



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Review Article

Computational analysis of interactomes: Current and future perspectives for bioinformatics approaches to model the host–pathogen interaction space

Roland Arnold^{a,1}, Kurt Boonen^{a,1}, Mark G.F. Sun^a, Philip M. Kim^{a,b,c,d,*}

^a Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada M5S 3E1

^b Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada M5S 3E1

^c Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada M5S 3E1

^d Department of Computer Science, University of Toronto, Toronto, ON, Canada M5S 3E1

ARTICLE INFO

Article history:

Available online 28 June 2012

Keywords:

Host–Pathogen
Protein Interaction
Computational Biology
Interaction Prediction

ABSTRACT

Bacterial and viral pathogens affect their eukaryotic host partly by interacting with proteins of the host cell. Hence, to investigate infection from a systems' perspective we need to construct complete and accurate host–pathogen protein–protein interaction networks. Because of the paucity of available data and the cost associated with experimental approaches, any construction and analysis of such a network in the near future has to rely on computational predictions. Specifically, this challenge consists of a number of sub-problems: First, prediction of possible pathogen interactors (e.g. effector proteins) is necessary for bacteria and protozoa. Second, the prospective host binding partners have to be determined and finally, the impact on the host cell analyzed. This review gives an overview of current bioinformatics approaches to obtain and understand host–pathogen interactions. As an application example of the methods covered, we predict host–pathogen interactions of *Salmonella* and discuss the value of these predictions as a prospective for further research.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Despite tremendous advances in biomedical science, infectious diseases are still a major health problem due to the rise of novel variants of pathogens that are resistant to available drugs just as the multi-resistant *Staphylococcus aureus* [1] or pathogenic *Escherichia coli* strains [2]. The approval of the oxazolidinone antibiotic Linezolid in 2000 was preceded by four decades without discovering any new structural class of antibacterial drugs [3], illustrating that there is an urgent need for new antiviral, antibacterial and antiprotozoal compounds. An understanding of host–pathogen interactions on the molecular level is crucial for the development of such therapeutics. While some experimental data for host–pathogen interactions exist (reviewed below), their relative paucity emphasizes the need for computational predictions. Modern technologies such as genome sequencing and high throughput methods for protein–protein interaction (PPI) detection provide a source of information that can be exploited by computational methods to provide short-cuts to laborious wet lab research [4]. These methods can support many different stages of knowledge generation, covering the generation of target lists for the investigation of

unknown virulence factors, prediction of the points of attack in the host cell and enlightenment of the modes of molecular interactions during the infection process.

On a molecular level, infection can be viewed as the interference of pathogenic proteins (most of the microbial toxins are proteins) with the hosts' interaction network. Viral proteins use the host protein synthesis machinery, whereas bacterial ones and proteins from parasitic protozoa are secreted or injected into host cells. Many pathogenic proteins follow general protein architectural rules and contain distinct functional and structural features, e.g., signal peptides that lead to targeted transport and pathogenic effector domains, which interact with the host system. A common pathogen strategy is to mimic eukaryotic PPI domains or binding motifs in order to remodel the host PPI network [5]. These structural mimics are often the result of horizontal gene transfer (HGT) [6] in combination with rapidly evolving pathogen genomes as part of host–pathogen coevolution [7]. Some well-known host–pathogen PPI interactions are mediated by host SH2, SH3 and PDZ domains. For example, the Influenza A non-structural protein 1 (NS1) contains regions that bind the SH2 and SH3 domains of the regulatory subunits of PI3K [8,9] and the class 1 PDZ-binding motifs of NS1 binds Scribble [10]. Moreover, these PPI motifs and domains can determine the virulence of the pathogen. For example, NS1 of avian and Spanish Influenza A has a SH3 domain-binding motif that preferably interacts with Crk-like protein, and this interaction increased Akt phosphorylation downstream [11]. Bacterial

* Corresponding author at: Department of Computer Science, University of Toronto, Toronto, ON, Canada M5S 3E1. Fax: +1 416 978 8287.

E-mail address: pi@kimlab.org (P.M. Kim).

¹ These authors contributed equally to this work.

molecular mimicries include the effector *espF* of the enteropathogenic *E. coli* which interacts with human sorting nexin 9 in order to achieve membrane re-modeling [12]. The effector *IpgB2* in *Shigella* as well as some of its homologs in other bacteria are functional mimics of Rho family GTPases and interfere with host signaling pathways [13], whereas the chlamydial *IncA* protein contains SNARE like motifs used to interfere with intra-cellular trafficking in order to establish the inclusion membrane of *Chlamydia* (an obligate intracellular parasite) [14]. An important feature of pathogenic proteins is their propensity to target proteins from pathways that aid infection and counteract the hosts defense systems (“anti-immunology”), e.g. by apoptosis evasion [15], cell cycle activation [16], proteasomal and ubiquitination alteration [17], actin and cytoskeleton reorganization [18] and innate immunity inhibition [19]. To conclude, interactions of pathogen proteins with host proteins are often mediated by domains and motifs and these interactions are not random but target specific proteins of the host. We can exploit this fact in computational approaches to detect host–pathogen interactions. The methods for doing so will be discussed in this review.

This review is structured as follows: First, we will summarize resources of host–pathogen interactions (HP-PPI) currently available and we will discuss text-mining methods that are capable of extracting large-scale interaction data from the literature. Second, we will review bioinformatics methods to predict bacterial effector proteins (i.e., proteins that are secreted by bacterial pathogens in order to interact with the host system). Third, we will discuss the prediction of HP-PPIs. We will focus on methods utilizing domain–domain interactions (DDIs) and short linear motifs (SLMs) since comprehensive databases of these types of interactions are available [20–23] and can be exploited for the prediction of HP-PPIs. Currently available HP-PPI networks are reviewed in the fourth part. The fifth part of the article provides an example analysis that shows how the discussed methods can contribute to the relatively under-explored field of host–pathogen interactions. Methods to increase the reliability and the pathogen class specificity of HP-PPI prediction methods are discussed last.

2. Databases and text-mining

Currently, there are several databases available that cover various aspects of HP-PPIs: broadly, these can be divided into databases of virulence related genes and of known HP-PPIs as detected by small or large scale screens. Some databases go beyond the scope of listing data and allow query-driven analyses by the user. We will also discuss HP-PPI specific text-mining approaches, as these approaches are valuable for mining HP-PPIs literature that is not contained in databases and can complement database searches.

2.1. Bacterial virulence related genes

For bacterial pathogens, the VFDB [24] and the MvirDB [25] databases comprise a collection of known bacterial virulence related genes such as toxins, resistance genes, and effector proteins which are delivered into the host cell by active transport. VFDB contains hand-curated data of experimentally validated virulence factors and covers currently data from 8 major bacterial pathogens. MvirDB is a meta-database that collects data from different sources including VFDB as well as databases dedicated to toxins, pathogenicity islands and antibiotic resistance and covers in addition to bacteria, eukaryotic and viral sequences. PATRIC (short for Pathosystems Resource Integration Center) is a meta-database that collects genomic data from bacterial and viral sequences and processes the data by a unified, automated annotation pipeline.

The resource is comprehensive in terms of the covered species (>5000 partially or completely sequenced bacteria, >2600 viral species with several thousand strains) and offers a variety of comparative analysis tools [26].

Several databases are dedicated to effector proteins secreted by the Type III secretion system and provide curated and predicted sets of bacterial effectors and cognate proteins as the Effective database [27], the T3SE database [28], and the T3DB database [29].

2.2. HP-PPI databases

PHI-base [30] is a database covering around 1800 verified bacterial, fungal, and oomycete host–pathogen interactions. HPIDB [31] is a meta-database that collects host–pathogen interactions from different interaction resources such as the DIP [32] or MINT [33] PPI collections, and from specialized resources as the PIG (pathogen interaction gateway) database [34]. The latter one collects host–pathogen PPI information between human and several viral and bacterial pathogens as well as additional data from GO annotations and Interpro assignments. It is integrated into the PATRIC resource [26]. iRefWeb is a very comprehensive meta-database that, despite not being especially designed for HP-PPIs, allows to search for interaction pairs measured between different species [35]. A number of databases are especially dedicated to protein–protein interactions between viral proteins and host proteins such as the Virusmint database [36]. This database consists of more than 5000 interactions as well as literature annotations. Virhostnet is a similar database, but contains, aside from host–pathogen interactions, the interactomes of the host (human) and of the respective viruses. CAPIH [37] is a database for analyzing HIV related interactions. In addition, it provides genetic variation data and predictions of phosphorylation and methylation sites. Another database, GPS-prot, is a meta-database that allows graphical browsing of HIV–human interaction networks [38]. Where some of the aforementioned databases just collect the interaction data-sets, others curate data by hand like PHI-base or PIG. A summary of these databases can be found in Table 1.

2.3. Integrated database and analysis systems

Several databases do not only list interaction data but also allow further exploration of the data. Most of the databases allow graphical exploration of the interaction networks (all mentioned databases except HPIDB). HPIDB allows BLASTing against the interaction sets implementing to detect homologs as well as conserved interacting pairs (Interologs, described below). CAPIH visualizes differences between orthologs of human, macaque, chimp, and mouse in terms of genetic variation and protein modifications. This information can potentially allow interpreting possible differences of HIV–human interactions in their model organism equivalents which can be very useful for translating animal based drug discovery to humans.

2.4. Text-mining

A promising approach to supplement the knowledge on host–pathogen interactions is automated mining of the literature (e.g. PubMed abstracts); such procedures can uncover functional relationships and interactions not deposited in the primary interaction databases and can aid the identification of relevant literature to improve the curation process of host–pathogen related databases. Many approaches exist to accomplish this for intra-species interactions, however, only quite recently the value of text-mining has been assessed for host–pathogen interactions: Yin and co-workers evaluated the usage of a certain group of classification algorithms, namely support vector machines (SVM) to retrieve documents

Table 1
Overview of HP-PPI databases.

Database	Species	Curated or meta-database	Special features	Size	Last update
PHI-base	97 pathogens, 76 host species	Yes	Search by phenotype possible	1335 interactions	2009
HPIDB	389 pathogens 58 host species	Meta-database	BLAST	22223 interactions	2012
PIG	21 bacteria in the host-pathogen section	Meta-database	BLAST & domain information, graphical representation of HP-PPI network	7180 interactions	2012
Virusmint	10 viral species, host: human	Yes	Graphical representation of HP-PPI network	5140 interactions	?
Virhostnet	180 viral species, host: human	Yes	–	2671 interactions	?
CAPIH	HIV/human	Meta-database	Resource for comparative analysis of HIV-host interactions with model organisms	1447 interactions	?
GPS-prot	HIV/human	Meta-database	Allows the integration of genetic interactions (RNAi screens)	?	2011

Host–Pathogen PPI databases. “Database”: the database name, “Species”: the amount and phyla of the species covered in the database, “Special features”: special technical aspects, “Size”: the amount of interactions, “Last update:” last update according to the prospective web-site.

related to host–pathogen interactions. They used cross-validation experiments to show that this can be accomplished with a rate of 50% relevant articles in the top 10% of the ranked documents, where a random guess would yield 10% [39]. In a different study, Thieu et al. [40] employed two approaches: An SVM based approach similar as the one used by Yin et al. and an approach based on semantic analysis, which allows the computational description of dependencies between words reflecting the logical structure of a sentence. In addition to Yin et al., who only discriminate abstracts which are relevant to host–pathogen interaction from abstracts of arbitrary articles, Thieu et al. formulate two additional problems: find relevant sentences that describe the mode of interaction, and find the actual interacting entities. It could be shown that the semantic approach has a higher performance on the latter two tasks compared to the SVM but still only received a precision (fraction of correct predictions) of 0.16 and a recall (fraction of found known positive instances) of 0.11 in the identification of actual HP-PPIs on the given test-set, compared to 0.03 and 0.05 in case of the SVM. Both studies have been performed on an initial, hand-curated set of articles and show the general feasibility of data mining in the domain of host–pathogen interactions. In theory, such algorithms could be applied to the complete set of available literature to generate a comprehensive knowledge-base for host–pathogen interactions.

3. Identification of secreted virulence factors

Before predicting interaction partners, the proteins that are actually employed by the pathogen for this purpose (the effectors) have to be identified. In case of viruses, all proteins are candidates since they are all expressed in the host system. In case of bacteria and protozoa, the detection of suitable candidates must be performed prior to further analysis. This prediction may be specific for the employed transport system. Three of the seven known bacterial transport-systems (Types III, IV and VI) penetrate the host cell membranes and inject proteins into the cytosol. Other bacterial transport systems can be used if effector entry into the hosts' cytosol is not required or mediated by alternative pathways: for example, the Type II secretion system has been shown to be relevant for bacterial virulence [41,42] but does not provide a direct way for translocation into the host's cytosol. Several of the bacterial transport systems can be employed by the same species. An example is *Pseudomonas syringae*, which delivers effectors by the Type III as well as by the Twin-Arginine system [43]. In the case of protozoa, the knowledge of the exact transport routes into the host are largely unknown [44] but effector specific motifs have

been identified in some species which can be exploited for prediction. There is a need for general prediction methods that do not refer to a certain transporter since not all routes to deliver effectors are known. In this paragraph, we describe methods to detect effectors based on sequence homology or on the presence of a signal peptide, a sequence that leads to active transport into the host. Moreover, we discuss detection methods by genomic and functional properties as domain signatures that hint to a role as effector protein.

3.1. Homology-based methods

With a list of known effectors at hand, homology searches can be applied to detect effector candidates. On the one hand, this approach has the drawback of being restricted to known families of effector proteins. This issue becomes especially evident when effectors are evolutionary “invented” by genome re-arrangements, as it has been shown for Type III secreted effector proteins in gram(–) bacteria [45]. On the other hand, effectors from closely related species or effectors that are transmitted by horizontal gene transfer (HGT) are detectable by homology searches. A good example for this method has been the *Shigella* toxin, which renders EHEC strains of *E. coli* especially harmful [46]. In this manuscript, 300 known and predicted effectors were used to detect homologous candidates in the proteome of an *E. coli* EHEC strain. This strategy identified 65 candidates of which 60% could be shown to be secreted. The sensitivity of homology searches in identifying effectors can be roughly estimated by the number of orthologous effectors known to date. The T3DB database [29] provides an up-to-date set of effectors with their respective orthologous relationships between different species. Where 90 effectors do not have well defined orthologs, 138 could potentially be detected by their evolutionary relationships to other effectors (resulting in a sensitivity of 60%). Since in this analysis genes from different strains are unified into one entry, the amount of successfully identified effectors between different strains might be higher.

3.2. Signal based detection

A more straightforward way to detect secreted proteins is the prediction of the secretion signal. This is a peptide sequence that is recognized by the respective secretion apparatus. This prediction is possible if basic information on the secretion signal is available. These data can be used to train a Hidden Markov Model or a binary classification algorithm. Substrates of two of the bacterial transport systems (namely, the Type II and V secretion system, commonly known as the sec-dependent pathways) can be detected

by both their cleavage signal and a hydrophobic N-terminal signal sequence. The latter is crucial for the transport into the periplasmic space [47]. Even if these systems do not deliver directly into the host cell, substrates of these secretion systems have been shown to play a role in infection, as can be shown for *Legionella* [48]. Computationally, this signal can be encoded in an Hidden Markov Model (HMM) which can detect transported proteins with high accuracy (sensitivity and selectivity >88%), as implemented in the SignalP detection software [49]. This software is also capable of predicting proteins that are secreted by the exocytosis pathway in Eukaryotes. In gram(–) bacteria, the transport of effectors is often mediated by the Type III secretion apparatus (TTSS), a molecular syringe spanning through the bacterial and eukaryotic host cells [50]. The recognition of substrates by the TTSS is mediated by the N-terminal end of the effector proteins [51–53] and is in some cases supported by TTSS specific chaperones [54,55]. Different studies applied binary classification algorithms (neural networks, Naive Bayes classifier, and SVMs) [56] for TTSS specific effector prediction resulting in predictions that are far better than random. This was determined in cross-validation experiments with a sensitivity up to 90% and a specificity of 88–97% [53,56,57]. Since the exact signal that leads to transport is unknown, these approaches use representations of amino-acid frequencies or frequencies of short amino-acid stretches from the proteins' N-termini as input features. The first 25 to the first 100 amino-acids are sufficient to initiate Type III secretion [52]. Samudrala et al. complemented this approach by implementing additional effector specific features such as their low sequence conservation and unusual phylogenetic distribution [53]. These additional features are discriminative between effectors and non-effectors and substantially contribute to the prediction performance by increasing the ROC-AUC (used to measure the general classification performance, which is 0.5 for a random prediction and 1.0 for a perfect classification) from 0.80 to 0.95. An interesting case is the agent of malaria, the eukaryote *Plasmodium falciparum* that infects erythrocytes and stays intracellular in a vacuole. The effectors of *Plasmodium* have one of two similar short motifs in common, called PEXEL and vacuolar transport signal (VTS) that lead to transport out of the pathogen, across the vacuole and into the host cell [58,59]. This motif can be utilized to detect novel virulence factors using a HMM or a sequence pattern. Using these motifs alone to detect secreted proteins would produce a large amount of false positives due to the shortness of the motifs, covering up to 28% of the *Plasmodium*'s proteome [60]. Hiss et al. reported that these motifs are flanked by patterns of amino acids with similar physico-chemical properties. These patterns can be used to classify the predicted effectors further, hereby reducing the amount of candidates to ~7% of the proteome [60]. Microarray data can be used to identify which effectors are active in certain stages of the life cycle, since malaria depends on the infection of erythrocytes and has a tissue specific life-cycle [59]. Another motif (the RxLR motif and flanking regions, similar to the PEXEL motif) has been described in several plant pathogens such as *Phytophthora infestans*, an oomycete. Remarkably, this signal is interchangeable with the PEXEL signal [61]. This finding might indicate a common mechanism used to deliver eukaryotic effectors. There is evidence that this mechanism works in the absence of the pathogen and therefore comprises a utilized host mechanism, but the topic is highly disputed [44,62].

3.3. “Genomic” and “function based” approaches

To detect effectors delivered by transport systems for which the signal of substrate recognition is completely unknown, more general features of effectors must be employed. For some bacterial species, effectors tend to be organized in pathogenicity islands on the chromosome and their genes exhibit a deviating CG content

[63]. Furthermore, their orthologs have an unusual phylogenetic distribution since they have been acquired by horizontal gene transfer [63]. They may therefore be more similar to eukaryotic domains as to other bacterial sequences [64]. A promising approach is based on machine learning that integrates several of such features. Such a method has been successfully applied for substrates of the Type IV secretion system in *Legionella pneumophila* [54], achieving a cross-validation performance of ROC-AUC 0.95 and a >90% success rate in wet-lab tests while testing for translocation. An interesting way to detect effectors is to identify domains suspected to interact with the host system as indicated by their binding mode or phylogenetic pattern. The latter case refers to so called “eukaryotic-like” domains unusual in bacteria but frequent in the eukaryotic host cell as ankyrin repeats [65–67]. In addition to serve as a basis for detection, these domains also indicate a function of the effector in the host and might mediate the interaction with the host system. Effector candidates can be detected by screening for these domains using HMM representations of such domains as they can be found in Pfam or SMART [68]. A convenient way to detect domain signatures is the use of Interpro-scan [69] which provides a comprehensive ‘umbrella-system’ for several domain signature databases. Many of the eukaryotic-like domains in effectors have been identified by expert knowledge in a case-to-case manner. Thousands of different domain signatures exist and comprehensively scanning these can provide additional domains of interest. The Effective database [27] provides a systematic screen for eukaryotic-like domains based on an evaluation of their taxonomic distribution between eukaryotes and pathogenic bacterial species. This database lists a frequency based Z-score for each domain in a pathogen which describes how more frequent the domain appears in the pathogenic species as in non-pathogenic ones. However, the reliability of this classification of eukaryotic domains has not been systematically assessed yet.

4. Prediction of HP-PPIs

The prediction of protein–protein interactions is a well-investigated problem for the intra-species case, and it is the next logical step to adapt these approaches for host–pathogen interactions. However, not all methods suitable for intra-species interaction prediction can be applied here. In this section, the most important methods for the predictions of HP-PPIs are discussed. These methods are based on mapping of known interactions, methods based on machine learning on sequence information, and methods based on domain–domain interactions. A special and very promising variant of the latter is the detection of short linear motifs (SLMs) that are the binding motifs for a number of eukaryotic interaction domains.

4.1. Interologs

The most evident approach is the mapping of known interactions onto homologous or orthologous pairs of sequences, based on the rationale of conserved interactions between the same or similar genes in different species. If an interaction of two genes is known in A, this interaction is likely to occur in B as well to maintain this common function. The transferred PPI is then called ‘Interolog’. The interolog method becomes weaker with larger evolutionary distances since genes may neo-functionalize during evolution and orthologous assignments will become more error-prone due to the lack of sequence similarity. In addition, many HP-PPIs will be unique inventions from the host–pathogen arms race that are not existent between orthologs of other species. Nevertheless, host–pathogen interactions may reveal known interactions, for instance if a gene acquired by HGT from the host could be re-utilized

to attack the host system. Since reliable mappings can only be achieved between sequences with very high sequence similarity (>30%) or very significant E-Values as shown in the studies by Aloy et al. [70] and Yu et al. [71], the original approach may miss many fast evolving effectors. In the case of HP-PPIs, several authors therefore combined a sensitive version of the interolog approach with pruning steps to remove false positives that are unlikely to interact due to spatial or temporal constraints. Using interologs, Wuchty [72] and Lee et al. [73] predicted HP-PPIs between *P. falciparum* and human, uncovering interactions that are meaningful for the maintenance of the parasites intracellular life-style. Using an additional machine learning step, Wuchty reports a 79% true positive rate while picking up only 4.7% false positives on a structurally inferred test-set from a previous study [74]. Lee et al. show that their predictions are significantly enriched with relevant Gene Ontology terms dealing with Calcium related processes which are relevant to the Plasmodium life cycle. Krishnadev et al. predicted interactions between human and three bacterial pathogens using a combined approach of stringent interologs and domain-based information (discussed below) hereby identifying several known as well as unknown relationships [75]. Tyagi et al. utilized the same approach to detect a small set of human–*Helicobacter* interactions [76]. Franzosa et al. implemented a structure based approach to detect human–virus interactions by first extracting human interacting pairs from the PDB and then mapping virus proteins to these structures by sequence similarity [77].

4.2. Genomic context based methods

Where text-mining as well as Interolog based approaches rely on known interactions, approaches to predict physically or at least functionally interacting proteins might uncover completely novel relationships. While genomic-context methods have been very successful to predict physical and functional interactions for species specific interactomes [78–80], direct application of genomic context methods on HP-PPIs lacks their underlying rationale (as genes should be neighbors on the chromosome to detect operons and conserved neighborhoods, or are under the same evolutionary constraints detectable by similar phylogenetic profiles, named the co-occurrence method). Nonetheless, interactions predicted using genomic context methods can be reliably mapped by orthology to other species, even if the genomes of said species did not contribute to the detection of the functional or physical interaction [79]. An application to a host–pathogen system would be an analogous procedure to the Interolog approach and is as such subjected to the same restrictions. In addition, the fraction of only functionally or indirectly interacting proteins in the predicted sets might be relatively big: in an early study of Huynen et al. [81], the fraction of predicted direct interactions has been reported to be between maximal ~50% (for tracking of gene fusion events) and minimal 34% for the co-occurrence method. In general, the usefulness of a genomic context based approach for HP-PPI prediction has not been determined but might provide an interesting source of information complementary to known physical interactions.

4.3. Sequence based methods

A promising way is the utilization of prediction methods that work on general properties of the amino-acid sequence. For example, machine learning techniques on representations of short stretches of amino-acids and their chemical–physical properties can be employed for a very general PPI prediction. Dyer and co-workers integrated such an approach for the prediction of human–HIV interactions [82,83]. In this work, the authors extended the sequence derived feature set with information on domains and information on the local network properties of the human

interaction partner. In contrast to the studies done on bacterial and eukaryotic pathogens, the high amount of experimental data for HIV allows the creation of a rigid test-set of reliable interactions. Dyer et al. evaluated several combinations of their features used for prediction and a combination of all (sequence, domain, and network information) achieved the best performance. Depending on the chosen cut-off, high precision and recall can be achieved, e.g. 80% precision at 50% recall.

4.4. Domain based prediction

Dyer et al. reported that a strong contributor to the successful prediction is the domain information [83]. This observation motivates the use of domain based methods to infer HP-PPI interactions. For the intra-species case, several methods to predict domain–domain interactions have been described. These methods are based on the rationale that domains are the mediators of interactions and, more importantly, are universally used by different proteins. The Protein Structural Interactome map (PSIMAP) and its cognate database, Psibase, is a resource that predicts interactions between folds as classified in the SCOP database by analyzing PDB structures [84]. A general resource for domain–domain interactions based on domain definitions given by Pfam models is the Domain Interaction Map (DIMA) [22]. It integrates several methods for predicting domain–domain interactions which rely on different principles of detection: in the Domain Pair Exclusion Method (DPEA, [85]) the frequency of known interactions of co-occurring domain pairs are analyzed. The DIPD [86] approach uses machine learning techniques to determine domain interaction pairs that are predictive for protein–interactions. Other integrated methods are based on correlated mutations and search for domain pairs that contain co-evolving residues [87]. The Domain Profile method (DPROF) is based on the same idea as the co-occurrence method to predict interactions between complete proteins: based on the assumption that interacting domains are under evolutionary pressure to be maintained concerted, the phylogenetic distributions of a pair of domains are compared and an interaction is reported if these are sufficiently high and informative [88]. IPfam [89] and 3DiD [90] analyze PDB structures and extract domain pairs that are in close contact in these structures. Where all these methods have been reported to result in meaningful predictions, their overall performance is difficult to assess due to the incompleteness of the domain and structural databases. However, they have been used in several host–pathogen screens since they allow to predict interactions over wider evolutionary distances as simple homology. Dyer et al. developed an own scoring system between domains in interactions using Bayesian statistics and predicted interactions between *H. sapiens* and *P. falciparum*. He reported over 500 PPIs which tend to significantly attack human proteins that are in close proximity to each other in the human interactome, that significantly share GO annotations, and that are co-expressed in the pathogen [91]. Tyagi et al. used, besides interologs, iPfam based predictions to increase the amount of detected *Helicobacter pylori*–human interactions. Kim et al. [92] developed XooNET, a pre-calculated resource for *Xanthomonas oryzae*, an important rice pathogen, comprising beside the intra-cellular interactome, predictions for 3400 host–pathogen interactions based on stringent cut-offs on PSIMAP, iPfam and interologs. One of the objectives for the utilization of domain–interactions to predict HP-PPIs is that an effector protein might have been acquired evolutionary by re-using a common interaction domain. The principle mechanism to achieve such a fusion is a mechanism of flexible re-shuffling of genes allowing to fuse a functional domain with a signal peptide for transport into the host. Such evolutionary re-shuffling events have been described for 18–32% of bacterial effectors, depending on the species of interest [45]. Some domains might not be part of the pathogens

genetic repertoire and have been reinvented as analogous molecular mimicry. Interesting candidates for such a process are short linear motifs, which might be easily gained by few mutations.

4.5. Short linear motifs

A wide range of protein domains recognize their substrates by short linear motifs (such as kinase domains and peptide recognition modules (PRMs), which include SH3 (Src homology 3), SH2 (Src homology 2), and PDZ (PSD-95/Discs-large/ZO-1) domains). Many play an important role in eukaryotic specific signal-transduction and regulation processes, making them an interesting target for pathogens. A general resource for these short linear motifs is the ELM database [93]. A recent survey of Davey et al. [94] identified at least 50 different short linear motifs that exist in viral proteins and that are used to manipulate their host system by acting as molecular mimicry. Examples from bacterial effector proteins have been described as well [79,94–96]. Dampier et al. reported that the outcome of HIV infection treatments depend on the existence of linear motifs in the respective HIV strain indicating that these are functional and interact with the host [97]. Methods for predicting short linear motifs are based on experimentally derived data on binding specificity from synthetic peptide arrays (SPOT), oriented peptide array libraries (OPAL), protein domain microarrays, and phage display based techniques [98–103]. The results of these experiments provide data on binding specificities which are captured in position weight matrices (PWMs), alternatively called position specific specificity profiles (PSSMs). These matrices describe the short linear motifs based on the frequencies of each measured amino-acid for each position [99]. Computational methods such as Scansite use PWMs to predict peptide binding motifs for a range of domain families such as kinase, SH3, SH2, PDZ, and 14–3–3 domains [104]. The scoring system of the PWMs can be exploited according to this rationale since higher binding specificities are reflected in higher scores assigned by the detection algorithm. In order to detect candidate molecular mimicry, scanning for short linear motifs in e.g. viral proteins and comparing them to the highest scoring human instances is needed, as done for PDZ domain recognition sites by Tonikian et al. [21]. They were able to detect known viral targets (such as SCRIB, whose functional disruption causes hyperproliferation) and the occurrence of binding motifs to the PDZ domain of this protein correlated with the oncogenic potential of HPV strains. Protein structure information such as protein disorder and solvent accessibility can be used in conjunction with PWMs to increase the accuracy of the prediction. MOTIPS is such a computational method that incorporates protein structure and domain specificity information in a Bayesian framework to predict binding partners for SH3 and S/T kinase domains [105].

5. Example: bacterial effector PPI prediction based on domain-domain interactions

In this section, we discuss an example analysis to deliver candidate HP-PPIs based on domain–domain interactions as a possible novel way to predict and rank bacteria–human interactions. We used publicly available data to predict human interaction partners of known effectors from *Salmonella* sp. Since the resulting lists of putative targets of a certain domain might be very large, we ranked the interactor candidates by the degree of the target proteins in the human PPI network. Due to the effectors tendency to interact with hubs [106], the actual interactors should appear enriched in the top of these lists.

For this approach, domain–domain interaction data have been downloaded from the DIMA 3.0 database [22]. A scan of all Pfam

domains in proteins from human and the respective pathogens was downloaded from SIMAP (a database that provides pre-calculated data on all Pfam models [107]). Data on the human interactome were downloaded from iRefWeb [35]. For each human protein, we computed the degree as the number of reported physical interactions from high-throughput data (filtering experiments with <100 interactions). The resulting data contain the interacting Pfam domains and the human and bacterial proteins that contain one of the corresponding domains, with the degree of the proteins in decreasing order. The data were searched for effectors that have been experimentally verified or are homologous to a verified effector in another bacteria. This list has been obtained from the literature [108,109]. In principle, this list could be extended by predicted effectors to get a more extended host–pathogen interaction network. Table 2 contains a summary of the possible interactions between *Salmonella* effectors and host proteins. It lists the human protein with the highest degree together with their function. The overall work-flow is pictured in Fig. 1.

SopE and SopE2 of *Salmonella typhimurium* and the SopE *Salmonella enterica* ortholog are predicted and known to interact with Cdc42. In the case of the *S. enterica*, it was the only interaction predicted. A 3D structure is available for this well documented interaction and the physiological effects are also known: SopE activates Cdc42 resulting in actin cytoskeletal rearrangements that promote bacterial uptake in non-phagocytotic cells [110]. The SipA effector from *S. typhimurium* is known to inhibit F-actin destabilizing proteins [111,112], such as vinculin, an F-actin binding protein. The SlrP and SspH2 effectors have a leucine rich repeat (PF00560) and are also known Cdc42 targeting proteins and inhibitor of interleukin expression. Correspondingly, they are predicted to bind the TNF α receptor and Cdc42. Moreover, SlrP and SspH2 are involved in the ubiquitin pathway [109,113] and are here predicted interaction partners of E3 ubiquitin–protein ligase SIAH1 and S-phase kinase-associated protein 1, which regulates the ubiquitination of NF κ B inhibitor alpha. SptP is a phosphatase and an inhibitor of Cdc42, helping the host cell to recover after bacterial invasion [114]. Its predicted targets are adapter proteins (most notably from the Ras pathway), the ubiquitin and proteasome pathway and with multiple cytoskeletal components. Of these cytoskeletal components, SptP is predicted to interact with WASP, a well-known actin filament stabilizing protein and target for bacterial effectors [5], with a WASP binding protein (Pre-mRNA-processing factor 40 homolog A), and with tubulin and with a protein involved in vesicular transport (Vesicle-fusing ATPase). SpvB is an enzyme that ribosylates actin and is predicted to interact with actin and Cdc42. SifA, SifB and SseJ all share a domain that mediates the interaction with Phosphoinositide phospholipase C-gamma-1, an interaction that is structurally resolved [115]. SifA and SifB are involved in the formation of *Salmonella* induced filaments (Sifs). SseJ has an additional lipase/acylhydrolase domain and is a negative regulator of Sif formation. Its activity is a virulence determinant [116]. Our prediction indicates a strong role in ubiquitin pathway alteration.

To conclude, the *Salmonella* data are consistent with previous work on effector protein interactions by predicting interactions with all known pathways that are targeted upon infection.

6. Increasing the precision of HP-PPI predictions

Many of the discussed approaches tend to over-predict the amount of interaction partners. For example, domain–domain interaction based methods will return many wrong positive predictions for frequent domain signatures. In general, the optimal trade-off between sensitivity and selectivity could be found with benchmark sets of HP-PPIs. These sets only exist for a couple of

Table 2
Effector–host PPI predictions for *Salmonella typhimurium*. A short description of the effector function is depicted beside their name. For each effector, the Pfam names of the interacting domains are depicted. The bacterial domain is named first. For the human domain, the protein with the highest degree, together with their function are given.

SipA interferes with host cell actin cytoskeleton		
PF09052–PF01044	VINC_HUMAN	Actin filament binding protein
sopE–sopE2 activator of CDC42 and RAC1		
PF07487–PF00071	CDC42_HUMAN	Cytoskeleton organiser
sptP tyrosine phosphatase and GTPase activating protein (GAP) activities		
PF00102–PF00017	GRB2_HUMAN	Adapter protein linking growth factor receptors and the Ras signaling pathway
PF00102–PF07714	EGFR_HUMAN	EGF receptor
PF00102–PF00244	1433G_HUMAN	Adapter protein that regulates signaling pathways
PF00102–PF02262	CBL_HUMAN	Adapter protein that negatively regulates signaling pathways
PF00102–PF00102	PTN12_HUMAN	Dephosphorylates cellular tyrosine kinases
PF00102–PF00647	EF1G_HUMAN	Molecular anchor
PF00102–PF01846	PR40A_HUMAN	Binds to WASP (actin cytoskeleton organiser)
PF00102–PF00493	MCM2_HUMAN	Component replicative helicase complex
PF00102–PF00041	INSR_HUMAN	Receptor tyrosine kinase which mediates the actions of insulin
PF00102–PF01049	CADH2_HUMAN	Calcium-dependent cell adhesion protein
PF00102–PF03810	IMB1_HUMAN	Nuclear protein import
PF00102–PF02933	NSF_HUMAN	Required for vesicle mediated transport
PF00102–PF00568	WASP_HUMAN	WASP (actin cytoskeleton organiser)
PF00102–PF00091	TBA3C_HUMAN	Tubulin forms microtubules
PF03545–PF00071	CDC42_HUMAN	Cytoskeleton organiser
sspH2–slrP E3 ubiquitin ligase that interferes with host's ubiquitination pathway		
PF00560–PF02344	MYC_HUMAN	Activates the transcription of growth-related genes
PF00560–PF08919	ABL1_HUMAN	Non-receptor tyrosine-protein kinase involved in cell growth and survival, cytoskeleton remodeling etc.
PF00560–PF00076	EWS_HUMAN	May act as arepressor in the tumorigenic process
PF00560–PF00071	CDC42_HUMAN	Cytoskeleton organiser
PF00560–PF00020	TNR1A_HUMAN	Receptor for TNFSF2/TNF-alpha (apoptosis)
PF00560–PF06068	RUVB2_HUMAN	Helicase activity
PF00560–PF00012	HSP7C_HUMAN	Repressor of transcriptional activation
PF00560–PF00149	PP2AA_HUMAN	Modulates activity MAP-2 kinase, dephosphorylation of p53
PF00560–PF00769	EZRI_HUMAN	Probably involved in connections of major cytoskeletal structures to the plasma membrane
PF00560–PF07679	RB_HUMAN	Regulator of cell division that acts as a tumor suppressor
PF00560–PF00028	CADH2_HUMAN	Calcium-dependent cell adhesion protein
PF00560–PF03931	SKP1_HUMAN	Essential component of the SCF ubiquitin ligase complex
PF00560–PF00134	CCNK_HUMAN	Cyclin-K: possible role in transcriptional regulation
PF00560–PF01749	IMA2_HUMAN	Functions in nuclear protein import
PF00560–PF03145	SIAH1_HUMAN	E3 ubiquitin-protein ligase
sifA–sifB promotes bacterial survival, formation of Salmonella-induced filaments		
PF06767–PF00169	PLCG1_HUMAN	Regulation of intracellular signaling cascades involving IP3
sseJ promotes bacterial survival, formation of Salmonella-induced filaments		
PF06767–PF00169	PLCG1_HUMAN	Regulation of intracellular signaling cascades involving IP3
PF00657–PF00400	2ABB_HUMAN	Promotes proapoptotic activity within the PP2A complex
PF00657–PF00240	UBQL4_HUMAN	Plays a role in the regulation of proteosomal protein degradation
PF00657–PF00097	TRAF6_HUMAN	E3 ubiquitin ligase
PF00657–PF02037	XRCC6_HUMAN	ATP-dependent helicase
SpvB actin cytoskeleton reorganiser		
PF03496–PF00071	CDC42_HUMAN	Cytoskeleton organiser
PF03496–PF00022	ACTG_HUMAN	Actin filaments

host–pathogen pairs (as HIV–human). Therefore, the challenge is to prune over-predicted interactions in a biological meaningful way to get a realistic picture of the interactome and to prioritize interesting targets for follow-up experiments (such as structural modeling). So, the precision of the HP–PPI predictions must be increased by removing biologically unlikely results. On the host side, this can be achieved by integrating data on tissue specific expression (if the pathogen is known to attack a certain tissue), and on the pathogen side by removing genes which are not expressed in the infection cycle. In addition, gene expression analysis of cells during infection may hint to genes activated due to the hosts defense mechanisms. Dyer et al. showed that genes of *P. falciparum* whose proteins are predicted to interact with human proteins tend to be co-expressed in the different life-cycle states of the parasite [91]. Wuchty used gene expression information on the *Plasmodium*'s life cycle, prediction of exported proteins, and human tissue specificity to prune candidate lists for the prediction of a *P. falciparum*–human PPI network, breaking down an initial network of >100000 candidates to a set of 2244 interacting pairs [72]. Davis

et al. refined predictions for 10 different bacterial pathogens by the use of host and pathogen expression data reducing the amount of predictions from several thousands to a few hundred high confidence candidates [74].

7. Different biology of viruses and bacteria – possible implications for HP–PPI predictions

Several discussed rationales behind the HP–PPI prediction may be limited to certain pathogen classes. Molecular mimics are found in viruses and bacteria, but they can originate by different ways and be subjected to different selection pressures.

Bacteria exchange genetic material with the host [117] and HGTs leading to bacterial virulence related genes have some well documented cases, e.g. some effectors in *Legionella* (RalF effector in particular [118,64]; for a review see Stebbins et al. [119]). However, not all mimicry originate by HGT, and several reported cases of functional or structural bacterial mimicry are probably *de novo*

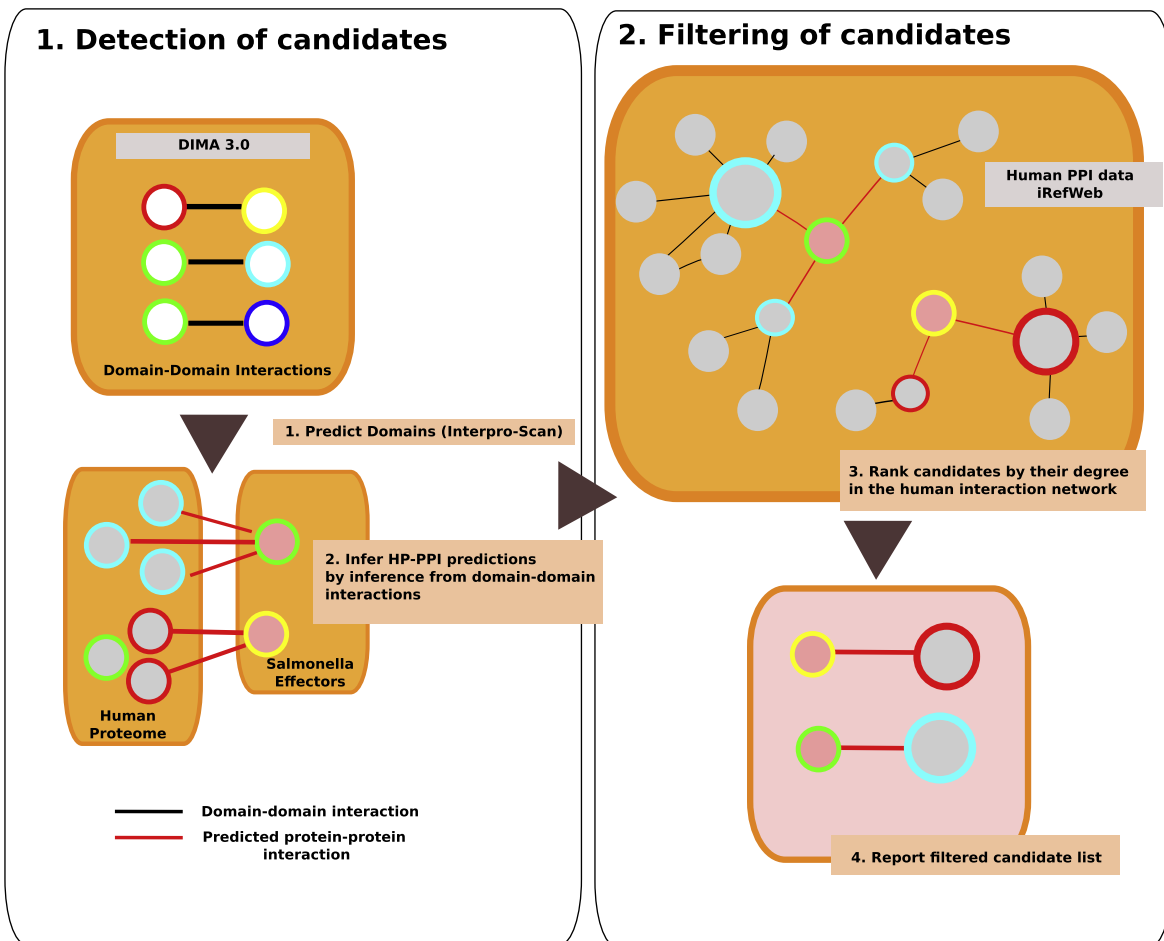


Figure 1. Work-flow of the example analysis. (1) Detection of host-pathogen interaction candidates by a domain-domain based approach, (2) ranking of the candidates according to the degree in the host interaction network.

evolutionary inventions, like the *Shigella* IpgB2 [13] or the *E. coli* (EPEC) EspF effector [12]. Detecting these effectors by Interolog or genomic context approaches is difficult. A possible solution for these cases would be the individual modeling of the three-dimensional structure of putative effector–target complexes, which is, for now, intractable for a large set of sequences. On the other hand, while focusing on domains, short eukaryotic interaction scaffold and signaling domains have been reported for several known or suspected effector proteins [65,120,121]. Evidently, these eukaryotic-like domains can also be detected in effectors with clearly eukaryotic origin (acquired by HGT), such as the *Legionella* LeS2 effector and others [121,122]. These examples were detected by the use of domain-signatures or sensitive sequence similarity searches. Bacteria exchange their genetic material frequently by conjugation and the exchange of plasmids, often exchanging virulence factors [117,123,124]. Some obligate intra-cellular bacteria as *Chlamydia* are excluded from this source of rapid genetic exchange as long as they do not share their habitat with other bacteria, as reported for bacterial symbionts in amoebae [125]. The exchange of genetic material results in a phylogenetic signal that can be used to predict effectors.

Viruses mutate very rapidly but have strong constraints on their genomic size. Such constraints may lead to a preference for mimicking short linear motifs. These SLMs can be easily achieved by mutations in viral effector proteins (e.g. the generation of a PxxP binding motif to interact with SH3 domains). The mutational potential and evolutionary pressure to limit the genome size of viruses makes the search for clear examples of HGT difficult. It is

thus not clear to what extent the large overlap of host–pathogen and host–host interaction interfaces [77] is caused by HGT (instead of being new evolutionary inventions). In a recent work, Rappoport et al. systematically investigated HGT events from eukaryotes to viruses and could identify several instances of dsDNA viruses that clearly acquired genetic material from the host. In addition, they report that these are often streamlined into single domain proteins without linker regions, a finding in congruence with the viruses' genomic size constraints [126].

To conclude, Interolog and domain signature based methods may be more applicable to non-viral pathogens, whereas detection of short linear motifs might be more successful for viral genomes for which many of these have been reported [127,128]. Genomic context and Interolog methods are most successful in clear cases of HGT. The structural modeling can be applied to all, if 3D structures of related host species are available. Domain–domain based interaction methods are applicable to detect interactions based on HGT from host to pathogen and might be fruitful for viruses considering the described streamlining effect.

8. What do the HP-PPI networks tell us?

The first studies of host pathogen PPI networks were based on high-throughput screens such as yeast-two-hybrid and more recently, genome-wide RNAi scans to screen for functional or genetic interactions. Their inherent potential makes them indispensable to understand the host–pathogen interaction on a genome-wide and

systems biology scale. Y2H based networks, sometimes backed up by literature mining (for HIV type 1 virus text mining see [129,130]), are the primal source of recent day physical HP-PPI networks of viruses, bacteria and parasitic protozoa. These networks are available for HIV type 1 virus [131–133], Hepatitis C virus [134,135], Influenza [136], SARS-coronavirus [137], Herpes simplex virus [138,139], Dengue virus [140], Epstein–Barr virus [141], *Yersinia pestis*, *Bacillus anthracis* and *Francisella tularensis* [142]. These studies have major implications for our understanding of infection, as the extent of the HP-PPI network and the effectiveness of pathogen interference can only be appreciated by genome-wide assays. Systems biology studies show that host pathways are targeted (i) specifically, (ii) by multiple effectors and (iii) preferably at hubs (highly connected proteins in the interaction networks). This in accordance with the observation, that effector proteins are rich in binding motifs and interacting domains, a fact that is seen in effectors of pathogens of animals as well as plants [106].

For example, acetylated HIV integrase has thirteen known binding partners and effects key proteins in several processes, ranging from chromatin remodeling to cytoskeleton rearrangement [133]. HCV and influenza HP-PPI networks also confirm the aforementioned view, whereby pathways are altered by direct PPI interactions from receptor scaffolds down to the level of transcription factors [136]. Navratil et al. showed that 22% of the interferon pathway, a major antiviral player, is directly interacting with viral proteins [143]. As mentioned in the introduction, the targeted pathways are related to cell cycle regulation, nuclear transport, ubiquitin, proteasomal degradation and immune response regulation [79,82,142,144,145]. These pathways are targeted in order to effectively manipulate the host system and circumvent the immune system. It could be shown that two evolutionary distinct pathogens of Arabidopsis, the bacterium *P. syringae* and the eukaryote *Hyaloperonospora arabidopsidis* target the same pathways as a result of convergent evolution [106]. The other recurring theme in HP-PPI network analysis, the targeting of hubs and bottlenecks by pathogen proteins, is usually explained as an evolutionary method of the pathogen to try to maximize their influence on the host to redirect the network priorities to increase pathogen fitness. Franzosa and Xia [77] reported that the viral proteins structurally mimic the interface of host regulatory protein domains that are involved in transient contacts. This way, pathogen proteins can out-compete host proteins for interactions [146].

For now, the most feasible structural approaches are investigating domain–domain interactions where other approaches such as 3-dimensional modeling of host–pathogen networks are still too computationally demanding to be easily automated for a large scale prediction. Known domain–domain interactions were the starting point for the PPI mapping of *P. falciparum* [147]. Likewise, the DIP and iPfam databases were used for the construction of the *H. pylori*–human interaction network [76]. We used a similar approach in this article to point out the merit of combining a domain based interaction prediction method and general network characteristics to get probable interaction partners of known bacterial effector proteins. The present-day challenge is the combination of structural biology and network biology to increase the reliability and completeness of HP-PPI networks. Another difficult task is the implementation of non-domain related interactions such as short linear motifs; this will complete our view on host–pathogen interactions as the motifs are known to be enriched in viral proteins [148].

9. Biological implications of HP interactomics and future perspectives

In this review, we focused on the prediction of HP-PPI and the analysis of their impact on the host. One important prospective

of this research is the development of novel treatment strategies. Structure-based predictions of domain–domain interactions and motif binding allow the development of specific compounds that interfere with HP-PPIs as it has been done with small molecules that interfere with the p53 [149], TNF α [150] and Wnt [151] pathways as possible cancer treatments. Interestingly, these pathways may also be targets of viruses and bacteria: the p53, TNF α and Wnt pathways are altered by pathogens to avoid immune induced apoptosis. This indicates that the pool of possible targets is large and to a degree overlapping with diseases other than infection.

The search for new antibiotics and infection-related drugs will be greatly supported by structural modeling of inhibitor–protein complexes to block elementary steps in the pathogens interaction with the host. An example of this approach is the inhibition of neuraminidase inhibitors for treatment of influenza. Since resistance of influenza to neuraminidase inhibitors such as Tamiflu[®] is already widespread [152], additional drugs are urgently needed. Every confirmed interaction in HP-PPI can be potentially used for the development of novel antiviral drugs.

Until recently, treatments focused on targeting a few key players in the etiology of the disease. Systems biology provides another treatment paradigm: can we use the information that HP-PPI networks provide us to design strategies to combat infection at the network level? The HP-PPI or host PPI networks can both be targeted for this purpose [153]. Systems biology approaches may help to identify the impact of virulence factors on the host system based on computational models of signal transduction and pathway analysis. Key in this approach is the application of small molecules against several nodes that are altered upon infection [154]. Both treatment viewpoints are yet to be explored and both are critically dependent on the ongoing optimization of the techniques presented in this review.

References

- [1] S. Parer, A. Lotthé, P. Chardon, R. Poncet, H. Jean-Pierre, E. Jumas-Bilak, *Infect. Control Hosp. Epidemiol.* 33 (2012) 167–174.
- [2] U. Buchholz, H. Bernard, D. Werber, M.M. Böhmer, C. Remschmidt, H. Wilking, Y. Deleré, M. Van der Heiden, C. Adlhoch, J. Dreesman, J. Ehlers, S. Ethelberg, M. Faber, C. Frank, G. Fricke, M. Greiner, M. Höhle, S. Ivarsson, U. Jark, M. Kirchner, J. Koch, G. Krause, P. Lubner, B. Rosner, K. Stark, M. Kühne, *N. Engl. J. Med.* 365 (2011) 1763–1770.
- [3] C. Walsh, *Nat. Rev. Microbiol.* 1 (2003) 65–70.
- [4] W. Magliani, S. Conti, A. Salati, S. Arseni, L. Ravanetti, R. Frazzi, L. Polonelli, *Curr. Med. Chem.* 11 (2004) 1793–1800.
- [5] N.C. Elde, H.S. Malik, *Nat. Rev. Microbiol.* 7 (2009) 787–797.
- [6] M.N. Lurie-Weinberger, L. Gomez-Valero, N. Merault, G. Glöckner, C. Buchrieser, U. Gophna, *Int. J. Med. Microbiol.* 300 (2010) 470–481.
- [7] G. Aguilera, G. Refrégier, R. Yockteng, E. Fournier, T. Giraud, *Infect. Genet. Evol.* 9 (2009) 656–670.
- [8] Y.-K. Shin, Y. Li, Q. Liu, D.H. Anderson, L.A. Babiuk, Y. Zhou, *J. Virol.* 81 (2007) 12730–12739.
- [9] B.G. Hale, I.H. Batty, C.P. Downes, R.E. Randall, *J. Biol. Chem.* 283 (2008) 1372–1380.
- [10] M. Thomas, C. Kranjec, K. Nagasaka, G. Matlashewski, L. Banks, *Virology* 42 (2011) 25.
- [11] L.S. Heikkinen, A. Kazlauskas, K. Melén, R. Wagner, T. Ziegler, I. Julkunen, K. Saksela, *J. Biol. Chem.* 283 (2008) 5719–5727.
- [12] N.M. Alto, A.W. Wefflen, M.J. Rardin, D. Yazar, C.S. Lazar, R. Tonikian, A. Koller, S.S. Taylor, C. Boone, S.S. Sidhu, S.L. Schmid, G.A. Hecht, J.E. Dixon, *J. Cell Biol.* 178 (2007) 1265–1278.
- [13] N.M. Alto, F. Shao, C.S. Lazar, R.L. Brost, G. Chua, S. Mattoo, S.A. McMahon, P. Ghosh, T.R. Hughes, C. Boone, J.E. Dixon, *Cell* 124 (2006) 133–145.
- [14] C. Delevoye, M. Nilges, P. Dehoux, F. Paumet, S. Perrinet, A. Dautry-Varsat, A. Subtil, *PLoS Pathog.* 4 (2008) e1000022.
- [15] L. Böhme, T. Rudel, *Microbes Infect.* 11 (2009) 1063–1070.
- [16] C. Swanton, G.L. Card, D. Mann, N. McDonald, N. Jones, *Trends Biochem. Sci.* 24 (1999) 116–120.
- [17] C.A. Collins, E.J. Brown, *Trends Cell Biol.* 20 (2010) 205–213.
- [18] K. Rottner, T.E.B. Stradal, J. Wehland, *Dev. Cell* 9 (2005) 3–17.
- [19] L. Diacovich, J.-P. Gorvel, *Nat. Rev. Microbiol.* 8 (2010) 117–128.
- [20] B. Raghavachari, A. Tasneem, T.M. Przytycka, R. Jothi, *Nucleic Acids Res.* 36 (2008) D656–D661.

- [21] R. Tonikian, Y. Zhang, S.L. Sazinsky, B. Currell, J.-H. Yeh, B. Reva, H.A. Held, B.A. Appleton, M. Evangelista, Y. Wu, X. Xin, A.C. Chan, S. Seshagiri, L.A. Lasky, C. Sander, C. Boone, G.D. Bader, S.S. Sidhu, *PLoS Biol.* 6 (2008) e239.
- [22] Q. Luo, P. Pagel, B. Vilne, D. Frishman, *Nucleic Acids Res.* 39 (2011) D724–D729.
- [23] D. Gfeller, F. Butty, M. Wierzbicka, E. Verschuereen, P. Vanhee, H. Huang, A. Ernst, N. Dar, I. Stagljar, L. Serrano, S.S. Sidhu, G.D. Bader, P.M. Kim, *Mol. Syst. Biol.* 7 (2011) 484.
- [24] L. Chen, J. Yang, J. Yu, Z. Yao, L. Sun, Y. Shen, Q. Jin, *Nucleic Acids Res.* 33 (2005) D325–D328.
- [25] C.E. Zhou, J. Smith, M. Lam, A. Zemla, M.D. Dyer, T. Slezak, *Nucleic Acids Res.* 35 (2007) D391–D394.
- [26] J.J. Gillespie, 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, *Infect. Immun.* 79 (2011) 4286–4298.
- [27] M.-A. Jehl, R. Arnold, T. Rattei, *Nucleic Acids Res.* 39 (2011) D591–D595.
- [28] D.M.M. Tay, K.R. Govindarajan, A.M. Khan, T.Y.R. Ong, H.M. Samad, W.W. Soh, M. Tong, F. Zhang, T.W. Tan, *BMC Bioinformatics* 11 (Suppl 7) (2010) S4.
- [29] Y. Wang, H. Huang, M. Sun, Q. Zhang, D. Guo, *BMC Bioinformatics* 13 (2012) 66.
- [30] R. Winnenburger, M. Urban, A. Beacham, T.K. Baldwin, S. Holland, M. Lindeberg, H. Hansen, C. Rawlings, K.E. Hammond-Kosack, J. Köhler, *Nucleic Acids Res.* 36 (2008) D572–D576.
- [31] R. Kumar, B. Nanduri, *BMC Bioinformatics* 11 (Suppl 6) (2010) S16.
- [32] I. Xenarios, L. Salwinski, X.J. Duan, P. Higney, S.-M. Kim, D. Eisenberg, *Nucleic Acids Res.* 30 (2002) 303–305.
- [33] L. Licata, L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A.P. Nardozza, E. Santonico, L. Castagnoli, G. Cesareni, *Nucleic Acids Res.* 40 (2012) D857–D861.
- [34] T. Driscoll, M.D. Dyer, T.M. Murali, B.W. Sobral, *Nucleic Acids Res.* 37 (2009) D647–D650.
- [35] B. Turner, S. Razick, A.L. Turinsky, J. Vlasblom, E.K. Crowdy, E. Cho, K. Morrison, I.M. Donaldson, S.J. Wodak, *Database (Oxford)*, 2010, baq023.
- [36] A. Chatr-aryamontri, A. Ceol, D. Peluso, A. Nardozza, S. Panni, F. Sacco, M. Tinti, A. Smolyar, L. Castagnoli, M. Vidal, M.E. Cusick, G. Cesareni, *Nucleic Acids Res.* 37 (2009) D669–D673.
- [37] F.-K. Lin, C.-L. Pan, J.-M. Yang, T.-J. Chuang, F.-C. Chen, *BMC Microbiol.* 9 (2009) 164.
- [38] M.E. Fahey, M.J. Bennett, C. Mahon, S. Jäger, L. Pache, D. Kumar, A. Shapiro, K. Rao, S.K. Chanda, C.S. Craik, A.D. Frankel, N.J. Krogan, *BMC Bioinformatics* 12 (2011) 298.
- [39] L. Yin, G. Xu, M. Torii, Z. Niu, J.M. Maisog, C. Wu, Z. Hu, H. Liu, *Artif. Intell. Med.* 49 (2010) 155–160.
- [40] T. Thieu, S. Joshi, S. Warren, D. Korin, *Bioinformatics* 28 (2012) 867–875.
- [41] O. Rossier, S.R. Starkenburg, N.P. Cianciotto, *Infect. Immun.* 72 (2004) 310–321.
- [42] N.P. Cianciotto, *Trends Microbiol.* 13 (2005) 581–588.
- [43] P.A. Bronstein, M. Marrichi, S. Cartinhour, D.J. Schneider, M.P. DeLisa, *J. Bacteriol.* 187 (2005) 8450–8461.
- [44] J.G. Ellis, P.N. Dodds, *Proc. Natl. Acad. Sci. USA* 108 (2011) 14381–14382.
- [45] J. Stavriniades, W. Ma, D.S. Guttman, *PLoS Pathog.* 2 (2006) e104.
- [46] T. Tobe, S.A. Beatson, H. Taniguchi, H. Abe, C.M. Bailey, A. Fivian, R. Younis, S. Matthews, O. Marches, G. Frankel, T. Hayashi, M.J. Pallen, *Proc. Natl. Acad. Sci. USA* 103 (2006) 14941–14946.
- [47] A.P. Pugsley, *Microbiol. Rev.* 57 (1993) 50–108.
- [48] L.M. Hales, H.A. Shuman, *Infect. Immun.* 67 (1999) 3662–3666.
- [49] H. Nielsen, A. Krogh, *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6 (1998) 122–130.
- [50] D.S.A. Beecman, D.C.G. Vranompraj, *Curr. Issues Mol. Biol.* 12 (2009) 17–42.
- [51] L.W. Cheng, D.M. Anderson, O. Schneewind, *Mol. Microbiol.* 24 (1997) 757–765.
- [52] A. Subtil, C. Delevoeye, M.-E. Balañá, L. Tastevin, S. Perrinet, A. Dautry-Varsat, *Mol. Microbiol.* 56 (2005) 1636–1647.
- [53] R. Samudrala, F. Heffron, J.E. McDermott, *PLoS Pathog.* 5 (2009) e1000375.
- [54] D. Burstein, T. Zusman, E. Degtyar, R. Viner, G. Segal, T. Pupko, *PLoS Pathog.* 5 (2009) e1000508.
- [55] M. Guo, S.T. Chancey, F. Tian, Z. Ge, Y. Jamir, J.R. Alfano, *J. Bacteriol.* 187 (2005) 4257–4269.
- [56] Y. Wang, Q. Zhang, M.-A. Sun, D. Guo, *Bioinformatics* 27 (2011) 777–784.
- [57] R. Arnold, S. Brandmaier, F. Kleine, P. Tischler, E. Heinz, S. Behrens, A. Niinikoski, H.-W. Mewes, M. Horn, T. Rattei, *PLoS Pathog.* 5 (2009) e1000376.
- [58] N.L. Hiller, S. Bhattacharjee, C. van Ooij, K. Liolios, T. Harrison, C. Lopez-Estrano, K. Haldar, *Science* 306 (2004) 1934–1937.
- [59] M. Marti, R.T. Good, M. Rug, E. Knuepfer, A.F. Cowman, *Science* 306 (2004) 1930–1933.
- [60] J.A. Hiss, J.M. Przyborski, F. Schwarte, K. Lingelbach, G. Schneider, *PLoS One* 3 (2008) e1560.
- [61] S. Bhattacharjee, N.L. Hiller, K. Liolios, J. Win, T.-D. Kanneganti, C. Young, S. Kamoun, K. Haldar, *PLoS Pathog.* 2 (2006) e50.
- [62] D. Dou, S.D. Kale, X. Wang, R.H.Y. Jiang, N.A. Bruce, F.D. Arredondo, X. Zhang, B.M. Tyler, *Plant Cell* 20 (2008) 1930–1947.
- [63] O. Gal-Mor, B.B. Finlay, *Cell. Microbiol.* 8 (2006) 1707–1719.
- [64] K.S. de Felipe, S. Pampou, O.S. Jovanovic, C.D. Pericone, S.F. Ye, S. Kalachikov, H.A. Shuman, *J. Bacteriol.* 187 (2005) 7716–7726.
- [65] P. Caturegli, K.M. Asanovich, J.J. Walls, J.S. Bakken, J.E. Madigan, V.L. Popov, J.S. Dumler, *Infect. Immun.* 68 (2000) 5277–5283.
- [66] S. Al-Khodor, C.T. Price, F. Habyarimana, A. Kalia, Y.A. Kwaik, *Mol. Microbiol.* 70 (2008) 908–923.
- [67] S. Blanié, J. Gelfi, S. Bertagnoli, C. Camus-Bouclainville, *Virol. J.* 7 (2010) 56.
- [68] I. Letunic, T. Doerks, P. Bork, *Nucleic Acids Res.* 37 (2009) D229–D232.
- [69] E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, *Nucleic Acids Res.* 33 (2005) W116–W120.
- [70] P. Aloy, H. Ceulemans, A. Stark, R.B. Russell, *J. Mol. Biol.* 332 (2003) 989–998.
- [71] H. Yu, N.M. Luscombe, H.X. Lu, X. Zhu, Y. Xia, J.-D.J. Han, N. Bertin, S. Chung, M. Vidal, M. Gerstein, *Genome Res.* 14 (2004) 1107–1118.
- [72] S. Wuchty, *PLoS One* 6 (2011) e26960.
- [73] S.-A. Lee, C.-h. Chan, C.-H. Tsai, J.-M. Lai, F.-S. Wang, C.-Y. Kao, C.-Y.F. Huang, *BMC Bioinformatics* 9 (Suppl. 12) (2008) S11.
- [74] F.P. Davis, D.T. Barkan, N. Eswar, J.H. Mckerrow, A. Sali, *Protein Sci.* 16 (2007) 2585–2596.
- [75] O. Krishnadev, N. Srinivasan, *Int. J. Biol. Macromol.* 48 (2011) 613–619.
- [76] N. Tyagi, O. Krishnadev, N. Srinivasan, *Mol. Biosyst.* 5 (2009) 1630–1635.
- [77] E.A. Franzosa, Y. Xia, *Proc. Natl. Acad. Sci. USA* 108 (2011) 10538–10543.
- [78] G. Kolesov, H.W. Mewes, D. Frishman, *J. Mol. Biol.* 311 (2001) 639–656.
- [79] C. von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, B. Snel, *Nucleic Acids Res.* 31 (2003) 258–261.
- [80] P.M. Bowers, M. Pellegrini, M.J. Thompson, J. Fierro, T.O. Yeates, D. Eisenberg, *Genome Biol.* 5 (2004) R35.
- [81] M. Huynen, B. Snel, W. Lathe, P. Bork, *Genome Res.* 10 (2000) 1204–1210.
- [82] M.D. Dyer, T.M. Murali, B.W. Sobral, *PLoS Pathog.* 4 (2008) e32.
- [83] M.D. Dyer, T.M. Murali, B.W. Sobral, *Infect. Genet. Evol.* 11 (2011) 917–923.
- [84] S. Gong, G. Yoon, I. Jang, D. Bolser, P. Dafas, M. Schroeder, H. Choi, Y. Cho, K. Han, S. Lee, H. Choi, M. Lappe, L. Holm, S. Kim, D. Oh, J. Bhak, *Bioinformatics* 21 (2005) 2541–2543.
- [85] R. Riley, C. Lee, C. Sabatti, D. Eisenberg, *Genome Biol.* 6 (2005) R89.
- [86] X.-M. Zhao, L. Chen, K. Aihara, *Proteins* 78 (2010) 1243–1253.
- [87] I. Kass, A. Horovitz, *Proteins* 48 (2002) 611–617.
- [88] P. Pagel, P. Wong, D. Frishman, *J. Mol. Biol.* 344 (2004) 1331–1346.
- [89] R.D. Finn, M. Marshall, A. Bateman, *Bioinformatics* 21 (2005) 410–412.
- [90] A. Stein, A. Panjkovich, P. Aloy, *Nucleic Acids Res.* 37 (2009) D300–D304.
- [91] M.D. Dyer, T.M. Murali, B.W. Sobral, *Bioinformatics* 23 (2007) i159–i166.
- [92] J.-G. Kim, D. Park, B.-C. Kim, S.-W. Cho, Y.T. Kim, Y.-J. Park, H.J. Cho, H. Park, K.-B. Kim, K.-O. Yoon, S.-J. Park, B.-M. Lee, J. Bhak, *BMC Bioinformatics* 9 (2008) 41.
- [93] H. Dinkel, S. Michael, R.J. Weatheritt, N.E. Davey, K.V. Roey, B. Altenberg, G. Toedt, B. Uyar, M. Seiler, A. Budd, L. Jödicke, M.A. Dammert, C. Schroeter, M. Hammer, T. Schmidt, P. Jehl, C. McGuigan, M. Dymecka, C. Chica, K. Luck, A. Via, A. Chatr-Aryamontri, N. Haslam, G. Grebnev, R.J. Edwards, M.O. Steinmetz, H. Meiselbach, F. Diella, T.J. Gibson, *Nucleic Acids Res.* 40 (2012) D242–D251.
- [94] N.E. Davey, G. Travé, T.J. Gibson, *Trends Biochem. Sci.* 36 (2011) 159–169.
- [95] B. Ravi Chandra, R. Gowthaman, R. Raj Akhouri, D. Gupta, A. Sharma, *Protein Eng Des Sel* 17 (2004) 175–182.
- [96] O. Aitio, M. Hellman, A. Kazlauskas, D.F. Vingadassalom, J.M. Leong, K. Saksela, P. Permi, *Proc. Natl. Acad. Sci. USA* 107 (2010) 21743–21748.
- [97] W. Dampier, P. Evans, L. Ungar, A. Tozeren, *BMC Med. Genomics* 2 (2009) 47.
- [98] T. Pawson, P. Nash, *Science* 300 (2003) 445–452.
- [99] R. Tonikian, X. Xin, C.P. Toret, D. Gfeller, C. Landgraf, S. Panni, S. Paoluzi, L. Castagnoli, B. Currell, S. Seshagiri, H. Yu, B. Windsor, M. Vidal, M.B. Gerstein, G.D. Bader, R. Volkmer, G. Cesareni, D.G. Drubin, P.M. Kim, S.S. Sidhu, C. Boone, *PLoS Biol.* 7 (2009) e1000218.
- [100] C. Wu, M.H. Ma, K.R. Brown, M. Geisler, L. Li, E. Tzeng, C.Y.H. Jia, I. Jurisica, S.S.-C. Li, *Proteomics* 7 (2007) 1775–1785.
- [101] M. Rodriguez, S.S.-C. Li, J.W. Harper, Z. Songyang, *J. Biol. Chem.* 279 (2004) 8802–8807.
- [102] A. Espejo, J. Côté, A. Bednarek, S. Richard, M.T. Bedford, *Biochem J* 367 (2002) 697–702.
- [103] C. Landgraf, S. Panni, L. Montecchi-Palazzi, L. Castagnoli, J. Schneider-Mergener, R. Volkmer-Engert, G. Cesareni, *PLoS Biol.* 2 (2004) E14.
- [104] J.C. Obenaus, L.C. Cantley, M.B. Yaffe, *Nucleic Acids Res.* 31 (2003) 3635–3641.
- [105] H.Y.K. Lam, P.M. Kim, J. Mok, R. Tonikian, S.S. Sidhu, B.E. Turk, M. Snyder, M.B. Gerstein, *BMC Bioinformatics* 11 (2010) 243.
- [106] M.S. Mukhtar, A.-R. Carvunis, M. Dreze, P. Epple, J. Steinbrenner, J. Moore, M. Tasan, M. Galli, T. Hao, M.T. Nishimura, S.J. Pevzner, S.E. Donovan, L. Ghamisari, B. Santhanam, V. Romero, M.M. Soulin, F. Gebreab, B.J. Gutierrez, S. Tam, D. Monachello, M. Boxem, C.J. Harbort, N. McDonald, L. Gai, H. Chen, Y. HeE.U.E.C., J. Vaaaandenhaute, F.P. Roth, D.E. Hill, J.R. Ecker, M. Vidal, J. Beynon, P. Braun, J.L. Dangl, *Science* 333 (2011) 596–601.
- [107] T. Rattei, P. Tischler, S. Gotz, M.A. Jehl, J. Hoser, R. Arnold, A. Conesa, H.W. Mewes, *Nucleic Acids Res.* 38 (2010) D223–D226.
- [108] P. Dean, *FEMS Microbiol. Rev.* 35 (2011) 1100–1125.
- [109] C.V. Srikanth, R. Mercado-Lubo, K. Hallstrom, B.A. McCormick, *Cell. Mol. Life Sci.* 68 (2011) 3687–3697.
- [110] G. Buchwald, A. Friebe, J.E. Galán, W.-D. Hardt, A. Wittinghofer, K. Scheffzek, *EMBO J.* 21 (2002) 3286–3295.
- [111] E.J. McGhie, R.D. Hayward, V. Koronakis, *Mol. Cell* 13 (2004) 497–510.
- [112] D. Zhou, M.S. Moosker, J.E. Galán, *Science* 283 (1999) 2092–2095.
- [113] I. Levin, C. Eakin, M.-P. Blanc, R.E. Klevit, S.I. Miller, P.S. Brzovic, *Proc. Natl. Acad. Sci. USA* 107 (2010) 2848–2853.

- [114] Y. Fu, J.E. Galán, *Nature* 401 (1999) 293–297.
- [115] M.B. Ohlson, Z. Huang, N.M. Alto, M.-P. Blanc, J.E. Dixon, J. Chai, S.I. Miller, *Cell Host Microbe* 4 (2008) 434–446.
- [116] M.B. Ohlson, K. Fluhr, C.L. Birmingham, J.H. Brumell, S.I. Miller, *Infect. Immun.* 73 (2005) 6249–6259.
- [117] E.V. Koonin, K.S. Makarova, L. Aravind, *Annu. Rev. Microbiol.* 55 (2001) 709–742.
- [118] H. Nagai, J.C. Kagan, X. Zhu, R.A. Kahn, C.R. Roy, *Science* 295 (2002) 679–682.
- [119] C.E. Stebbins, J.E. Galán, *Nature* 412 (2001) 701–705.
- [120] M. Eugster, C.-A.H. Roten, G. Greub, *BMC Evol. Biol.* 7 (2007) 231.
- [121] C.P. Ponting, L. Aravind, J. Schultz, P. Bork, E.V. Koonin, *J. Mol. Biol.* 289 (1999) 729–745.
- [122] E. Degtyar, T. Zusman, M. Ehrlich, G. Segal, *Cell. Microbiol.* 11 (2009) 1219–1235.
- [123] P. Mazodier, J. Davies, *Annu. Rev. Genet.* 25 (1991) 147–171.
- [124] C.M. Thomas, K.M. Nielsen, *Nat. Rev. Microbiol.* 3 (2005) 711–721.
- [125] G. Gimenez, C. Bertelli, C. Moliner, C. Robert, D. Raoult, P.-E. Fournier, G. Greub, *BMC Genomics* 12 (2011) 542.
- [126] N. Rappoport, M. Linial, *PLoS Comput. Biol.* 8 (2012) e1002364.
- [127] H. Dutartre, M. Harris, D. Olive, Y. Collette, *Virology* 247 (1998) 200–211.
- [128] R.T. Javier, A.P. Rice, *J. Virol.* 85 (2011) 11544–11556.
- [129] R.G. Ptak, W. Fu, B.E. Sanders-Beer, J.E. Dickerson, J.W. Pinney, D.L. Robertson, M.N. Rozanov, K.S. Katz, D.R. Maglott, K.D. Pruitt, C.W. Dieffenbach, *AIDS Res. Hum. Retroviruses* 24 (2008) 1497–1502.
- [130] W. Fu, B.E. Sanders-Beer, K.S. Katz, D.R. Maglott, K.D. Pruitt, R.G. Ptak, *Nucleic Acids Res.* 37 (2009) D417–D422.
- [131] R. König, Y. Zhou, D. Elleder, T.L. Diamond, G.M.C. Bonamy, J.T. Irelan, C.-Y. Chiang, B.P. Tu, P.D. De Jesus, C.E. Lilley, S. Seidel, A.M. Opaluch, J.S. Caldwell, M.D. Weitzman, K.L. Kuhen, S. Bandyopadhyay, T. Ideker, A.P. Orth, L.J. Miraglia, F.D. Bushman, J.A. Young, S.K. Chanda, *Cell* 135 (2008) 49–60.
- [132] J.-C. Rain, A. Cribier, A. Gérard, S. Emiliani, R. Benarous, *Methods* 47 (2009) 291–297.
- [133] A. Allouch, A. Cereseto, *Amino Acids* 41 (2011) 1137–1145.
- [134] B. de Chassey, V. Navratil, L. Tafforeau, M.S. Hiet, A. Aublin-Gex, S. Agaogué, 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. André, C. Rabourdin-Combe, V. Lotteau, *Mol. Syst. Biol.* 4 (2008) 230.
- [135] L.P. Tripathi, C. Kataoka, S. Taguwa, K. Moriishi, Y. Mori, Y. Matsuura, K. Mizuguchi, *Mol. Biosyst.* 6 (2010) 2539–2553.
- [136] S.D. Shapira, I. Gat-Viks, B.O.V. Shum, A. Dricot, M.M. de Grace, L. Wu, P.B. Gupta, T. Hao, S.J. Silver, D.E. Root, D.E. Hill, A. Regev, N. Hacohen, *Cell* 139 (2009) 1255–1267.
- [137] S. Pfefferle, J. Schöpf, M. Kögl, C.C. Friedel, M.A. Müller, J. Carbajo-Lozoya, T. Stellberger, E. von Dall'Armi, P. Herzog, S. Kallies, D. Niemeyer, V. Ditt, T. Kuri, R. Züst, K. Pumpor, R. Hilgenfeld, F. Schwarz, R. Zimmer, I. Steffen, F. Weber, V. Thiel, G. Herrler, H.-J. Thiel, C. Schwegmann-Wessels, S. Pöhlmann, J. Haas, C. Drosten, A. von Brunn, *PLoS Pathog.* 7 (2011) e1002331.
- [138] P. Uetz, Y.-A. Dong, C. Zeretzke, C. Atzler, A. Baiker, B. Berger, S.V. Rajagopala, M. Roupelieva, D. Rose, E. Fossum, J. Haas, *Science* 311 (2006) 239–242.
- [139] E. Fossum, C.C. Friedel, S.V. Rajagopala, B. Titz, A. Baiker, T. Schmidt, T. Kraus, T. Stellberger, C. Rutenberg, S. Suthram, S. Bandyopadhyay, D. Rose, A. von Brunn, M. Uhlmann, C. Zeretzke, Y.-A. Dong, H. Boulet, M. Koegl, S.M. Bailer, U. Koszinowski, T. Ideker, P. Uetz, R. Zimmer, J. Haas, *PLoS Pathog.* 5 (2009) e1000570.
- [140] Khadka, S., Vangeloff, A.D., Zhang, C., Siddavatam, P., Heaton, N.S., Wang, L., Sengupta, R., Sahasrabudhe, S., Randall, G., Gribskov, M., Kuhn, R.J., Perera, R., and LaCount, D.J. *Mol Cell Proteomics*, 10, (2011) M111.012187.
- [141] M.A. Calderwood, 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, *Proc. Natl. Acad. Sci. USA* 104 (2007) 7606–7611.
- [142] M.D. Dyer, C. Neff, M. Dufford, C.G. Rivera, D. Shattuck, J. Bassaganya-Riera, T.M. Murali, B.W. Sobral, *PLoS One* 5 (2010) e12089.
- [143] V. Navratil, B. de Chassey, L. Meyniel, F. Pradezynski, P. André, C. Rabourdin-Combe, V. Lotteau, *J. Proteome Res.* 9 (2010) 3527–3536.
- [144] D. van Dijk, G. Ertaylan, C.A. Boucher, P.M. Sloot, *BMC Syst. Biol.* 4 (2010) 96.
- [145] J.I. MacPherson, J.E. Dickerson, J.W. Pinney, D.L. Robertson, *PLoS Comput. Biol.* 6 (2010) e1000863.
- [146] M. Sarmady, W. Dampier, A. Tozeren, *PLoS One* 6 (2011) e23293.
- [147] S. Wuchty, *BMC Genomics* 7 (2006) 122.
- [148] M. Sarmady, W. Dampier, A. Tozeren, *PLoS One* 6 (2011) e20735.
- [149] L.T. Vassilev, *Cell Cycle* 3 (2004) 419–421.
- [150] L. Li, R.M. Thomas, H. Suzuki, J.K. De Brabander, X. Wang, P.G. Harran, *Science* 305 (2004) 1471–1474.
- [151] Y. Zhang, B.A. Appleton, C. Wiesmann, T. Lau, M. Costa, R.N. Hannoush, S.S. Sidhu, *Nat. Chem. Biol.* 5 (2009) 217–219.
- [152] K. Thorlund, T. Awad, G. Boivin, L. Thabane, *BMC Infect. Dis.* 11 (2011) 134.
- [153] S.-L. Tan, G. Ganji, B. Paepier, S. Proll, M.G. Katze, *Nat. Biotechnol.* 25 (2007) 1383–1389.
- [154] T. Pawson, R. Linding, *FEBS Lett.* 582 (2008) 1266–1270.