses (I)	5	74	11	53
Poxviruses (I)	11	213	36	149
Siphoviruses (I)	2	2	2	2
TOTAL	84	1682	299	835

Table 5.2. Contents of RNA viruses-human PHI data.

RNA virus family (Group #)	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
Arenaviruses (V)	5	16	5	13
Arteriviruses (IV)	1	56	1	56
Bunyaviruses (V)	7	174	7	138
Coronaviruses (IV)	2	7	6	6
Filoviruses (V)	3	115	5	100
Flaviviruses (IV)	21	1049	172	416
Hepeviruses (IV)	2	3	2	2
Orthomyxoviruses (V)	21	828	47	346
Paramyxoviruses (V)	11	856	22	638
Picornaviruses (IV)	6	9	8	5
Reoviruses (III)	7	61	9	55
Rhabdoviruses (V)	4	9	5	6
Togaviruses (IV)	1	1	1	1
TOTAL	91	3184	290	1359

5.1.2. Human Protein Sets

A total of 12 sets of human proteins interacting with viral pathogens were constructed from PHI data to analyze the properties of targeted human proteins. Firstly, the sets targeted by DNA viruses (DNA viruses-targeted set), RNA viruses (RNA viruses-targeted set), only DNA viruses, not targeted by any RNA viruses, (only DNA viruses-targeted set) and only RNA viruses, not targeted by any DNA viruses, (only RNA viruses-targeted set) were constructed to observe the attacking characteristics specific to DNA and RNA viruses.

For a deeper comparison between DNA and RNA viral infections, human proteins interacting with at least two DNA virus families (two DNA viruses-targeted set) and three DNA virus families (three DNA viruses-targeted set) and also human proteins interacting with at least two RNA virus families (two RNA viruses-targeted set), three RNA virus families (three RNA viruses-targeted set), four RNA virus families (four RNA viruses-targeted set) and five RNA virus families (five RNA viruses-targeted set) were used. To obtain common infection strategies of viral pathogens, sets of human proteins targeted by DNA and/or RNA viruses (viruses-targeted set) and by both DNA and RNA viruses together (DNA-RNA viruses-targeted set) were also analyzed.

5.2. All DNA Viruses versus All RNA Viruses

In virus-human PHI data, RNA viruses target more than 1.5 times greater number of human proteins resulting nearly two times greater number of PHIs than DNA viruses with almost same number of strains and viral proteins (Table 5.3).

Table 5.3. Comparison of viral PHI data contents: DNA viruses versus RNA viruses.

Virus type	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
DNA viruses	84	1682	299	835
RNA viruses	91	3184	290	1359
TOTAL	175	4866	589	1823

Despite their very small encoding capacity, RNA viruses have the ability to interact with lots of human proteins owing to their multi-functional and multi-domain viral proteins. It is known that RNA virus proteins have on average more binding partners than host proteins of the same size (Shapira *et al.*, 2009). This higher connectivity implies that RNA viruses should encode proteins with more but smaller binding interfaces to be able to hit multiple cellular targets (Vidalain and Tangy, 2010).

5.2.1. Targeted Human Proteins

The distribution of 1823 viruses-targeted human proteins based on their targeting virus types is shown in the Venn diagram (Figure 5.1).

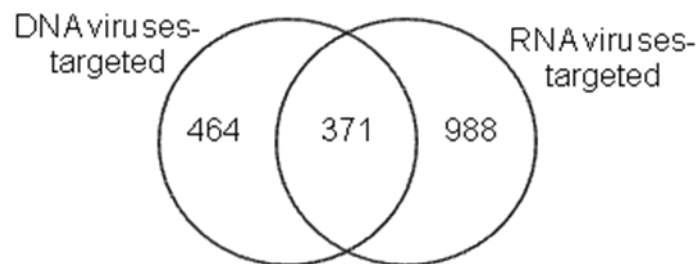


Figure 5.1. The number of human proteins targeted by DNA viruses and/or RNA viruses.

Targeting virus families for highly targeted human proteins such as three DNA viruses-targeted set and five RNA viruses-targeted set are given in Tables 5.4 and 5.5, respectively.

Table 5.4. Human proteins targeted by at least three DNA virus families.

Human protein (Degree)	Targeting DNA virus family
RB-Retinoblastoma-associated protein (149)	Adenoviruses, Herpesviruses, Papillomaviruses, Polyomaviruses
CDK1-Cyclin-dependent kinase 1 (190)	Adenoviruses, Herpesviruses, Papillomaviruses, Polyomaviruses
TBP-TATA-box-binding protein (147)	Adenoviruses, Herpesviruses, Papillomaviruses, Polyomaviruses
P85A-Phosphatidylinositol 3-kinase 85 kDa regulatory subunit alpha (402)	Circoviruses, Herpesviruses, Polyomaviruses, Poxviruses
IMA2-Importin subunit alpha-2 (106)	Herpesviruses, Papillomaviruses, Parvoviruses, Polyomaviruses
EP300-Histone acetyltransferase p300 (123)	Adenoviruses, Herpesviruses, Papillomaviruses, Polyomaviruses
HDAC1-Histone deacetylase 1 (189)	Adenoviruses, Herpesviruses, Papillomaviruses, Polyomaviruses
CBP-CREB-binding protein (147)	Adenoviruses, Herpesviruses, Papillomaviruses, Polyomaviruses
A8K3Z8-RAN, member RAS oncogene family, isoform CRA_b (NA)	Adenoviruses, Papillomaviruses, Polyomaviruses
MYC-Myc proto-oncogene protein (382)	Herpesviruses, Papillomaviruses, Polyomaviruses
P53-Cellular tumor antigen p53 (347)	Adenoviruses, Papillomaviruses, Polyomaviruses
DPOLA-DNA polymerase alpha catalytic subunit (52)	Herpesviruses, Papillomaviruses, Polyomaviruses
TAF1-Transcription initiation factor TFIID subunit 1 (58)	Adenoviruses, Papillomaviruses, Polyomaviruses
CDK2-Cyclin-dependent kinase 2 (151)	Herpesviruses, Papillomaviruses, Polyomaviruses
HNRPK-Heterogeneous nuclear ribonucleoprotein K (222)	Asfarviruses, Herpesviruses, Poxviruses
RAN-GTP-binding nuclear protein Ran (111)	Adenoviruses, Papillomaviruses, Polyomaviruses
SUMO1-Small ubiquitin-related modifier 1 (103)	Herpesviruses, Papillomaviruses, Poxviruses
YBOX1-Nuclease-sensitive element-binding protein 1 (173)	Herpesviruses, Polyomaviruses, Poxviruses
CSK2-Casein kinase II subunit alpha (80)	Herpesviruses, Papillomaviruses, Polyomaviruses
TF2B-Transcription initiation factor IIB (69)	Herpesviruses, Papillomaviruses, Polyomaviruses
BAK-Bcl-2 homologous antagonist/killer (23)	Adenoviruses, Papillomaviruses, Poxviruses

Table 5.5. Human proteins targeted by at least five RNA virus families.

Protein name (Degree)	Targeting RNA virus family
TBB5-Tubulin beta chain (195)	Bunyaviruses, Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
HNRH1-Heterogeneous nuclear ribonucleoprotein H (211)	Arteriviruses, Bunyaviruses, Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses
EF1A1-Elongation factor 1-alpha 1 (262)	Bunyaviruses, Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
TBA1A-Tubulin alpha-1A chain (140)	Arenaviruses, Bunyaviruses, Filoviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
H2A1B-Histone H2A type 1-B/E (NA)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
ROA1-Heterogeneous nuclear ribonucleoprotein A1 (246)	Arteriviruses, Coronaviruses, Filoviruses, Flaviviruses, Orthomyxoviruses
H2A1-Histone H2A type 1 (5)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
H2A1D-Histone H2A type 1-D (NA)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
ROA2-Heterogeneous nuclear ribonucleoproteins A2/B1 (189)	Arteriviruses, Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses
HNRH3-Heterogeneous nuclear ribonucleoprotein H3 (80)	Arteriviruses, Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses
HS71L-Heat shock 70 kDa protein 1-like (NA)	Bunyaviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
FUS-RNA-binding protein FUS (201)	Arteriviruses, Bunyaviruses, Filoviruses, Orthomyxoviruses, Paramyxoviruses
TCPE-T-complex protein 1 subunit epsilon (NA)	Arenaviruses, Bunyaviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
HNRPF-Heterogeneous nuclear ribonucleoprotein F (186)	Arteriviruses, Bunyaviruses, Filoviruses, Orthomyxoviruses, Paramyxoviruses
HNRH2-Heterogeneous nuclear ribonucleoprotein H2 (152)	Bunyaviruses, Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses
H4-Histone H4 (122)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
H2A2C-Histone H2A type 2-C (3)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
EF1A3-Putative elongation factor 1-alpha-like 3 (NA)	Bunyaviruses, Filoviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
H2A2A-Histone H2A type 2-A (3)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
H2A3-Histone H2A type 3 (3)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses

Table 5.5. Human proteins targeted by at least five RNA virus families (cont.)

Protein name (Degree)	Targeting RNA virus family
H2A1C-Histone H2A type 1-C (2)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
H2A1H-Histone H2A type 1-H (1)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
H2A1J-Histone H2A type 1-J (NA)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses
TBA1C-Tubulin alpha-1C chain (103)	Arenaviruses, Bunyaviruses, Filoviruses, Paramyxoviruses, Reoviruses
H2AJ-Histone H2A.J (NA)	Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses, Reoviruses

An important observation about these human proteins is that there are no common proteins in the two lists although there are 371 human proteins targeted by both DNA and RNA viruses in the PHI data analyzed (Figure 5.1). Therefore it can be said that these two lists of human proteins are specific targets for DNA and RNA virus infections separately, indicating their different infection strategies. Degree values of the highly targeted proteins (Tables 5.4 and 5.5) in the human PPI network was obtained from the results of Section 4.2.1. Compared to RNA viruses, DNA viruses are observed to interact with human proteins of higher connectivity values in the human network. The interesting point about these highly targeted human proteins is that the similar behaviour of DNA virus families (Table 5.4) and RNA virus families (Table 5.5) as they usually attack to same targets as follows: DNA viruses such as Adenoviruses, Herpesviruses, Papillomaviruses, Polyomaviruses and RNA viruses such as Filoviruses, Flaviviruses, Orthomyxoviruses, Paramyxoviruses are observed to behave similarly. It was suggested that piecing together targeted human proteins across multiple pathogens has the potential to provide insights into common molecular mechanisms of infection used by different pathogens (Dyer *et al.*, 2008). In this case, using the suggested approach, the highly targeted human proteins by DNA and RNA viruses may provide insights into common infection strategies used by different families belonging to same virus type.

In fact, these similarly-attacking (to human proteins) virus families cause very distinct diseases in human. *Adenoviruses* cause *infections* in the ear, in the *respiratory tract* and *gastrointestinal tract* (Martin *et al.*, 2007). Herpesviruses cause various human

diseases ranging from Kaposi sarcoma, B-cell lymphomas, chickenpox, shingles, and nasopharyngeal carcinoma (Davison and Scott, 1986; Chang *et al.*, 1994; Dolan *et al.*, 2006) to facial cold-sores and genital infections (Chayavichitsilp *et al.*, 2009). Papillomaviruses cause benign tumors and may cause cancerous ones (Munoz *et al.*, 2006). Polyomaviruses are potentially oncogenic, they often cause latent infections in hosts without causing disease, but may produce tumors in a host with an ineffective immune system (Egli *et al.*, 2009). On the other hand, Filoviruses cause hemorrhagic fevers (Kiley *et al.*, 1982). Flaviviruses cause a variety of threatening diseases such as Dengue fever, Japanese encephalitis, Hepatitis C virus infection, West Nile infection and yellow fever (Tuma *et al.*, 2009). Orthomyxoviruses are causes of Influenza (Hay *et al.*, 2001). Paramyxoviruses cause upper and lower respiratory tract infections (Galinski, 1991). The similar behaviour of these different virus families may be used in developing new-age infection therapeutics which target human proteins required by pathogens to replicate and persist within the host. Conventional high-throughput antiviral discovery often targets the activities of specific viral enzymes. These approaches have been ineffective in stemming the emergence of drug-resistant variants, especially in the face of rapidly-mutating RNA viruses (Murali *et al.*, 2011). If common human proteins can be identified as host factors (on which the members of these similarly-behaving family groups depend indispensably), generic antiviral drugs may be developed against a number of viruses effectively. Therefore, it can be proposed here that the common human targets of these two groups of DNA and RNA families (Adenoviruses-Herpesviruses-Papillomaviruses-Polyomaviruses and Filoviruses-Flaviviruses-Orthomyxoviruses-Paramyxoviruses) should be further investigated in terms of their host factor potential.

5.2.2. GO Enrichment Analysis

GO enrichments of all human protein sets were performed as explained in Section 4.2.2. The enriched GO process terms can be used to point out the human processes that are attacked by DNA/RNA viruses. Special attention should be paid to the results of sets of human proteins interacting with only DNA virus proteins (Table 5.6) and only RNA virus proteins (Table 5.7) to retrieve specific attack strategies of these different virus types. The human proteins targeted by multiple viral families reflect more specificity to infection

mechanisms of the corresponding virus types. Therefore, the results for three DNA viruses-targeted set (Table 5.8) and for five RNA viruses-targeted set (Table 5.9) are also given.

Table 5.6. First 20 enriched GO Process terms in proteins targeted by only DNA viruses.

GO Process Term	p-value
interspecies interaction between organisms	4.34E-28
positive regulation of cellular process	3.51E-25
regulation of cellular process	3.51E-25
regulation of biological process	6.18E-25
biological regulation	1.66E-24
cellular macromolecule metabolic process	7.53E-24
positive regulation of biological process	2.16E-23
macromolecule metabolic process	8.18E-21
regulation of macromolecule metabolic process	1.60E-19
multi-organism process	1.08E-18
regulation of molecular function	1.90E-18
regulation of metabolic process	2.03E-18
negative regulation of macromolecule metabolic process	1.06E-17
regulation of cellular metabolic process	8.58E-17
regulation of primary metabolic process	8.58E-17
cellular process	1.53E-16
negative regulation of cellular process	1.90E-16
regulation of cell cycle	4.89E-16
negative regulation of metabolic process	4.89E-16
negative regulation of biological process	1.09E-15

Table 5.7. First 20 enriched GO Process terms in proteins targeted by only RNA viruses.

GO Process Term	p-value
intracellular transport	3.94E-16
establishment of localization in cell	6.46E-15
establishment of protein localization	8.22E-15
protein transport	8.22E-15
macromolecule localization	3.29E-14
protein localization	2.30E-13
cellular localization	2.68E-13
cellular process	1.08E-12
cellular metabolic process	1.17E-10
establishment of localization	9.20E-10

Table 5.7. First 20 enriched GO Process terms in human proteins targeted by only RNA viruses (cont.)

GO Process Term	p-value
localization	5.57E-09
transport	7.51E-09
interspecies interaction between organisms	1.12E-08
metabolic process	3.14E-08
mitochondrial transport	5.10E-08
biosynthetic process	1.12E-07
mRNA transport	2.77E-07
generation of precursor metabolites and energy	5.09E-07
mitochondrial ATP synthesis coupled proton transport	8.48E-07
primary metabolic process	9.05E-07

Table 5.8. First 20 enriched GO Process terms in human proteins targeted by at least three DNA virus families.

GO Process Term	p-value
interspecies interaction between organisms	6.54E-15
multi-organism process	2.69E-10
positive regulation of transcription	1.64E-07
positive regulation of gene expression	1.96E-07
positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.42E-07
positive regulation of macromolecule biosynthetic process	3.42E-07
positive regulation of nitrogen compound metabolic process	3.42E-07
positive regulation of macromolecule metabolic process	3.94E-07
positive regulation of cellular metabolic process	4.34E-07
positive regulation of cellular biosynthetic process	4.34E-07
positive regulation of biosynthetic process	4.56E-07
positive regulation of metabolic process	6.01E-07
cell cycle	8.53E-07
cell cycle process	9.54E-07
positive regulation of cellular process	4.39E-06
positive regulation of transcription, DNA-dependent	5.14E-06
positive regulation of RNA metabolic process	5.30E-06
interphase of mitotic cell cycle	7.12E-06
interphase	8.18E-06
positive regulation of biological process	1.06E-05

Table 5.9. First 20 enriched GO Process terms in human proteins targeted by at least five RNA virus families.

GO Process Term	p-value
nucleosome assembly	1.92E-15
chromatin assembly	1.92E-15
protein-DNA complex assembly	1.94E-15
nucleosome organization	1.94E-15
cellular macromolecular complex assembly	3.26E-15
cellular macromolecular complex subunit organization	1.29E-14
DNA packaging	1.35E-14
chromatin assembly or disassembly	2.53E-14
DNA conformation change	5.72E-14
macromolecular complex assembly	3.02E-11
macromolecular complex subunit organization	6.56E-11
chromosome organization	1.03E-09
cellular component assembly	1.03E-09
chromatin organization	2.09E-09
cellular component biogenesis	4.29E-09
organelle organization	1.76E-06
cellular component organization	2.62E-05
RNA splicing	3.43E-05
protein polymerization	3.96E-04
mRNA processing	7.48E-04

Observation of the enriched GO process terms in Tables 5.6 and 5.8 concluded that DNA viruses have evolved mechanisms to attack human cellular and metabolic processes simultaneously. This can be achieved by their eukaryotic originated genetic ability. These PHI mechanisms allow DNA viruses to subvert the cellular and metabolic mechanisms of infected host cells to their advantage resulting generally chronic infectious diseases in human. Further investigation of the enriched processes of human proteins targeted highly by DNA viruses (Table 5.8), concluded their tendency to target cell cycle mechanisms. Pichlmair *et al.* (2012) reported that viORFs from DNA viruses preferentially target human proteins functioning in the cell cycle with either transcription or chromosome biology, possibly reflecting the necessity of uncoupling viral replication from cellular growth.

Enriched GO process terms in human proteins targeted by only RNA viruses (Table 5.7) and by five and more RNA virus families (Table 5.9) give crucial insights on infection strategies of RNA viruses. Firstly, the results in Table 5.7 indicate that only RNA viruses need the host mechanisms of intracellular transport and localization within the cell. It can be concluded that DNA viruses have their own transport and localization mechanisms, owing to their larger encoding capacity. Secondly, the enriched process terms in Table 5.9 show the specific processes within the cellular mechanisms highly targeted by RNA viruses. In contrast to DNA viruses, RNA viruses do not highly target metabolic processes of human resulting in generally acute infections. Pichlmair *et al.* (2012) reported that RNA viruses-targeted human proteins involved in RNA metabolism and also protein and RNA transport, while promoting efficient viral RNA processing and disrupting the link between signaling and immunity-related processes, for a successful infection.

Table 5.10. First 20 enriched GO Process terms in human proteins targeted by both DNA and RNA viruses.

GO Process Term	p-value
translational elongation	2.49E-61
translation	1.34E-41
gene expression	1.21E-37
cellular macromolecule metabolic process	1.66E-33
protein-DNA complex assembly	2.16E-31
nucleosome assembly	7.16E-29
nucleosome organization	1.69E-28
chromatin assembly or disassembly	1.69E-28
chromatin assembly	2.95E-28
cellular macromolecular complex assembly	2.27E-27
cellular protein metabolic process	1.79E-26
macromolecule metabolic process	2.97E-26
cellular macromolecular complex subunit organization	7.70E-26
macromolecular complex assembly	6.72E-25
interspecies interaction between organisms	7.05E-25
cellular macromolecule biosynthetic process	1.77E-24
macromolecular complex subunit organization	6.26E-24
macromolecule biosynthetic process	6.48E-24
DNA packaging	9.92E-24
cellular process	6.43E-23

On the other hand, the results for human proteins interacting with both DNA and RNA viruses (Table 5.10) are also important to highlight common infection mechanisms shared by the two types of viruses. The most interesting point here is the very low p-values, indicating statistically meaningful results for enriched GO process terms in human proteins targeted by both DNA and RNA viruses, together. The list includes specific cellular and metabolic processes which can be considered as targets of both virus types.

5.3. Virus Families with the Largest PHI Data

For a family-based analysis of PHI data, three families, with the largest amount of data were selected for each virus type: Herpesviruses, Papillomaviruses and Poxviruses for DNA virus families and Flaviviruses, Orthomyxoviruses and Paramyxoviruses for RNA virus families. The PHI data belonging to these families (Tables 5.11 and 5.12) cover more than 85% of the total data (Tables 5.1 and 5.2) in each case and therefore can reflect the behaviour of the corresponding virus types. To observe the family-based interaction characteristics of viral proteins with human proteins, the targeted human proteins are presented with respect to their interactions with the families (Figures 5.2 and 5.3).

5.3.1. DNA Virus Families

Herpesviruses, Papillomaviruses and Poxviruses are DNA virus families which have the largest amount of PHI data (Table 5.1). The contents of their PHI data and the Venn diagram for the distribution of human proteins on interactions with these families are given in Table 5.11 and Figure 5.2.

Table 5.11. Contents of PHI data belonging to DNA virus families with the largest data.

DNA virus family	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
Herpesviruses	31	922	149	579
Papillomaviruses	15	330	57	133
Poxviruses	11	213	36	149
TOTAL	57	1465	242	770

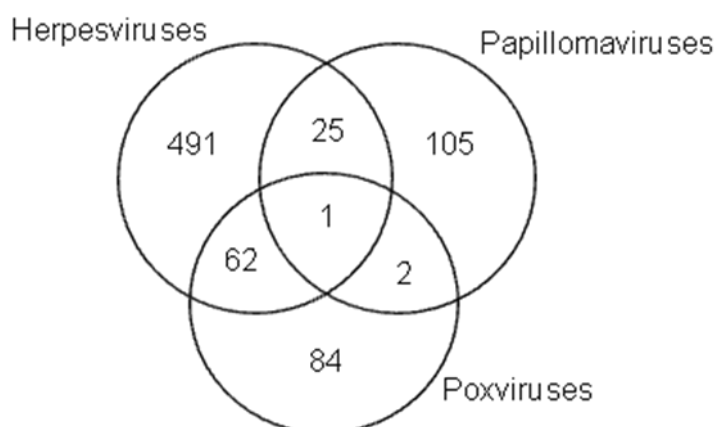


Figure 5.2. The number of DNA virus families-targeted human proteins that are grouped based on their interactions with families; Herpesviruses, Papillomaviruses and Poxviruses.

The only human protein targeted by all three DNA virus families under investigation is SUMO1 (Small ubiquitin-related modifier 1) which has role in RNA transport pathway. SUMO1 is also a common target of bacteria and viruses other than these three families (Table 4.5).

Herpesviruses and Poxviruses act more similarly on attacking to human proteins (having larger intersection for targeted human proteins in Figure 5.2) during infections.

5.3.2. RNA Virus Families

Flaviviruses, Orthomyxoviruses and Paramyxoviruses are RNA virus families which have the largest amount of PHI data (Table 5.2). The contents of their PHI data and the Venn diagram for the distribution of human proteins on interactions with these families are given in Table 5.12 and Figure 5.3.

Table 5.12. Contents of PHI data belonging to RNA virus families with the largest data.

RNA virus family	# of strains	# of PHIs	# of pathogen proteins	# of human proteins
Flaviviruses	21	1049	172	416
Orthomyxoviruses	21	828	47	346
Paramyxoviruses	11	856	22	638
TOTAL	53	2733	241	1204

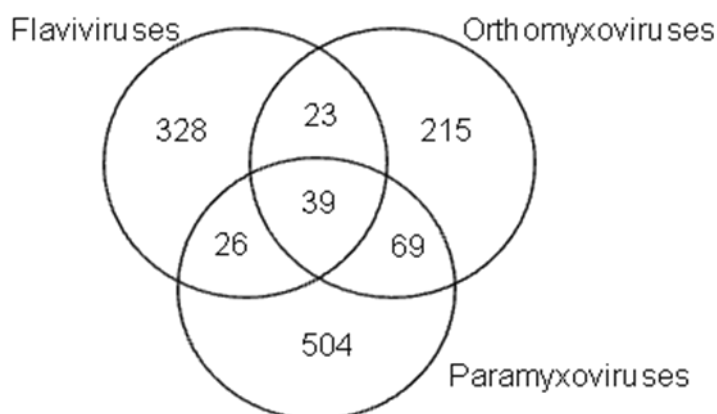


Figure 5.3. The number of RNA virus families-targeted human proteins that are grouped based on their interactions with families; Flaviviruses, Orthomyxoviruses and Paramyxoviruses.

Compared to DNA virus families, RNA virus families have much more common targeted human proteins (Figures 5.2 and 5.3). Especially, Orthomyxoviruses and Paramyxoviruses behave more similarly, each has interactions with same 108 human proteins. This observation may indicate that RNA viruses have more common infection strategies with other RNA viruses than in the case of DNA viruses.

This makes sense considering more complex viral mechanisms and therefore more specialized infection strategies for each member of DNA viruses. Compared to RNA viruses, DNA viruses have larger encoding capacity. These large viruses can encode up to 900 proteins (Van Etten *et al.*, 2010). DNA viruses can have more specialized functioning proteins, in contrast to multi-functional RNA virus proteins. Therefore the needs may be similar in RNA viruses case, lacking most of the mechanisms, depending largely on host's mechanisms. During infections they interact with large number of human proteins despite their very small number of viral proteins. On the other hand, the needs in the life cycle may differ largely in DNA viruses case and they can meet some of their requirements using their own mechanisms. It can be concluded that different DNA virus families target less number of common human proteins since their needs and the dependency on the host differ largely compared to RNA virus families.

5.4. Virus Strains with the Largest PHI Data

For a strain-based analysis of PHI data, three virus strains, with the largest amount of data were selected for each family analyzed in Section 5.3. Totally nine strains were selected for each DNA and RNA virus types. These three strains with the largest PHI data (Tables 5.13 and 5.14) have more than 70% of all data belonging to corresponding families (Tables 5.11 and 5.12) in each case and therefore can reflect the behaviour of the virus families. To observe the strain-based interaction characteristics of viral proteins with human proteins, the targeted human proteins are presented with respect to their interactions with the strains (Figures 5.4-5.9).

5.4.1. DNA Virus Strains

The largest PHI data-related strains belonging to the families; Herpesviruses, Papillomaviruses and Poxviruses were analyzed. The contents of their PHI data and the Venn diagrams for the distribution of human proteins on interactions with these strains are given in Table 5.13 and Figures 5.4-5.6.

Table 5.13. Contents of PHI data belonging to DNA virus strains with the largest data.

DNA virus family	Virus strains (NCBI taxonomy ID)	# of PHIs	# of pathogen proteins	# of human proteins
Herpes viruses	HHV4 Strain B95-8 (10377)	302	40	245
	HHV1 Strain 17 (10299)	264	15	239
	HHV4 (10376)	121	31	92
	TOTAL	687	86	471
Papilloma viruses	HPV 16 (333760)	127	12	106
	HPV11 (10580)	67	11	40
	HPV18 (333761)	46	5	43
	TOTAL	240	28	125
Pox viruses	VacV Strain WR (10254)	166	14	132
	Orf virus Strain SA00 (647330)	16	5	4
	VacV Strain Copenhagen (10249)	9	4	9
	TOTAL	191	23	139

5.4.1.1. Herpesviruses. The commonly targeted human proteins by all three Herpesvirus strains are IGHG1 (Ig gamma-1 chain C region, functioning in innate immune responses), RL4 (60S ribosomal protein L4, functioning in translation), PP1A (Serine/threonine-protein phosphatase PP1-alpha catalytic subunit, functioning in mRNA surveillance pathway, oocyte meiosis and in other signaling pathways) and LG3BP (Galectin-3-binding protein, functioning in promotion of integrin-mediated cell adhesion) (Figure 5.4).

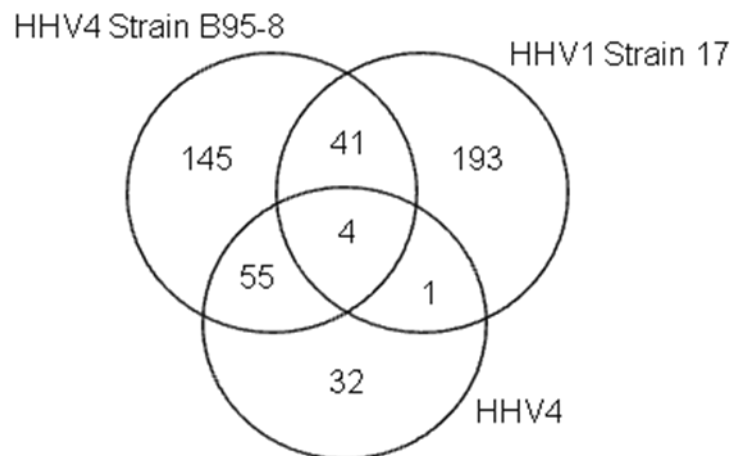


Figure 5.4. The number of Herpesvirus strains-targeted human proteins that are grouped based on their interactions with strains; Human Herpesvirus 4 Strain B95-8, Human Herpesvirus 1 Strain 17 and Human Herpesvirus 4.

5.4.1.2. Papillomaviruses. Papillomavirus strains target more common human proteins (as percentage of total human proteins targeted by the three strains) than Herpesviruses (see intersections of the Venn diagrams in Figures 5.4 and 5.5). In other words, the numbers of targeted human proteins specific to strains (targeted by only one strain) are smaller in Papillomaviruses than Herpesviruses. This may explain the distinct diseases caused by different Herpesviruses whereas members of Papillomaviruses cause more similar diseases.

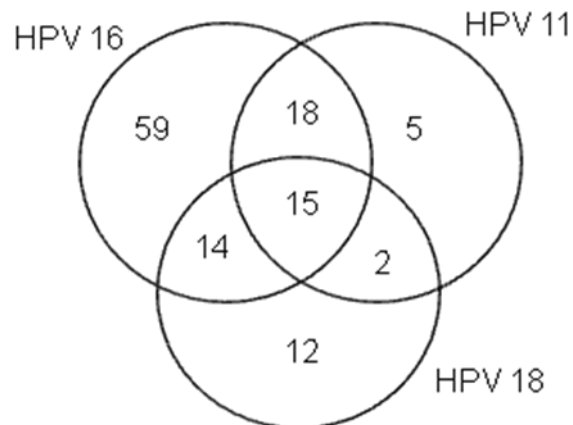


Figure 5.5. The number of Papillomavirus strains-targeted human proteins that are grouped based on their interactions with strains; Human Papillomavirus 16, Human Papillomavirus 11 and Human Papillomavirus 18.

5.4.1.3. Poxviruses. In Poxviruses any conclusion cannot be drawn since very limited data are available for two Poxvirus strains; Orf virus Strain SA00 and Vaccinia virus Strain Copenhagen (Figure 5.6).

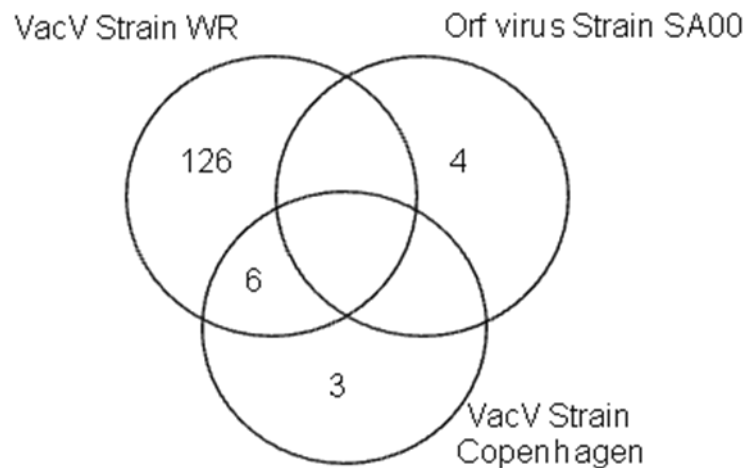


Figure 5.6. The number of Poxvirus strains-targeted human proteins that are grouped based on their interactions with strains; Vaccinia virus Strain WR, Orf virus Strain SA00 and Vaccinia virus Strain Copenhagen.

5.4.2. RNA Virus Strains

The largest PHI data-related strains belonging to the families; Flaviviruses, Orthomyxoviruses, and Paramyxoviruses were analyzed. The contents of their PHI data and the Venn diagrams for the distribution of human proteins on interactions with these strains are given in Table 5.14 and Figures 5.7-5.9.

Table 5.14. Contents of PHI data belonging to RNA virus strains with the largest data.

RNA virus family	Virus strains (NCBI taxonomy ID)	# of PHIs	# of pathogen proteins	# of human proteins
Flavi viruses	HCV (11103)	609	148	43
	HCV Isolate Con1 (333284)	323	1	323
	HCV Isolate H (11108)	48	1	48
	TOTAL	980	150	384
Orthomyxo viruses	Influenza A virus H1N1-1934 (211044)	385	16	241
	Influenza A virus H5N1-1997 (130763)	108	1	108
	Influenza A virus H3N2-1972 (381517)	102	8	87
	TOTAL	595	25	288
Paramyxo viruses	Measles virus Strain Schwarz (132487)	448	3	443
	HRSV Strain A2 (11259)	221	2	221
	Hendra virus Isolate Horse (928303)	67	4	65
	TOTAL	736	9	619

5.4.2.1. Flaviviruses. Although RNA virus strains have more common human protein targets compared to DNA viruses, strains of Hepatitis C virus (HCV) species, do not target commonly as much human proteins as other RNA virus strains belonging to same families (Figures 5.7-5.9). Observing this attacking strategy of viral strains belonging to HCV species, it can be speculated that the strains can mutate largely to have different characteristics (attacking to different human targets). It is known that, the infection mechanisms of HCV are differing largely from other RNA viruses. In general, RNA viruses are responsible for acute infections; HCV is one of the few noticeable exceptions of this, causing chronic infection similar to DNA viruses (Vidalain and Tangy, 2010).

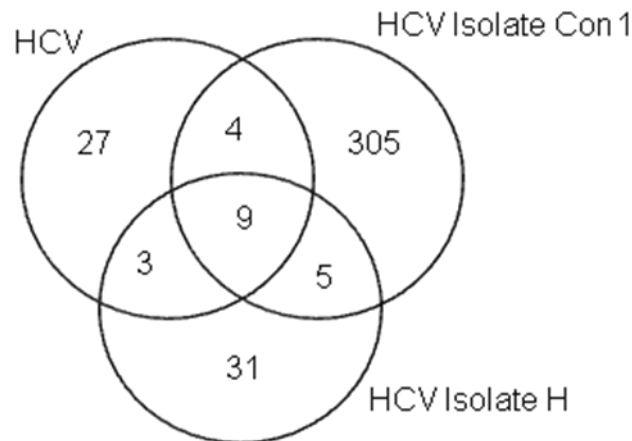


Figure 5.7. The number of Flavivirus strains-targeted human proteins that are grouped based on their interactions with strains; Hepatitis C virus, Hepatitis C virus Isolate Con1 and Hepatitis C virus Isolate H.

5.4.2.2. Orthomyxoviruses. The commonly targeted human proteins by all three Orthomyxovirus strains are PRKRA (Interferon-inducible double stranded RNA-dependent protein kinase activator A) and H4 (Histone H4, one of core components of nucleosome) (Figure 5.8).

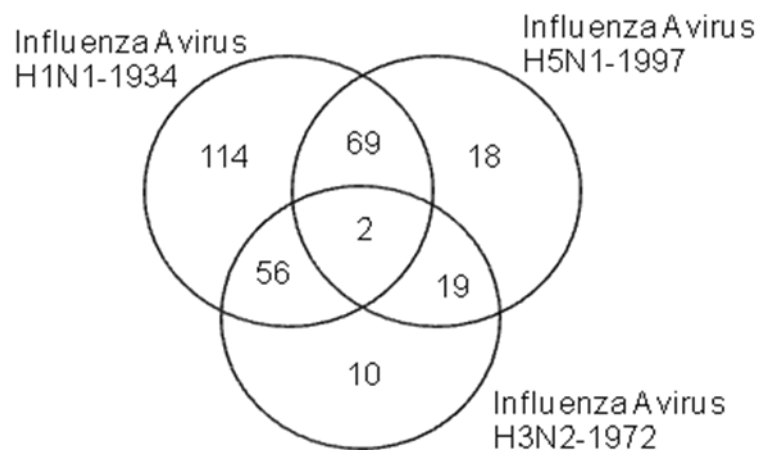


Figure 5.8. The number of Orthomyxovirus strains-targeted human proteins that are grouped based on their interactions with Influenza A virus strains; H1N1-1934, H5N1-1997 and H3N2-1972.

The interesting point about the distribution of human proteins targeted by different strains of Influenza A virus (cause of flu), is that they commonly target lots of proteins. Because of mutations of Influenza A viruses (continuously emerging new strains, may be resistant to current antiviral therapeutics), flu became a threatening disease in the 21st century. The desired therapeutic solution should allow handling strains together. This may be possible by identifying host factors among commonly-targeted human proteins. In other words, silencing the host factor on which more than one Influenza A strain depend, may inhibit their infection effectively.

5.4.2.3. Paramyxoviruses. The commonly targeted human proteins by all three Paramyxovirus strains are 1433B (14-3-3 protein beta/alpha), 1433E (14-3-3 protein epsilon), 1433Z (14-3-3 protein zeta/delta) (all these three proteins have functions in cell cycle, oocyte meiosis, PI3K-Akt signaling pathway and Hippo signaling pathway) and TBA1C (Tubulin alpha-1C chain, functioning in phagosome and gap junction) (Figure 5.9).

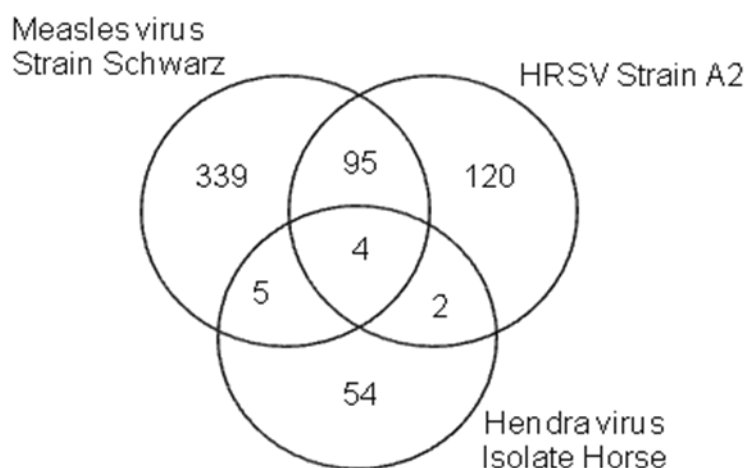


Figure 5.9. The number of Paramyxovirus strains-targeted human proteins that are grouped based on their interactions with strains; Measles virus Strain Schwarz, Human respiratory syncytial virus Strain A2 and Hendra virus Isolate Horse.

Measles virus and HRSV act more similarly on attacking to human proteins (having larger intersection for targeted human proteins in Figure 5.9) during infections.

5.5. Graph Analysis of Viral PHI Networks

Topology of intracellular protein interaction networks was related to the cell's functional organizations successfully (Barabasi and Oltvai, 2004; Przulj *et al.*, 2004). In parallel, the topological characteristics of PHI networks may contribute largely to enlighten infection mechanisms. However, the current available data on interspecies protein interactions between pathogenic microorganisms and their hosts are not enough to deduce topological network properties. For instance, the first attempt to fit the degree distribution of EBV-human interaction network failed due to scarcity of data (Figure 2.2), the degree distribution of the proteins in the virus-human PHI network could not be fitted to any model (Calderwood *et al.*, 2007). Since this first attempt until today, considerable amounts of experimentally-verified PHI data have not been produced to model degree distributions of pathogen/human proteins. However, the degree values of each interacting pathogen and human proteins were calculated for the PHI networks for 18 virus strains analyzed in Section 5.4. The top highly interacting pathogen and human proteins are tabulated in Table 5.15.

The highly interacting pathogen and human proteins may be proposed as putative antiviral drug targets. Here, only one human protein and one pathogen protein having the largest degree values in the PHI networks are listed. The following highly-targeted human proteins should be further investigated for their potential to be human factors on which corresponding viral strains are dependent. To serve host factor property, human protein must be non-essential for human. Host factors are indispensable for pathogens, but not essential for host cells (Murali *et al.*, 2011). On the other hand, highly interacting viral proteins may serve as targets for antiviral therapeutics. In this case, their homology with human proteins should be investigated since anti-infection therapeutics should target essential genes in the pathogens which have no homology with human genes (Barker, 2006).

Table 5.15. Proteins with the highest degree values for PHI data of DNA viruses.

Family	Virus Strains	Viral protein (# of interaction)	Human protein (# of interaction)
Herpes viruses	HHV4 Strain B95-8	EBNA5 (119)	HOME3 (10)
	HHV1 Strain 17	RNB (141)	LMNA, TBB5, RS18 (3)
	HHV4	DEN (16)	PSA3 (6)
Papilloma viruses	HPV 16	VE7 (49)	RB, IMA2, IMB1, TNPO1 (3)
	HPV11	VE6 (12)	EP300 (4)
	HPV18	VE6 (23)	EP300, DLG1, KAT2B (2)
Pox viruses	VacV Strain WR	E3 (45)	TBB5, TBA1C (4)
	Orf virus Strain SA00	Q6TVJ1 (4)	SKP1, CUL1, RBX1 (5)
	VacV Strain Copenhagen	K7, VPK1 (3)	no protein with more than 1 degree

Table 5.16. Proteins with the highest degree values for PHI data of RNA viruses.

Family	Virus Strains	Viral protein (# of interaction)	Human protein (# of interaction)
Flavi viruses	HCV	Q6QZG4 (25)	P53 (140)
	HCV Isolate Con1	POLG (323)	no protein with more than 1 degree
	HCV Isolate H	POLG (48)	no protein with more than 1 degree
Orthomyxo viruses	Influenza A virus H1N1-1934	NS1 (162)	LNK2 (11)
	Influenza A virus H5N1-1997	NS1 (108)	no protein with more than 1 degree
	Influenza A virus H3N2-1972	NS1 (44)	LNK2 (5)
Paramyxo viruses	Measles virus Strain Schwarz	Q77M19 (436)	TBA1C, GRP78, TBA1A, FOXK1, FOXK2 (2)
	HRSV Strain A2	NS1 (220)	no protein with more than 1 degree
	Hendra virus Isolate Horse	W (53)	STAT1, STAT2 (2)

6. CONCLUSIONS AND RECOMMENDATIONS

Within the framework of the present thesis on investigation of infection mechanisms through pathogen-human protein-protein interactions, firstly a novel software, PHISTO was developed to contribute the research field. Then two comparative interactomics analyses were performed to elaborate on infection mechanisms of bacteria versus viruses and DNA viruses versus RNA viruses, using PHI data from PHISTO.

6.1. Conclusions

An initial system-level understanding of infection mechanisms through PHI networks was pursued in this thesis by bioinformatics approaches. The following conclusions can be drawn on the basis of the results presented.

- PHISTO was developed to serve as an up-to-date and functionally-enhanced source of comprehensive PHI data through a user-friendly interface (www.phisto.org).

- Integrated tools for visualization of PHI networks, graph-theoretical analysis of targeted human proteins, BLAST search, and text mining for detecting missing experimental methods were implemented in PHISTO.

- The preference of interacting with hub and bottleneck proteins was found to be a common infection strategy of all types of pathogens to manipulate essential mechanisms in human. Compared to bacteria, viruses tend to interact with human proteins of much higher connectivity and centrality values in the human network.

- GO enrichment analysis of the human proteins targeted by pathogens indicates crucial clues about the infection mechanisms of bacteria and viruses. As the main infection strategy, bacteria interact with human proteins that function in immune response to disrupt human defense mechanisms. Indispensable viral strategy, on the other hand, is the manipulation of human cellular processes in order to use that transcriptional machinery for their own genetic material transcription. A novel observation about pathogen-human

systems is that the human proteins targeted by both pathogens are enriched in the regulation of metabolic processes in order to use human metabolism via destroying its regulatory mechanisms.

- Compared to RNA viruses, DNA viruses were observed to interact with human proteins of higher connectivity values in the human network.

- GO enrichment analysis of the human proteins targeted by DNA and RNA viruses indicates: DNA viruses have evolved mechanisms to attack human cellular and metabolic processes simultaneously. RNA viruses attack human processes especially involved in protein transport/localization and specific cellular mechanisms indicating their lacking own abilities. Translation-related processes are targeted highly by both DNA and RNA viruses.

6.2. Recommendations

This study aims to provide a framework for further investigations of infection mechanisms through PHI networks with increasing amount of precise information on pathogen-host systems in future.

6.2.1. Improvements on PHISTO

PHISTO was developed to fulfill the requirement for a comprehensive public PHI data source. Its first version with only PHI data content was released in July 2011. The second version enriched with basic bioinformatics tools was released in August 2012. The analysis tools planned to be implemented into PHISTO, for its third version in the near future are :

- Visualization of PHI networks as bipartite graphs is currently available for only small networks which are composed of up to 75 interactions between pathogen proteins and human proteins. Visualization of larger PHI networks as bipartite graphs will be available.

- "GO Enrichment" tool will be implemented for computations similar to analysis performed within 4th and 5th chapters of this thesis (Sections 4.2.2 and 5.2.2). Users will be able to obtain enriched GO process, function and component terms for targeted human proteins by the selected pathogen families/species/strains.

- "Comparative Interactomics" tool will be implemented for computations similar to analysis performed within sections 5.3 and 5.4. Users will be able to access targeted human proteins commonly or specifically by selected two or more pathogen families/species/strains.

- "Graph Analysis of PHI Networks" tool will be implemented for computations similar to analysis performed within section 5.5. Users will be able to access degree values of all pathogen/human proteins within the PHI network under investigation to observe the highly interacting proteins easily. Furthermore, degree distributions of pathogen/human proteins separately will be available to observe the general interaction characteristics of the PHI networks.

With PHISTO it is aimed to provide a centralized platform for studying PHI systems with additional relevant data content and advanced bioinformatics analysis tools. It is planned to release a newer version with enriched properties nearly once in a year. Additional bioinformatics analysis tools will be continuously developed and implemented into PHISTO to assist researchers during investigation of PHI networks. The functionality of the tools will be designed with needs following the increase in available data in future.

6.2.2. Future Works on PHI Analysis

For the majority of pathogenic microorganisms, the elucidation of interactomes with their hosts is still in its infancy and countless PHIs remain to be discovered and characterized. With the currently available PHI data, large-scale multi-species comparisons (e.g. analyses in the 4th and 5th chapters of this thesis), provide crucial insights into general mechanisms of PPIs between host and pathogen. However a lot more effort is required for a complete understanding of pathogenesis of infections with increasing data. The needed future works on the results of comparative analyses performed in this thesis as follows:

- In order to investigate further the preference of pathogens on targeting hub and bottleneck proteins, it might be worthwhile to differentiate bacteria into intracellular and extracellular pathogens and to order the pathogens according to their respective coding capacity and/or the number of pathogen effector proteins to see the effect of these on their preferences.

- Highly targeted human proteins (found in Chapters 4 and 5) should be investigated experimentally in terms of their potential to be host factors on which pathogens depend on.

- RNA viruses should be further analyzed in a Group-based fashion (Group III: dsRNA viruses, Group IV: (+) ssRNA viruses and Group V: (-) ssRNA viruses) to compare the effects of their strandedness, sense and method of replication on infection strategies.

- Retroviruses (which are not considered as RNA viruses despite their RNA genetic material, from Group VI of the Baltimore classification) including HIV, should be comparatively analyzed with RNA viruses to point out the differences.

- Molecular pathogen-host interaction data other than PPIs (e.g. pathogenic microRNAs interact largely with host molecular machinery during infection) should be integrated and analyzed.

- Intraspecies PPIs of interacting pathogen and human proteins within PHI networks should be analyzed together with PHIs to provide a system view of PHI systems with intra- and interspecies interaction networks underlying pathogenesis of infections.

Continued studies and increasing amounts of experimentally-verified PHI data will further improve our understanding of the interplay between pathogens and human and hopefully indentify novel and effective therapeutics for infectious diseases.

REFERENCES

- Albert, R., H. Jeong, A. L. Barabasi, 2000, "Error and Attack Tolerance of Complex Networks", *Nature*, Vol. 406, pp. 378-382.
- Alcami, A. and U. H. Koszinowski, 2000, "Viral Mechanisms of Immune Evasion" *Immunology Today*, Vol. 21, pp. 447-455.
- Alfarano, C., C. E. Andrade, K. Anthony, N. Bahroos, M. Bajec, K. Bantoft, D. Betel, B. Bobechko, K. Boutilier, E. Burgess, K. Buzadzija, R. Caverio, C. D'Abreo, I. Donaldson, D. Dorairajoo, M. J. Dumontier, M. R. Dumontier, V. Earles, R. Farrall, H. Feldman, E. Garderman, Y. Gong, R. Gonzaga, V. Grytsan, E. Gryz, V. Gu, E. Haldorsen, A. Halupa, R. Haw, A. Hrvojic, L. Hurrell, R. Isserlin, F. Jack, F. Juma, A. Khan, T. Kon, S. Konopinsky, V. Le, E. Lee, S. Ling, M. Magidin, J. Moniakakis, J. Montojo, S. Moore, B. Muskat, I. Ng, J. P. Paraiso, B. Parker, G. Pintilie, R. Pirone, J. J. Salama, S. Sgro, T. Shan, Y. Shu, J. Siew, D. Skinner, K. Snyder, R. Stasiuk, D. Strumpf, B. Tuekam, S. Tao, Z. Wang, M. White, R. Willis, C. Wolting, S. Wong, A. Wrong, C. Xin, R. Yao, B. Yates, S. Zhang, K. Zheng, T. Pawson, B. F. Ouellette, C. W. Hogue, 2005, "The Biomolecular Interaction Network Database and Related Tools: 2005 update", *Nucleic Acids Research*, Vol. 33, pp. D418-D424.
- Alonso, A., N. Bottini, S. Bruckner, S. Rahmouni, S. Williams, S. P. Schoenberger, T. Mustelin, 2004, "Lck Dephosphorylation at Tyr-394 and Inhibition of T Cell Antigen Receptor Signaling by *Yersinia* Phosphatase YopH", *The Journal of Biological Chemistry*, Vol. 279, pp. 4922-4928.
- Anthony, S., V. Sintchenko, E. Coiera, 2010, "Text Mining for Discovery of Host-Pathogen Interactions", In: Sintchenko, V. (Ed.), *Infectious Disease Informatics*, Springer Science+Business Media, New York, pp. 149-165.

- Aoki-Kinoshita, K. F. and M. Kanehisa, 2007, "Gene Annotation and Pathway Mapping in KEGG", In: Nicholas, H. G. (Ed.) *Comparative Genomics, Methods in Molecular Biology*, Humana Press, New Jersey, pp. 71–91.
- Apweiler, R., A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, L. -S. L. Yeh, 2004, "UniProt: The Universal Protein Knowledgebase", *Nucleic Acids Research*, Vol. 32, pp. D115-119.
- Aranda, B., H. Blankenburg, S. Kerrien, F. S. L. Brinkman, A. Ceol, E. Chautard, J. M. Dana, J. D. L. Rivas, M. Dumousseau, E. Galeota, A. Gaulton, J. Goll, R. E. W. Hancock, R. Isserlin, R. C. Jimenez, J. Kerssemakers, J. Khadake, D. J. Lynn, M. Michaut, G. O'Kelly, K. Ono, S. Orchard, C. Prieto, S. Razick, O. Rigina, L. Salwinski, M. Simonovic, S. Velankar, A. Winter, G. Wu, G. D. Bader, G. Cesareni, I. M. Donaldson, D. Eisenberg, G. J. Kleywegt, J. Overington, S. Richard-Blum, M. Tyers, M. Albrecht, H. Hermjakob, 2011, "PSICQUIC and PSISCORE: Accessing and Scoring Molecular Interactions", *Nature Methods*, Vol. 8, pp. 528-529.
- Arifuzzaman, M., M. Maeda, A. Itoh, K. Nishikata, C. Takita, R. Saito, T. Ara, K. Nakahigashi, H.-C. Huang, A. Hirai, K. Tsuzuki, S. Nakamura, M. Altaf-Ul-Amin, T. Oshima, T. Baba, N. Yamamoto, T. Kawamura, T. Ioka-Nakamichi, M. Kitagawa, M. Tomita, S. Kanaya, C. Wada, H. Mori, 2006, "Large Scale Identification of Protein-Protein Interaction of *Escherichia coli* K-12", *Genome Research*, Vol. 16, pp. 686-691.
- Ashburner, M., C. Ball, J. Blake, D. Botstein, 2000, "Gene Ontology: Tool for the Unification of Biology", *Nature*, Vol. 25, pp. 25-29.
- Bailer, S. M. and J. Haas, 2009, "Connecting Viral with Cellular Interactomes", *Current Opinion in Microbiology*, Vol. 12, pp. 453-459.
- Baltimore, D., 1971, "Expression of Animal Virus Genomes", *Bacteriological Reviews*, Vol. 35, pp. 235-241.

- Barabasi, A. L. and R. Albert, 1999, "Emergence of Scaling in Random Networks", *Science*, Vol. 286, pp. 509-512.
- Barabasi, A. L. and Z. N. Oltvai, 2004, "Network Biology: Understanding the Cell's Functional Organization" *Nature Reviews Genetics*, Vol. 5, pp. 101-113.
- Barker, J. J., 2006, "Antibacterial Drug Discovery and Structure-Based Design", *Drug Discovery Today*, Vol. 11, pp. 391-404.
- Bartel, P., J. Roecklein, D. SenGupta, 1996, "A Protein Linkage Map of *Escherichia coli* Bacteriophage T7", *Nature Genetics*, Vol. 12, pp. 72-77.
- Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, E. W. Sayers, 2013, "GenBank", *Nucleic Acids Research*, Vol. 41, pp. D36-D42.
- Brass, A. L., D. M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R. J. Xavier, J. Lieberman, S. J. Elledge, 2008, "Identification of Host Proteins Required for HIV Infection through a Functional Genomic Screen", *Science*, Vol. 319, pp. 921-926.
- Brennan, R. O. and D. T. Durack, 1981, "Gay Compromise Syndrome", *Lancet*, Vol. 2, pp. 1338-1339.
- Buck, D., M. E. Spencer, J. R. Guest, 1985, "Primary Structure of the Succinyl- CoA Synthetase of *Escherichia coli*", *Biochemistry*, Vol. 24, pp. 6245-6252.
- Bushman, F. D., N. Malani, J. Fernandes, I. D'Orso, G. Cagney, T. L. Diamond, H. Zhou, D. J. Hazuda, A. S. Espeseth, R. König, S. Bandyopadhyay, T. Ideker, S. P. Goff, N. J. Krogan, A. D. Frankel, J. A. Young, S. K. Chanda, 2009, "Host Cell Factors in HIV Replication: Meta-Analysis of Genome-Wide Studies", *PLoS Pathogens*, Vol. 5, e1000437.
- Butland, G., J. M. Peregrin-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattie, N. Krogan, M. Davey, J. Parkinson, J. Greenblatt, A. Emili,

- 2005, "Interaction Network Containing Conserved and Essential Protein Complexes in *Escherichia coli*", *Nature*, Vol. 433, pp. 531-537.
- Bünnige, M., 2011, *New Strategies to Fight Infectious Diseases - Arms Race on a Microscale*, http://www.infection-research.de/perspectives/detail/pressrelease/new_strategies_to_fight_infectious_diseases_arms_race_on_a_microscale, accessed at April 2012.
- Calderwood, M. A., K. Venkatesan, L. Xing, M. R. Chase, A. Vazquez, A. M. Holthaus, A. E. Ewence, N. Li, T. Hirozane-Kishikawa, D. E. Hill, M. Vidal, E. Kieff, E. Johannsen, 2007, "Epstein-Barr Virus and Virus Human Protein Interaction Maps", *Proceedings of the National Academy of Sciences*, Vol. 104, pp. 7606-7611.
- Carrillo, E., E. Garrido, P. Gariglio, 2004, "Specific in vitro Interaction Between Papillomavirus E2 Proteins and TBP-Associated Factors", *Intervirology*, Vol. 47, pp. 342-349.
- Chang, Y., E. Cesarman, M. S. Pessin, F. Lee, J. Culpepper, D. M. Knowles, P. S. Moore, 1994, "Identification of Herpesvirus-Like DNA Sequences in AIDS-Associated Kaposi's Sarcoma", *Science*, Vol. 266, pp. 1865-1869.
- Chatr-aryamontri, A., A. Ceol, D. Peluso, A. Nardozza, S. Panni, M. Tinti, A. Smolyar, L. Castagnoli, M. Vidal, M. E. Cusick, G. Cesareni, 2009, "Virus-MINT: A Viral Protein Interaction Database", *Nucleic Acids Research*, Vol. 37, pp. D669-673.
- Chayavichitsilp, P., J. V. Buckwalter, A. C. Krakowski, S. F. Friedlander, 2009, "Herpes Simplex", *Pediatrics in Review*, Vol. 30, pp.119-129.
- Croft, D., G. O'Kelly, G. Wu, R. Haw, 2011, "Reactome: A Database of Reactions, Pathways and Biological Processes", *Nucleic Acids Research*, Vol. 39, pp. D691-D697.

- Cui, T., L. Zhang, X. Wang, Z. -G. He, 2009, "Uncovering New Signaling Proteins and Potential Drug Targets through the Interactome Analysis of *Mycobacterium tuberculosis*", *BMC Genomics*, Vol. 10, p. 118.
- Datsenko, K. A. and B. L. Wanner, 2000, "One-Step Inactivation of Chromosomal Genes in *Escherichia coli* K-12 Using PCR Products", *Proceedings of the National Academy of Sciences*, Vol. 97, pp. 6640-6645.
- Davis, F. P., D. T. Barkan, N. Eswar, J. H., McKerrow, A. Sali, 2007, "Host Pathogen Protein Interactions Predicted by Comparative Modeling", *Protein Science*, Vol. 16, pp. 2585-2596.
- Davison, A. J. and J. E. Scott, 1986, "The Complete DNA Sequence of Varicella-Zoster Virus", *Journal of General Virology*, Vol. 67, pp. 1759-1816.
- de Chassey, B., V. Navratil, L. Tafforeau, M. S. Hiet, A. Aublin-Gex, S. Agaue, G. Meiffren, F. Pradezynski, B. F. Faria, T. Chantier, M. Le Breton, J. Pellet, N. Davoust, P. E. Mangeot, A. Chaboud, F. Penin, Y. Jacob, P. O. Vidalain, M. Vidal, P. Andre, C. Rabourdin-Combe, V. Lotteau, 2008, "Hepatitis C Virus Infection Protein Network", *Molecular Systems Biology*, Vol.4, p.230.
- Dimitrova, M., I. Imbert, M. P. Kieny, C. Schuster, 2003, "Protein-Protein Interactions between Hepatitis C Virus Nonstructural Proteins", *Journal of Virology*, Vol. 77, pp. 5401-5414.
- Dolan, A., C. Addison, D. Gatherer, A. J. Davison, D. J. McGeoch, 2006, "The Genome of Epstein-Barr Virus Type 2 Strain AG876", *Virology*, Vol. 350, pp. 164-170.
- Doolittle, J. M. and S. M. Gomez, 2010, "Structural Similarity-Based Predictions of Protein Interactions between HIV-1 and *Homo Sapiens*", *Virology Journal*, Vol.7, p.82.

- Duigou, S., S. D. Ehrlich, P. Noirot, M. -F. Noirot-Gros, 2005, "DNA Polymerase I Acts in Translesion Synthesis Mediated by the Y-Polymerases in *Bacillus Subtilis*", *Molecular Microbiology*, Vol. 57, pp. 678-690.
- Durmuş Tekir, S., T. Çakır, K. Ö. Ülgen, 2012, "Infection Strategies of Bacterial and Viral Pathogens through Pathogen-Human Protein-Protein Interactions", *Frontiers in Microbiology*, Vol.3, p.46.
- Durmuş Tekir, S. and K. Ö. Ülgen, 2013a, "System Biology of Pathogen-Host Interaction: Networks of Protein-Protein Interaction within Pathogens and Pathogen-Human Interactions in the Post-Genomic Era", *Biotechnology Journal*, Vol. 8, pp. 85-96.
- Durmuş Tekir, S., T. Çakır, E. Ardiç, A. S. Sayılırbaş, G. Konuk, M. Konuk, H. Sarıyer, A. Uğurlu, İ. Karadeniz, A. Özgür, F. E. Sevilgen, K. Ö. Ülgen, 2013b, "PHISTO: Pathogen-Host Interaction Search Tool", *Bioinformatics*, Vol. 29, pp. 1357-1358.
- Dyer, M. D., T. M. Murali, B. W. Sobral, 2007, "Computational Prediction of Host-Pathogen Protein-Protein Interactions", *Bioinformatics*, Vol. 23, pp. i159-i166.
- Dyer, M. D., T. M. Murali, B. W. Sobral, 2008, "The Landscape of Human Proteins Interacting with Viruses and Other Pathogens", *PLoS Pathogens*, Vol. 4, e32.
- Dyer, M. D., C. Neff, M. Dufford, C. G. Rivera, D. Shattuck, J. Bassaganya-Riera, T. M. Murali, B. W. Sobral, 2010, "The Human-Bacterial Pathogen Protein Interaction Networks of *Bacillus Anthracis*, *Francisella Tularensis*, and *Yersinia Pestis*", *PLoS One*, Vol. 5, e12089.
- Egli, A., L. Infanti, A. Dumoulin, A. Buser, J. Samaridis, C. Stebler, R. Gosert, H. H. Hirsch, 2009, "Prevalence of Polyomavirus BK and JC Infection and Replication in 400 Healthy Blood Donors", *The Journal of Infectious Diseases*, Vol. 199, pp. 837-846.
- Eigen, M., 1993, "Viral Quasispecies", *Scientific American*, Vol. 269, pp. 42-49.

- Eisenberg, D., E. M. Marcotte, I. Xenarios, T. O. Yeates, 2000, "Protein Function in the Post-Genomic Era", *Nature*, Vol. 405, pp. 823–826.
- Evans, P., W. Dampier, L. Ungar, A. Tözeren, 2009, "Prediction of HIV-1 Virus-Host Protein Interactions Using Virus and Host Sequence Motifs", *BMC Medical Genomics*, Vol. 2, p. 27.
- Fauci, A. S., 2001, "Infectious Diseases: for Considerations the 21st Century" *Clinical Infectious Diseases*, Vol. 32, pp. 675-685.
- Fields, S. and O. Song, 1989, "A Novel Genetic System to Detect Protein-Protein Interactions", *Nature*, Vol. 340, pp. 245-246.
- Filippova, M., L. Parkhurst, P. J. Duerksen-Hughes, 2004, "The Human Papillomavirus 16 E6 Protein Binds to Fas-Associated Death Domain and Protects Cells from Fas-Triggered Apoptosis", *The Journal of Biological Chemistry*, Vol. 279, pp. 25729-25744.
- Finley, R. L. and R. Brent, 1994, "Interaction Mating Reveals Binary and Ternary Connections between Drosophila Cell Cycle Regulators", *Proceedings of the National Academy of Sciences*, Vol. 91, pp. 12980-12984.
- Flajolet, M., G. Rotondo, L. Daviet, F. Bergametti, G. Inchauspe, P. Tiollais, C. Transy, P. Legrain, 2000, "A Genomic Approach of the Hepatitis C Virus Generates a Protein Interaction Map", *Gene*, Vol. 242, pp. 369-379.
- Flannick, J., A. Novak, B. S. Srinivasan, H. H. Mcadams, S. Batzoglou, 2006, "Græmlin: General and Robust Alignment of Multiple Large Interaction Networks", *Genome Research*, Vol. 16, pp. 1169-1181.
- Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, 1995, "Whole-

- Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd", *Science*, Vol. 269, pp. 496-512.
- Forst, C. V., 2006, "Host-Pathogen Systems Biology", *Drug Discovery Today*, Vol. 11, pp. 220-227.
- Fraser, C. M., 1998, "Complete Genome Sequence of *Treponema Pallidum*, the Syphilis Spirochete", *Science*, Vol. 281, pp. 375-388.
- Fromont-Racine, M., J. Rain, P. Legrain, 1997, "Toward a Functional Analysis of the Yeast Genome through Exhaustive Two-Hybrid Screens", *Nature Genetics*, Vol. 16, pp. 277-282.
- Galinski, M. S., "Paramyxoviridae: Transcription and Replication", 1991, *Advances in Virus Research*, Vol.39, pp. 129-163.
- Gandhi, T., J. Zhong, S. Mathivanan, L. Karthick, 2006, "Analysis of the Human Protein Interactome and Comparison with Yeast, Worm and Fly Interaction Datasets", *Nature Genetics*, Vol. 38, pp. 285-293.
- Gavin, A.-C., M. Bösch, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Höfert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, G. Superti-Furga, 2002, "Functional Organization of the Yeast Proteome by Systematic Analysis of Protein Complexes", *Nature*, Vol. 415, pp. 141-147.
- Gillespie, J. J., A. R. Wattam, S. A. Cammer, J. L. Gabbard, M. P. Shukla, O. Dalay, T. Driscoll, D. Hix, S. P. Mane, C. Mao, E. K. Nordberg, M. Scott, J. R. Schulman, E. E. Snyder, D. E. Sullivan, C. Wang, A. Warren, K. P. Williams, T. Xue, H. S. Yoo, C. Zhang, Y. Zhang, R. Will, R. W. Kenyon, B. W. Sobral, 2011, "PATRIC: The

- Comprehensive Bacterial Bioinformatics Resource with a Focus on Human Pathogenic Species", *Infection and Immunity*, Vol. 79, pp. 4286-4298.
- Goebel, S. J., G. P. Johnson, M. E. Perkus, S. W. Davis, J. P. Winslow, E. Paoletti, 1990, "The Complete DNA Sequence of Vaccinia Virus", *Virology*, Vol. 179, pp. 517-563.
- Goff, S. P., 2007, "Host Factors Exploited by Retroviruses", *Nature*, Vol. 5, pp. 253-263.
- Goll, J., S. V. Rajagopala, S. C. Shiau, H. Wu, B. T. Lamb, P. Uetz, 2008 "MPIDB: The Microbial Protein Interaction Database", *Bioinformatics*, Vol. 24, pp. 1743-1744.
- Guiguemde, W. A., A. A. Shelat, D. Bouck, S. Duffy, G. J. Crowther, P. H. Davis, D. C. Smithson, M. Connelly, J. Clark, F. Zhu, M. B. Jimenez-Diaz, M. S. Martinez, E. B. Wilson, A. K. Tripathi, J. Gut, E. R. Sharlow, I. Bathurst, F. E. Mazouni, J. W. Fowble, I. Forquer, P. L. McGinley, S. Castro, I. Angulo-Barturen, S. Ferrer, P. J. Rosenthal, J. L. DeRisi, D. J. Sullivan, J. S. Lazo, D. S. Roos, M. K. Riscoe, M. A. Phillips, P. K. Rathod, W. C. Van Voorhis, V. M. Avery, R. K. Guy, 2010, "Chemical Genetics of *Plasmodium falciparum*", *Nature*, Vol. 465, pp. 311-315.
- Güldener, U., M. Münsterkötter, M. Oesterheld, P. Pagel, A. Ruepp, H. W. Mewes, V. Stümpflen, 2006, "MPact: The MIPS Protein Interaction Resource on Yeast", *Nucleic Acids Research*, Vol. 34, pp. D436-D441.
- Gully, D., 2003, "New Partners of Acyl Carrier Protein Detected in *Escherichia coli* by Tandem Affinity Purification", *FEBS Letters*, Vol. 548, pp. 90-96.
- Hale, B. G., R. E. Randall, J. Ortín, D. Jackson, 2008, "The Multifunctional NS1 Protein of Influenza A Viruses", *Journal of General Biology*, Vol. 89, pp. 2359-2376.
- Hay, A., V. Gregory, A. Douglas, Y. Lin, 2001 "The Evolution of Human Influenza Viruses", *Philosophical Transactions of the Royal Society of London*, Vol. 356, pp. 1861-1870.

- He, B., X. Qiao, P. J. Klasse, A. Chiu, A. Chadburn, D. M. Knowles, J. P. Moore, A. Cerutti, 2006, "HIV-1 Envelope Triggers Polyclonal Ig Class Switch Recombination through a CD40-Independent Mechanism Involving BAFF and C-Type Lectin Receptors", *The Journal of Immunology*, Vol. 176, pp. 3931-3941.
- Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkel, B. -C. Li, R. Herrmann, 1996, "Complete Sequence Analysis of the Genome of the Bacterium *Mycoplasma pneumoniae*", *Nucleic Acids Research*, Vol. 24, pp. 4420-4449.
- Ho, Y., A. Gruhler, A. Heilbut, G. Bader, L. Moore, 2002, "Systematic Identification of Protein Complexes in *Saccharomyces cerevisiae* by Mass Spectrometry", *Nature*, Vol. 415, pp. 2-5.
- Houghton, M., M. Selby, A. Weiner, Q. L. Choo, 1994, "Hepatitis C virus: Structure, Protein Products and Processing of the Polyprotein Precursor", *Current Studies in Hematology and Blood Transfusion*, Vol. 61, pp. 1-11.
- Hu, P., S. C. Janga, M. Babu, J. J. Diaz-Mejia, G. Butland, W. Yang, O. Pogoutse, X. Guo, S. Phanse, P. Wong, S. Chandran, C. Christopoulos, A. Nazarians-Armavil, N. K. Nasser, G. Musso, M. Ali, N. Nazemof, V. Eroukova, A. Golshani, A. Paccanaro, J. F. Greenblatt, G. Moreno-Hagelsieb, A. Emili, 2009, "Global Functional Atlas of *Escherichia coli* Encompassing Previously Uncharacterized Proteins", *PLoS Biology* Vol. 7, e96.
- Ito, T., K. Tashiro, S. Muta, R. Ozawa, T. Chiba, M. Nishizawa, K. Yamamoto, S. Kuhara, Y. Sakaki, 2000, "Toward a Protein-Protein Interaction Map of the Budding Yeast: A Comprehensive System to Examine Two-Hybrid Interactions in All Possible Combinations between the Yeast Proteins", *Proceedings of the National Academy of Sciences*, Vol. 97, pp. 1143-1147.
- Itzhaki, Z., 2011, "Domain-Domain Interactions Underlying Herpesvirus Human Protein-Protein Interaction Networks", *PLoS ONE*, Vol. 6, e21724.

- Jager, S., P. Cimermancic, N. Gülbahçe, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H. Hernandez, G. M. Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D'Orso, J. Fernandes, M. Fahey, C. Mahon, A. J. O'Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, J. A. Young, S. K. Chanda, W. I. Sundquist, T. Kortemme, R. D. Hernandez, C. S. Craik, A. Burlingame, A. Sali, A. D. Frankel, N. J. Krogan, 2012, "Global Landscape of HIV-Human Protein Complexes", *Nature*, Vol. 481, pp. 365-370.
- Karlas, A., N. Machuy, Y. Shin, K. P. Pleissner, 2010, "Genome-Wide RNAi Screen Identifies Human Host Factors Crucial for Influenza Virus Replication", *Nature*, Vol. 463, pp. 818-822.
- Kaufmann, S. H. E., 1993, "Immunity to Intracellular Bacteria" *Annual Review of Immunology*, Vol. 11, pp. 129-163.
- Kerrien, S., S. Orchard, L. Montecchi-Palazzi, B. Aranda, A. F. Quinn, N. Vinod, G. D. Bader, I. Xenarios, J. Wojcik, D. Sherman, M. Tyers, J. J. Salama, S. Moore, A. Ceol, A. Chatr-aryamontri, M. Oesterheld, V. Stümpflen, L. Salwinski, J. Nerothin, E. Cerami, M. E. Cusick, M. Vidal, M. Gilson, J. Armstrong, P. Woollard, C. Hogue, D. Eisenberg, G. Cesareni, R. Apweiler, H. Hermjakob, 2007, "Broadening the Horizon - Level 2.5 of the HUPO-PSI Format for Molecular Interactions", *BMC Biology*, Vol. 5, p.44.
- Kerrien, S., B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, C. Jandrasits, R. C. Jimenez, J. Khadake, U. Mahadevan, P. Masson, I. Pedruzzi, E. Pfeifferberger, P. Porras, A. Raghunath, B. Roechert, S. Orchard, H. Hermjakob, 2012, "The IntAct Molecular Interaction Database in 2012", *Nucleic Acids Research*, Vol. 40, pp. D841-D846.
- Kiley, M. P., E. T. W. Bowen, G. A. Eddy, M. Isaacson, K. M. Johnson, J. B. McCormick, F. A. Murphy, S. R. Pattyn, D. Peters, O. W. Prozesky, R. L. Regnery, D. I. H.,

- Simpson, W. Slenczka, P. Sureau, G. van der Groen, P. A. Webb, H. Wulff, 1982, "Filoviridae: A Taxonomic Home for Marburg and Ebola Viruses?" *Intervirology*, Vol. 18, pp. 24-32.
- Klase, Z., R. Winograd, J. Davis, L. Carpio¹, R. Hildreth, M. Heydarian, S. Fu, T. McCaffrey, E. Meiri, M. Ayash-Rashkovsky, S. Gilad, Z. Bentwich, F. Kashanchi, 2009, "HIV-1 TAR miRNA Protects Against Apoptosis by Altering Cellular Gene Expression", *Retrovirology*, Vol. 6, p. 18.
- Knipe, D. M. and P. M. Howley, 2007, *Flaviviridae: The Viruses and Their Replication*, Lippincott–Raven Publishers, Philadelphia.
- König, R., Y. Zhou, D. Elleder, T. Diamond, 2008, "Global Analysis of Host Pathogen Interactions that Regulate Early-Stage HIV-1 Replication", *Cell*, Vol. 135, pp. 49-60.
- König, R., S. Stertz, Y. Zhou, A. Inoue, H. H. Hoffmann, S. Bhattacharyya, J. G. Alamares, D. M. Tscherne, M. B. Ortigoza, Y. Liang, Q. Gao, S. E. Andrews, S. Bandyopadhyay, P. De Jesus, B. P. Tu, L. Pache, C. Shih, A. Orth, G. Bonamy, L. Miraglia, T. Ideker, A. Garcia-Sastre, J. A. Young, P. Palese, M. L. Shaw, S. K. Chanda, 2010, "Human Host Factors Required for Influenza Virus Replication", *Nature*, Vol. 463, pp. 813-817.
- Korkin, D., T. Thieu, S. Joshi, 2006, "Mining Host-Pathogen Interactions", In: Yang, N.-S. (Ed.), *Systems and Computational Biology - Molecular and Cellular Experimental Systems*, InTech, Rijeka, pp. 163–184.
- Krishnadev O. and N. Srinivasan, 2008, "A Data Integration Approach to Predict Host-Pathogen Protein-Protein Interactions: Application to Recognize Protein Interactions between Human and a Malarial Parasite" *In Silico Biology*, Vol. 8, pp. 235–250.
- Krishnan, M., A. Ng, B. Sukumaran, F. Gilfoy, 2008, "RNA Interference Screen for Human Genes Associated with West Nile Virus Infection", *Nature*, Vol. 455, pp. 242-245.

- Kühner, S., V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castano-Diez, W. -H. Chen, D. Devos, M. Güell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Böttcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork, A. -C. Gavin, 2009, "Proteome Organization in a Genome-Reduced Bacterium", *Science*, Vol. 326, pp. 1235-1240.
- Kumar, D., L. Nath, A. Kamal, A. Varshney, A. Jain, S. Singh, K. V. S. Raol, 2010, "Genome-Wide Analysis of the Host Intracellular Network that Regulates Survival of *Mycobacterium tuberculosis*", *Cell*, Vol. 140, pp. 731–743.
- Kumar, R. and B. Nanduri, 2010, "HPIDB - A Unified Resource for Host-Pathogen Interactions", *BMC Bioinformatics*, Vol. 11, S16.
- LaCount, D. J., M. Vignali, R. Chettier, A. Phansalkar, R. Bell, J. R. Hesselberth, L. W. Schoenfeld, I. Ota, S. Sahasrabudhe, C. Kurschner, S. Fields, R. E. Hughes, 2005, "A Protein Interaction Network of the Malaria Parasite *Plasmodium falciparum*", *Nature*, Vol. 438, pp. 103-107.
- Lai, X. and I. Golovliov, 2001, "*Francisella tularensis* Induces Cytopathogenicity and Apoptosis in Murine Macrophages via a Mechanism that Requires Intracellular Bacterial Multiplication", *Infection and Immunity*, Vol. 69, pp. 4691-4694.
- Lee, S. -A., C. Chan, C. -H Tsai, J. -M. Lai, F. -S. Wang, C. -Y Kao, C. -Y. F. Huang, 2008, "Ortholog-Based Protein-Protein Interaction Prediction and its Application to Interspecies Interactions", *BMC Bioinformatics*, Vol. 9, S11.
- Lechner, M. S. and L. A. Laimins, 1994, "Inhibition of p53 DNA Binding by Human Papillomavirus E6 Proteins", *Journal of Virology*, Vol. 68, pp. 4262-4273.
- Lehner, B. and A. G. Fraser, 2004, "A First-Draft Human Protein-Interaction Map", *Genome Biology*, Vol. 5, R63.

- Lengeling, A., K. Pfeffer, R. Balling, 2001, "The Battle of Two Genomes: Genetics of Bacterial Host/Pathogen Interactions in Mice", *Mammalian Genome*, Vol. 12, pp. 261-271.
- Li, D., J. Li, S. Ouyang, J. Wang, S. Wu, P. Wan, Y. Zhu, X. Xu, F. He, 2006, "Protein Interaction Networks of *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster*: Large-Scale Organization and Robustness", *Proteomics*, Vol. 6, pp. 456–461.
- Li, Q., A. L. Brass, A. Ng, Z. Hu, R. J. Xavier, T. J. Liang, S. J. Elledge, 2009, "A Genome-Wide Genetic Screen for Host Factors Required for Hepatitis C Virus Propagation", *Proceedings of the National Academy of Sciences*, Vol. 106, pp. 16410-16415.
- Licata, L., L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galetto, F. Sacco, A. Palma, A. P. Nardoza, e. Santonico, L. Castagnoli, G. Cesareni, 2012, "MINT, the Molecular Interaction Database: 2012 update", *Nucleic Acids Research*, Vol. 40, pp. D857-D861.
- Maere, S., K. Heymans, M. Kuiper, 2005, "BiNGO: A Cytoscape Plugin to Assess over Representation of Gene Ontology Categories in Biological Networks", *Bioinformatics*, Vol. 21, pp. 3448-3449.
- Marchadier, E., R. Carballido-Lopez, S. Brinster, C. Fabret, P. Mervelet, P Bessie`res, M. - F. Noirot-Gros, V. Fromion, P. Noirot, 2011, "An Expanded Protein-Protein Interaction Network in *Bacillus Subtilis* Reveals A Group of Hubs: Exploration by an Integrative Approach", *Proteomics*, Vol. 11, pp. 2981-2991.
- Martin, M. A., D. M. Knipe, B. N. Fields, P. M. Howley, D. Griffin, R. Lamb, 2007, "Fields' Virology", Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, p. 2395.

- Maslov, S. and K. Sneppen, 2002, "Specificity and Stability in Topology of Protein Networks", *Science*, Vol. 296, pp. 910-913.
- McCraith, S., T. Holtzman, B. Moss, S. Fields, 2000, "Genome-Wide Analysis of Vaccinia Virus Protein-Protein Interactions", *Proceedings of the National Academy of Sciences*, Vol. 97, pp. 4879-4884.
- Mogensen, T. H., S. R. Paludan, M. Kilian, L. Østergaard, 2006, "Live *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis* Activate the Inflammatory Response through Toll-Like Receptors 2, 4, And 9 in Species-Specific Patterns", *Journal of Leukocyte Biology*, Vol. 80, pp. 267-277.
- Moss, B., 1996, "Genetically Engineered Poxviruses for Recombinant Gene Expression, Vaccination, and Safety", *Proceedings of the National Academy of Sciences*, Vol. 93, pp. 11341-11348.
- Munoz, N., X. Castellsague, A. Berrington de Gonzalez, L. Gissmann, 2006 "Chapter 1: HPV in the Etiology of Human Cancer", *Vaccine*, Vol. 24, pp. S1–S10.
- Münter, S., M. Way, F. Frischknecht, 2006, "Signaling During Pathogen Infection", *Science's STKE*, Vol. 335, re5.
- Murali, T. M., M. D. Dyer, D. Badger, B. M. Tyler, M. G. Katze, 2011, "Network-Based Prediction and Analysis of HIV Dependency Factors", *PLoS Computational Biology*, Vol. 7, e1002164.
- Navratil, V., B. de Chassey, L. Meyniel, S. Delmotte, C. Gautier, P. Andre, V. Lotteau, C. Roubardin-Combe, 2009, "VirHost-Net: A Knowledge Base for the Management and the Analysis of Proteome-Wide Virus-Host Interaction Networks", *Nucleic Acids Research*, Vol. 37, pp. D661-D668.

- Nisole, S., B. Krust, A. G. Hovanessian, 2002, "Anchorage of HIV on Permissive Cells Leads to Coaggregation of Viral Particles with Surface Nucleolin at Membrane Raft Microdomains", *Experimental Cell Research*, Vol. 276, pp. 155-173.
- Noirot-Gros, M. -F., E. Dervyn, L. J. Wu, P. Mervelet, J. Errington, S. D. Ehrlich, P. Noirot, 2002, "An Expanded View of Bacterial DNA Replication", *Proceedings of the National Academy of Sciences*, Vol. 99, pp. 8342-8347.
- Oda, K. and H. Kitano, 2006, "A Comprehensive Map of the Toll-Like Receptor Signaling Network", *Molecular Systems Biology*, Vol. 2, 2006.0015.
- Pan, J., X. Peng, Y. Gao, Z. Li, X. Lu, Y. Chen, M. Ishag, D. Liu, M. L. Dediego, L. Enjuanes, D. Guo, 2008, "Genome-Wide Analysis of Protein-Protein Interactions and Involvement of Viral Proteins in SARSCoV Replication", *PLoS One*, Vol. 3, e3299.
- Park, J. M., F. R. Greten, Z. -W Li, M. Karin, 2002, "Macrophage Apoptosis by Anthrax Lethal Factor through P38 MAP Kinase Inhibition", *Science*, Vol. 297, pp. 2048-2051.
- Parrish, J. R., J. Yu, G. Liu, J. A. Hines, J. E. Chan, B. A. Mangiola, H. Zhang, S. Pacifico, F. Fotouhi, V. J. DiRita, T. Ideker, P. Andrews, R. L. Finley, 2007, "A Proteome-Wide Protein Interaction Map for *Campylobacter jejuni*", *Genome Biology*, Vol. 8, R130.
- Pichlmair, A., K. Kandasamy, G. Alvisi, O. Mulhern, R. Sacco, M. Habjan, M. Binder, A. Stefanovic, C. -A. Eberle, A. Goncalves, T. Bürckstümmer, A. C. Müller, A. Fauster, C. Holze, K. Lindsten, S. Goodbourn, G. Kochs, F. Weber, R. Bartenschlager, A. G. Bowie, K. L. Bennett, J. Colinge, G. Superti-Furga, 2012, "Viral Immune Modulators Perturb the Human Molecular Network by Common and Unique Strategies", *Nature*, Vol. 487, pp. 486-490.

- Prasad, T. S. K., R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. H. Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, A. Pandey, 2009, "Human Protein Reference Database", *Nucleic Acids Research*, Vol. 37, pp. D767-D772.
- Prieto, C. and J. D. L. Rivas, 2006, "APID: Agile Protein Interaction Data Analyzer" *Nucleic Acids Research*, Vol. 34, pp. W298-W302.
- Przulj, N., D. A. Wigle, I. Jurisica, 2004, "Functional Topology in a Network of Protein Interactions", *Bioinformatics*, Vol. 20, pp. 340-348.
- Rain, J. -C., L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, g. Lenzen, F. Petel, J. Wojcik, V. Schachter, Y. Chemama, A. Labigne, P. Legrain, 2001, "The Protein-Protein Interaction Map of *Helicobacter pylori*", *Nature*, Vol. 409, pp. 6817, 211–215.
- Raman, K. and N. Chandra, 2008, "*Mycobacterium tuberculosis* Interactome Analysis Unravels Potential Pathways to Drug Resistance", *BMC Microbiology*, Vol. 8, p. 234.
- Razick, S., G. Magklaras, I. M. Donaldson, 2008, "iRefIndex: A Consolidated Protein Interaction Database with Provenance", *BMC Bioinformatics*, Vol. 9, p. 405.
- Rual, J. -F., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G. F. Berriz, F. D. Gibbons, M. Dreze, N. Ayivi-Guedehoussou, N. Klitgord, C. Simon, M. Boxern, S. Milstein, J. Rosenberg, D. S. Goldberg, L. V. Zhang, S. L. Wong, G. Franklin, S. Li, J. S. Albala, J. Lim, C. Fraughton, E. Liamosas, S. Cevik, C. Bex, P. Lamesch, R. S. Sikorski, J. Vandenhaute, H. Y. Zoghbi, A. Smolyar, S. Bosak, R. Sequerra, L. Doucette-Stamm, M. E. Cusick, D. E. Hill, F. P. Roth, M. Vidal, 2005,

- "Towards a Proteome-Scale Map of the Human Protein-Protein Interaction Network", *Nature*, Vol. 437, pp. 1173-1178.
- Salwinski, L., C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, D. Eisenberg., 2004, "The Database of Interacting Proteins: 2004 update", *Nucleic Acids Research*, Vol. 32, pp. D449-D451.
- Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, J. C. Fiddes, C. A. Hutchison, P. M. Slocombe, M. Smith, 1977, "Nucleotide Sequence of Bacteriophage ϕ X174 DNA", *Nature*, Vol. 265, pp. 687-695.
- Schleker, S. and Trilling, M., 2013, "Data-Warehousing of Protein-Protein Interactions Indicates that Pathogens Preferentially Target Hub and Bottleneck Proteins", *Frontiers in Microbiology*, Vol. 4, p. 51.
- Shapira, S., I. Gat-Viks, B. Shum, A. Dricot, 2009, "A Physical and Regulatory Map of Host-Influenza Interactions Reveals Pathways in H1N1 Infection", *Cell*, Vol. 139, pp. 1255-1267.
- Sharan, R., S. Suthram, R. M. Kelley, T. Kuhn, S. McCuine, P. Uetz, T. Sittler, R. M. Karp, T. Ideker, 2005, "Conserved Patterns of Protein Interaction in Multiple Species", *Proceedings of the National Academy of Sciences*, Vol. 102, pp. 1974-1979.
- Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. Spaan, A. E. Gorbalenya, 2003, "Unique and Conserved Features of Genome and Proteome of SARS-Coronavirus, an Early Split-Off from the Coronavirus Group 2 Lineage" *Journal of Molecular Biology*, Vol. 331, pp. 991-1004.
- Soufo, C. D., H. J. D. Soufo, M. -F., Noirot-Gros, A. Steindorf, P. Noirot, P. L. Graumann, 2008, "Cell-Cycle-Dependent Spatial Sequestration of the DnaA Replication Initiator Protein in *Bacillus subtilis*" *Developmental Cell*, Vol. 15, pp. 935-941.

- Stark, C., B. J. Breitkreutz, A. Chatr-Aryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Regul, J. M. Rust, A. Winter, K. Dolinski, M. Tyers, 2011, "The BioGRID Interaction Database: 2011 Update", *Nucleic Acids Research*, Vol. 39, pp. D698-D704.
- Stebbins, C. E., 2005, "Structural Microbiology at the Pathogen-Host Interface", *Cellular Microbiology*, Vol. 7, pp. 1227-1236.
- Stelzl, U., U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksöz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, E. E. Wanker, 2005, "A Human Protein Protein Interaction Network: A Resource for Annotating the Proteome", *Cell*, Vol. 122, pp. 957-968.
- Stumptner-Cuvelette, P., S. Morchoisne, M. Dugast, S. L. Gall, G. Raposo, O. Schwartz, P. Benaroch, 2001, "HIV-1 Nef Impairs MHC Class II Antigen Presentation and Surface Expression", *Proceedings of the National Academy of Sciences*, Vol. 98, pp. 12144-12149.
- Szklarczyk, D., A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguéz, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen, C. von Mering, 2011, "The STRING Database in 2011: Functional Interaction Networks of Proteins Globally Integrated and Scored", *Nucleic Acids Research*, Vol. 39, pp. D561-D568.
- Tastan, O., Y. Qi, J. G. Carbonell, J. Klein-Seetharaman, 2009, "Prediction of Interactions between HIV-1 and Human Proteins by Information Integration", *Pacific Symposium on Biocomputing*, Vol. 527, pp. 516-527.
- Taylor, D. J., R. W. Leach, J. Bruenn, 2010, "Filoviruses are Ancient and Integrated into Mammalian Genomes", *BMC Evolutionary Biology*, Vol. 10, p. 193.

- Thieu, T., S. Joshi, S. Warren, D. Korkin, 2012, "Literature Mining of Host Pathogen Interactions: Comparing Feature-Based Supervised Learning and Language-Based Approaches", *Bioinformatics*, Vol. 28, pp. 867-875.
- Thomas, M., P. Massimi, C. Navarro, J. P. Borg, L. Banks, 2005, "The hScrib/Dlg Apico-Basal Control Complex is Differentially Targeted by HPV-16 and HPV-18 E6 Proteins", *Oncogene*, Vol. 24, pp. 6222-6230.
- Thompson, D. A., G. Belinsky, T. H. -T Chang, D. L. Jones, R. Schlegel, K. Munger, 1997, "The Human Papillomavirus-16 E6 Onco-Protein Decreases the Vigilance of Mitotic Checkpoints", *Oncogene*, Vol. 15, pp. 3025-3035.
- Titz, B., S. V. Rajagopala, J. Goll, R. Hauser, M. T. McKevitt, T. Palzkill, P. Uetz, 2008, "The Binary Protein Interactome of *Treponema pallidum* - The Syphilis Spirochete", *PLoS One*, Vol. 3, e2292.
- Tuma, P., P. Rivas, E. Vispo, P. Barreiro, S. Puente, J. Medrano, A. Madejon, D. M. Herrero, P. Labarga, J. Garcia-Samaniego, V. Soriano, 2009, "New Drug Targets for Hepatitis C and Other Flaviviridae Viruses". *Infectious Disorders-Drug Targets*, Vol. 9, pp. 133-47.
- Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, J. M. Rothberg, 2000, "A Comprehensive Analysis of Protein-Protein Interactions in *Saccharomyces cerevisiae*", *Nature*, Vol. 403, pp. 623-627.
- Uetz, P., Y. A. Dong, C. Zeretzke, C. Atzler, A. Baiker, B. Berger, S. V. Rajagopala, M. Roupelieva, D. Rose, E. Fossum, J. Haas, 2006, "Herpesviral Protein Networks and Their Interaction with the Human Proteome", *Science*, Vol. 311, pp. 239-242.

Van Etten, J. L., L. C. Lane, D. D. Dunigan, "DNA Viruses: The Really Big Ones (Giruses)", 2010, *Annual Review of Microbiology*, Vol. 64, pp. 83-99.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. G. Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. -R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Y. Wang, A. Wang, X. Wang, J. Wang, M. -H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. -H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A.

- Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. -H. Chiang, M. Coyne, C. Dahlke, A. D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, X. Zhu, 2001, "The Sequence of the Human Genome", *Science*, Vol. 291, pp. 1304-1351.
- Vidalain, P. -O. and F. Tangy, 2010, "Virus-Host Interactions in RNA Viruses", *Microbes and Infection*, Vol. 12, pp. 1134-1143.
- von Brunn, A., C. Teepe, J. C. Simpson, R. Pepperkok, C. C. Friedel, R. Zimmer, R. Roberts, R. Baric, J. Haas, 2007, "Analysis of Intraviral Protein-Protein Interactions of the SARS Coronavirus ORFeome", *PLoS One*, Vol. 2, e459.
- Walhout, J. M. A., R. Sordella, X. Lu, J. L. Hartley, G. F. Temple, M. A. Brasch, N. Thierry-Mieg, M. Vidal, 2000, "Protein Interaction Mapping in *C. Elegans* Using Proteins Involved in Vulval Development", *Science*, Vol. 87, pp. 116-122.
- Wang, Y., T. Cui, C. Zhang, M. Yang, Y. Huang, W. Li, L. Zhang, C. Gao, Y. He, Y. Li, F. Huang, J. Zeng, C. Huang, Q. Yang, Y. Tian, C. Zhao, H. Chen, H. Zhang, Z. G. He, 2010, "Global Protein-Protein Interaction Network in the Human Pathogen *Mycobacterium tuberculosis* H37Rv", *Journal of Proteome Research*, Vol. 9, pp. 6665-6677.
- Watanabe, T., S. Watanabe, Y. Kawaoka, 2010, "Cellular Networks Involved in the Influenza Virus Life Cycle", *Cell Host Microbe*, Vol. 7, pp. 427-439.
- Weiss, R. A., 1993, "How Does HIV Cause AIDS?", *Science*, Vol. 260, pp. 1273-1279.

- Winnenburg, R., M. Urban, A. Beacham, T. K. Baldwin, S. Holland, M. Lindeberg, H. Hansen, C. Rawlings, K. E. Hammond-Kosack, J. Köhler, 2008, "PHI-base Update: Additions to the Pathogen Host Interaction Database", *Nucleic Acids Research*, Vol. 36, pp. D572-D576.
- World Health Organization, 2009, "Global Tuberculosis Control 2009: Epidemiology, Strategy, Financing", WHO Press, Geneva.
- World Health Organization, 2012, "World Health Statistics-2012", WHO Press, Geneva.
- Xiang, Z., Y. Tian, Y. He, 2007, "PHIDIAS: A Pathogen-Host Interaction Data Integration and Analysis System", *Genome Biology*, Vol. 8, R150.
- Yang, H., Y. Ke, J. Wang, Y. Tan, S. K. Myeni, D. Li, Q. Shi, Y. Yan, H. Chen, Z. Guo, Y. Yuan, X. Yang, R. Yang, Z. Du, 2011, "Insight into Bacterial Virulence Mechanisms against Host Immune Response via the *Yersinia Pestis*-Human Protein-Protein Interaction Network", *Infection and Immunity*, Vol. 79, pp. 4413-4424.
- Yedavalli, V. S., H. M. Shih, Y. P. Chiang, C. Y. Lu, L. Y. Chang, M. Y. Chen, C. Y. Chuang, A. I. Dayton, K. T. Jeang, L. M. Huang, 2005, "Human Immunodeficiency Virus Type 1 Vpr Interacts with Antiapoptotic Mitochondrial Protein HAX-1", *Journal of Virology*, Vol. 79, pp. 13735-13746.
- Yellaboina, S. and K. Goyal, 2007, "Inferring Genome-Wide Functional Linkages in *E. Coli* by Combining Improved Genome Context Methods: Comparison with High-Throughput Experimental Data", *Genome Research*, Vol. 17, pp. 527-535.
- Yin, L., G. Xu, M. Torii, Z. Niu, J. Maisog, 2010, "Document Classification for Mining Host Pathogen Protein-Protein Interactions", *Artificial Intelligence in Medicine*, Vol. 49, pp. 155-160.

- Yu, H., P. M. Kim, E. Sprecher, V. Trifonov, M. Gerstein, 2007, "The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics", *PloS Computational Biology*, Vol. 3, e59.
- Zhang, G. and S. Ghosh, 2000, "Molecular Mechanisms of NF- κ B Activation Induced by Bacterial Lipopolysaccharide through Toll-Like Receptors", *Journal of Endotoxin Research*, Vol. 6, pp. 453-457.
- Zhang, Y., A. T. Ting, K. B. Marcu, B. James, J. B. Bliska, 2005, "Inhibition of MAPK and NF-KB Pathways is Necessary for Rapid Apoptosis in Macrophages Infected with Yersinia" *Journal of Immunology*, Vol. 174, pp. 7939-7949.
- Zhou, H., M. Xu, Q. Huang, A. T. Gates, X. D. Zhang, J. C. Castle, E. Stec, M. Ferrer, b. Strulovici, D. J. Hazuda, A. S. Espeseth, 2008, "Genome-Scale RNAi Screen for Host Factors Required for HIV Replication", *Cell Host & Microbe*, Vol. 4, pp. 495-504.
- Zhu, H., M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein, M. Snyder, 2001, "Global Analysis of Protein Activities using Proteome Chips", *Science*, Vol. 293, pp. 2101-2105.