Protein-protein interactions: network analysis and applications in drug discovery.

Running title: PPI networks and drug discovery

Jennyfer Bultinck\textsuperscript{1,2}, Sam Lievens\textsuperscript{1,2} & *Jan Tavernier\textsuperscript{1,2}

\textsuperscript{1}Department of Medical Protein Research, VIB, A. Baertsoenkaai 3, 9000 Ghent, Belgium
\textsuperscript{2}Department of Biochemistry, Faculty of Medicine and Health Sciences, Ghent University, A. Baertsoenkaai 3, 9000 Ghent, Belgium

* Corresponding author: Jan Tavernier
e-mail: jan.tavernier@vib-ugent.be
tel: +32 9 264 93 02
fax: +32 9 264 94 92
Abstract

Physical interactions among proteins constitute the backbone of cellular function, making them an attractive source of therapeutic targets. Although the challenges associated with targeting protein-protein interactions (PPIs) -in particular with small molecules - are considerable, a growing number of functional PPI modulators is being reported and clinically evaluated. An essential starting point for PPI inhibitor screening or design projects is the generation of a detailed map of the human interactome and the interactions between human and pathogen proteins. Different routes to produce these biological networks are being combined, including literature curation and computational methods. Experimental approaches to map PPIs mainly rely on the yeast two-hybrid (Y2H) technology, which have recently shown to produce reliable protein networks. However, other genetic and biochemical methods will be essential to increase both coverage and resolution of current protein networks in order to increase their utility towards the identification of novel disease-related proteins and PPIs, and their potential use as therapeutic targets.

Keywords: protein-protein interaction, interactome, drug discovery, small molecule, oncology, virology.
1. Introduction

Despite huge academic and industrial efforts, many disease areas still face considerable unmet medical needs, which can be attributed to the disappointing output of the pharmaceutical industry with only around 5-7 mechanistically novel small molecule therapeutics being approved each year [1]. The 1,204 small molecule drugs that are already on the market today act through only 248 distinct molecular targets, with half of them targeting only 3 protein families: GPCRs, ion channels and nuclear receptors [2]. Although there clearly is an urgent need for more innovative targets, the opportunities were suggested to be limited based on the estimation that there are as few as 600 disease-associated protein targets susceptible to interference by Lipinski rule-of-5 drugs [3]. However, this point of view might be too pessimistic since the pharmaceutical industry now has small molecule therapeutics in the pipeline against targets that were previously categorized as undruggable. The best example of such novel drug targets are PPIs.

Interactions between proteins mediate their function and play a key role in virtually every cellular process. The interface of two interacting proteins mostly consists of a large, flat and non-continuous surface without discrete binding pockets and is shielded by solvent molecules and ions [4]. Therefore, it seemed unlikely that these interfaces could be disrupted by the classical small molecule drugs. With the identification of ‘hotspots’, compact, centralized regions of residues within the PPI contact surface that account for most of the binding energy [5, 6], and the detection of grooves in certain conformations of adaptive protein surfaces [7], it became apparent that PPIs still might be valid targets for small molecule-based therapeutic intervention. In addition, the interaction between 2 proteins can also be blocked through allosteric sites, or by interfering with an activity that is essential for the PPI to form. AG-1478, for example, indirectly interferes with EGFR-p85 interaction by inhibiting EGFR kinase activity which is critical for the phosphorylation-dependent binding of p85 to EGFR [8]. The apparent druggability of PPIs opens up a wide range of opportunities to treat disease. As the estimated number of PPIs among human proteins ranges between 130,000 [9] and 650,000 [10], this significantly expands the target portfolio of pharmaceutical R&D, even when only a small fraction of these interactions were disease relevant. Particularly in areas where the number of classical single protein targets is limited, like virology and bacteriology, this may turn out to be of vital importance, as will be discussed later. Moreover, it allows developing drugs with improved specificity and fewer side effects. In contrast to the conserved active site of enzymes that is often hit by ‘classical’ drugs, PPIs contain very specific binding pockets and recognition sites due to their structural and conformational diversity, allowing the development of very specific drugs. Furthermore, compared to drugs targeting single protein, additional specificity can be obtained by disrupting specifically a disease-related PPI, while keeping the protein’s interactions with substrates that are required for normal cellular functions intact.

In the past decade, R&D experienced a boost in the search for modulators of PPIs with almost every pharmaceutical company having PPI targeted programs. As a result, various drugs that modulate extracellular PPIs are now on the market. Several biopharmaceuticals act by antagonizing the interaction between receptors and their (protein) ligand, the best known probably being herceptin (trastuzumab), an antibody that binds to the extracellular domain of the human EGFR2 (HER2).
Herceptin was approved by the FDA in 1998 for use against HER2-positive metastatic breast cancer [11]. Other examples are antibodies that inhibit the binding of TNFα or IL1β to their receptor for treatment of autoimmune and autoinflammatory disorders, respectively (e.g. infliximab and adalimumab for TNFα; canakinumab for IL1β), antibodies interfering with EGF-EGFR1 (e.g. cetuximab and panitumumab) or VEGF-VEGFR (e.g. bevacizumab) binding for application in metastatic colorectal cancer [11]. An important drawback of such biopharmaceuticals compared to small molecules is that they cannot be administrated orally because they are degraded in and poorly absorbed by the gastrointestinal tract. They are generally also more expensive. Therefore, efforts are made to screen for small molecular functional analogues of such biopharmaceutical drugs (e.g. VEGF-VEGFR1 [12]) as a cheaper alternative or for use in small molecule-antibody combination therapy aiming at higher therapeutic efficiency [13]. Although small molecules can be developed that are cell permeable and should therefore be able to target PPIs inside the cell, most FDA-approved small molecule PPI inhibitors inhibit the interaction of extracellular proteins. Examples are the anti-HIV drug maraviroc that allosterically inhibits the interaction between the viral gp120 protein and the human CCR5 receptor [14]; tirofiban that antagonizes fibrinogen binding to the glycoprotein IIb/IIIa receptor and is used as a platelet aggregation inhibitor; and degarelix, an inhibitor of the interaction between gonadotropin-releasing hormone and its receptor for the treatment of prostate cancer [11]. Since quite some intracellular small molecule PPI inhibitors are currently being clinically evaluated, predominantly in the oncology field, and more should soon enter phase I trials based on their thorough pre-clinical validation, it is likely just a matter of time before some of them will gain FDA approval. In the next section, we will give an update of the intracellular PPI inhibitor compounds that are presently progressing through the (pre-) clinical pipeline.

2. Small molecule PPI inhibitors in the drug discovery pipeline

Targeting oncology-related PPIs

In PPI-based drug development, the inhibition of the p53-MDM2 interaction gained a lot of attention, as evidenced by the number of small molecule disruptors that have been developed so far as a strategy to treat cancer [15]. MDM2 binds and negatively regulates the p53 tumor suppressor protein and is overexpressed in many human tumors. The nutlin (imidazoline derivatives) are by far the best characterized p53-MDM2 disruptors and displace p53 from the MDM2 protein with an IC50 in the low nanomolar range (e.g. nutlin-3 IC50 = 90 nM, Figure 1) [16]. Roche plans to bring the nutlins on the market as RG-7112, which is currently in phase I clinical trials for solid tumors and hematological malignancies [17]. p53-MDM2 antagonists developed by a number of other companies like J&J, Boehringer Ingelheim and Sanofi Aventis have also reached the clinical trial stage [17]. Because the MDM2-related protein MDM4 exerts complementary but distinct functions in the p53 response in tumor cells, Reed et al. developed a high throughput (HT) Fluorescence Polarization Assay (FPA) and identified the first p53-MDM4 small molecule inhibitor, SJ-172550, which was shown to efficiently kill MDM4-amplified retinoblastoma cells. This tumoricidal effect is p53-dependent and additive when combined with nutlin-3a. SJ-172550 requires further optimization to improve its potency as the binding constant is only in the micromolar range [18].
Proteins of the BCL2 family are involved in tumor development and resistance to anticancer therapies [19]. Members of this family can be functionally classified as either anti-apoptotic (BCL2, BCL-xL, BCL-w, MCL1 and A1) or pro-apoptotic (effectors BAK and BAX; BH3-only direct activators BIM and BID; and BH3-only sensitizers BAD, BIK, PUMA, NOXA, HRP, BMF). Small molecule BH3 mimetics were developed that would displace the anti-apoptotic BCL2 members from the pro-apoptotic effectors, in a similar way as the BH3-only members do. This ultimately results in cell death due to homo-oligomerization of the pro-apoptotic effectors, pore formation in the outer mitochondrial membrane and subsequent mitochondrial cytochrome C release. Three such BH3 mimetics, navitoclax (ABT-263, Abbott), AT-101 (Ascenta Therapeutics) and obatoclax (GX15-070, GeminX), are currently in phase II trials for a number of cancer-related indications [20, 21]. The structure of these compounds can be found in Figure 2. ABT-263 antagonizes BCL2, BCL-xL and BCL-w but not MCL1, which is a major flaw of this drug as there is a direct correlation between MCL1 expression and resistance to the drug. Interestingly, a recent publication by Albershardt and colleagues showed that other BH3 mimetics activate the ER stress response rather than inhibiting the interaction between BCL2 and pro-apoptotic members, as ABT-263 does. Because this stress response results in NOXA-mediated inhibition of MCL1, combining ABT-263 with another BH3 mimetic could bypass the pharmaceutical R&D programs to develop MCL1-specific BH3-mimetics to overcome resistance to ABT-263 [22].

Figure 2. BH3 mimetics in phase II clinical trials

Another means of targeting the apoptotic pathway in cancer cells is by mimicking the binding of the pro-apoptotic protein Smac/DIABLO to IAP family members, an interaction that results in the release of cell death-inducing caspases. Tumor cells overexpress IAP proteins to protect themselves against a number of pro-apoptotic stimuli, including chemotherapeutics, and this correlates with bad outcome in a variety of tumors. Several pharmaceutical companies developed small molecule analogues of the IAP-binding motif of Smac/DIABLO, four of which have currently entered phase I
clinical trials: GDC-0152 (Genentech), LCL161 (Novartis), AEG40826/HGS1029 (Aegera Therapeutics/Human Genome Sciences) and AT-406 (Ascenta Therapeutics) [23].

Constitutive activation of the WNT pathway is found in many human malignancies. By maintaining the stem cell properties of cancer cells, WNT is involved in different phases of cancer progression like tumor initiation, metastasis, recurrence and chemoresistance. So, this signaling pathway is an attractive anti-oncogenic target and several WNT-specific PPI inhibitors have already been identified and pre-clinically validated [24]. In vitro HT screens (HTS) for disruptors of the interaction between β-catenin and the TCF transcription factor identified two small molecules, PKF115-584 and CGP049090 (Figure 3), that were later shown to suppress the growth of cancer cells in vitro as well as in animal models [24]. With the cell-based TOPFLASH reporter assay, Emami et al. screened for small molecule inhibitors of the WNT pathway and found ICG-011 that specifically disrupts the interaction of β-catenin with CBP but not with the highly homologous co-activator p300. Min (multiple intestinal neoplasia)-mice developed less intestinal adenomas when treated with ICG-001, and the compound also reduced the tumor volume in a xenograft colon cancer model [25]. A more potent, specific CBP-β-catenin antagonist, PRI-724, was later developed by Prism Biolab and a phase I clinical trial in patients with advanced solid tumors was opened beginning 2011 [21].

Figure 3. Small-molecule disruptors of the interaction between TCF and β-catenin

The transcription factor cMyc drives cell proliferation, regulates apoptosis, determines the differentiation state of the cell and is overexpressed in many types of cancers. Because cMyc requires heterodimerization with Max for most of these functions, several groups have developed cMyc-Max antagonists as a strategy to treat cancer patients, but none of them reached clinical evaluation due to their inadequate potency and selectivity (reviewed in [26]).

The interaction between the well known Retinoblastoma tumor suppressor, Rb, and the proto-oncogene RAF1 was put forward as a promising target for cancer treatment [27]. An ELISA-based HTS identified RRD-251 (Figure 4A) as a potent and selective small molecule disruptor of Rb-RAF1 interaction. Its potential as anti-cancer drug was underscored by the fact that peroral treatment of tumor-bearing xenografted mice significantly suppressed tumor growth by inhibiting angiogenesis and cell proliferation [28].

Other small molecule PPI inhibitors that were shown to have potent anti-tumor effects in human cancer cell lines and in animal models are chetomin (Figure 4B) and its analogues that inhibit
the Hif1α-p300 interaction [29, 30]; STX-0119 (Figure 4C), a STAT3 dimerization inhibitor [31]; EPI-001 that attenuates androgen-stimulated N/C interaction of and CBP interaction with the androgen receptor [32]; and two independently identified small molecule disruptors of eIF4E-eIF4G, 4EGI-1 (Figure 4D) [33] and 4E1RCat (Figure 4E) [34].

Finally, Zhao and coworkers picked up FOBISIN101 (Figure 4F) from a fluorescence polarization-based 14-3-3 binding assay. This compound might have therapeutic potential in several pathologies, including cancer. Although the authors do not functionally validate this compound, they did a remarkable finding when resolving the crystal structure of the 14-3-3 protein in complex with FOBISIN101. The X-rays caused the compound to become covalently attached to 14-3-3, leading to the persistent inhibition of the latter. It was subsequently suggested that FOBISIN101 could be used as a radiation-triggered therapeutic agent [35].

Figure 4. Small-molecule PPI inhibitors with anti-tumor effects

PPI inhibitors for the treatment of infectious diseases

Due to the limited size of viral proteomes, targeted approaches at the level of single viral proteins are rapidly exhausted and targeting PPIs, especially those with host proteins, provides a critically important alternative. Moreover, current anti-viral drugs often lose their effectiveness due to the selection of drug-resistant mutants, which could be circumvented with drugs that act through host factors that bind essential viral proteins, because then both virus and host are under chemotherapeutic pressure. Another advantage is that such virus-host inhibitory therapy is probably less toxic because it doesn’t interfere with essential host-specific PPIs.
Indandiones and repaglinide are disruptors of the interaction between the E1 helicase domain and the transactivation domain of E2, both proteins from human papillomavirus (HPV), identified through \textit{in vitro} screens performed by Boehringer Ingelheim. These compounds could indeed inhibit HPV genome replication \textit{in vivo} and are thus potential intraviral PPI inhibitory drugs for the treatment of HPV-associated benign lesions [36, 37]. The therapeutic potential of these compounds is however restricted because the expression of E1 and E2 is often lost during malignant transformation, this is when the viral episome is integrated into the host genome. In line with this, care should also be taken when the interaction of HPV E2 with host Brd4 is envisaged as a target for the treatment of HPV infection as was recently suggested, because of the crucial role of this PPI for viral episome maintenance and transcriptional activation [38]. Indeed, E2-Brd4 disruptors might enhance tumor development due to the repressive effect of the E2-Brd4 interaction on the expression of E6 and E7 oncogenes from the integrated viral genome [39, 40].

LEDGIns are rationally designed 2-(quinolin-3-yl)acetic acid derivatives that disrupt the interaction between host LEDGF and HIV1 integrase and as such block viral integration since LEDGF is required for tethering the preintegration complex to the chromatin [41]. By measuring the levels of p24, a protein that is part of the HIV1 capsid, it was shown that LEDGIns inhibit HIV1 replication in human PBMC and macrophages at submicromolar concentration [41]. Also Boehringer Ingelheim reported anti-HIV activity of 2-(quinolin-3-yl)acetic acid derivatives in a patent application, without however elucidating the mechanism of action [42].

Another attractive target for anti-HIV treatment is the human Apobec3G (hA3G)-HIV1 Vif interaction. G-to-A hypermutation induced by hA3G in newly synthesized viral DNA inhibits viral replication. Upon interaction with hA3G, HIV1 escapes the immune response by the Vif-mediated degradation of hA3G. Two independent groups established a HT cell-based assay using YFP-fused hA3G as a reporter to identify inhibitors of Vif-mediated hA3G degradation. Nathans \textit{et al.} found a small molecule, RN-18, that prevents hA3G degradation in a Vif-dependent manner. This compound promotes cytidine deamination of the viral genome and by increasing hA3G incorporation into virions, it reduces the replication and infectivity of HIV1 with an IC$_{50}$ in the micromolar range. However, the authors did not address whether RN-18 directly affects the hA3G-Vif interaction nor did they test its \textit{in vivo} safety [43]. Later, Cen \textit{et al.} discovered two other compounds, IMB-26 and IMB-35, that were 100-fold more potent against HIV1 infection of hA3G-expressing human T cell lines compared to RN-18. Different PPI detection assays show that both compounds directly bind to hA3G and thereby prevent the interaction with HIV1 Vif. Importantly, no cellular toxicity could be observed, not even with a 200-fold higher concentration than required for an efficient antiviral response. Moreover, doses up to 1 g/kg injected intraperitoneally into mice do not cause significant weight loss or toxicity in any of the tested organs (liver, kidney, heart and lungs) [44].

Coronaviruses, a family of enveloped single-stranded RNA viruses, cause gastroenteritis and are also believed to be responsible for a lot of common colds in human adults. Because the translation of coronavirus RNA is cap-dependent and requires eIF4F, a HT time resolved-FRET (Forster Resonance Energy Transfer) screen was developed by Cencic and colleagues to discover compounds that block the interaction of two eIF4F subunits, eIF4E (the cap-binding protein) and eIF4G (a large
scaffolding protein). One of the identified small molecules, 4E2RCat, significantly reduced the number of human embryonic pulmonary epithelial cells infected with coronavirus 229E. Both the intra- and extracellular production of virus was prevented in a dose- and time-dependent manner, with hardly any detectable extracellular virus at 48 h after infection when the cells were treated with 6.25 µM 4E2RCat. There is however room to improve its potency as the compound interferes with only four out of the five ‘hot spots’ present in the eIF4G binding region [45]. Although the efficacy of this eIF4F disruptor in animal infection models remains to be determined, it is known from other studies that targeting this host factor is well tolerated in vivo [34, 46, 47].

Due to the increasing prevalence of antibiotic-resistant bacteria as well as the very low output of novel antibacterial drug classes over the past 25 years, bacterial infections become increasingly difficult to treat [48]. Interactions among bacterial proteins and, even more so, between host proteins and bacterial virulence factors are promising but underexplored new therapeutic targets. This is especially valid for Gram-negative pathogenic bacteria for which only a few potential drugs are currently in clinical development [49]. It is extremely challenging to obtain sufficient antibiotics inside Gram-negative bacteria because of the additional outer membrane and the efflux pumps. In addition, Gram-negative bacterial enzymes modify both drug and target as a means to reduce the bactericidal potential [48]. These limitations could be circumvented by targeting the interaction of secreted bacterial virulence factors with proteins of the host cell; such drugs would indirectly kill the bacterial pathogen, for example through a more efficient host immune response. An important bottleneck here is the poorly elucidated intra-bacterial as well as host-bacterial PPI networks, as will also be discussed later. It is not surprising that none of the antibiotics in clinical trials in 2011 are targeting PPIs [49], as only one reported HTS aimed at identifying bacterial PPI disruptors [50]. The essential role of RNA polymerase for bacterial growth, the conservation of the interface between σ factors and the core polymerase in bacteria, and the absence of σ factors in mammalian cells motivated André and coworkers to identify specific inhibitors of core–σ interaction. Using an in vitro HT ELISA, they identified synthetic small molecule compounds (SB2 and SB8) that inhibit this interaction with an IC₅₀ in the low micromolar range. These molecules were shown to be bactericidal for a wide range of Gram-positive bacteria such as (multidrug-resistant) *S. aureus, S. epidermidis, B. cereus, B. anthracis, S. pneumoniae* as well as some Gram-negative bacteria such as *Bacteroides* spp, but were unfortunately inactive against *E. coli* or *P. aeruginosa* due to low bacterial entry of the compounds [50, 51].

**Other therapeutic applications of PPI inhibitors**

The TLR pathway is a crucial component of the immune response against pathogenic organisms but when it is not tightly regulated, chronic and systemic inflammatory diseases might develop. TAK-242 (Resatorvird), a small molecule disruptor of the interaction between TLR4 and its adaptor proteins TIRAP or TRAM [52], reached phase III clinical trials for severe sepsis but the study was ended after the DSMB (Data monitoring committee, an independent group of experts who monitor patient safety and treatment efficacy data during a clinical trial) determined there was insufficient cytokine suppression in Resatorvird-treated patients. Also the phase III trial for sepsis-induced cardiovascular
and respiratory failure was terminated, which was based on economical considerations rather than safety or efficacy concerns [21].

JNK1-JIP1 interaction was put forward as a promising target to treat insulin resistance in type II diabetes patients [53, 54]. Two recent publications describe small-molecule antagonists that target this PPI. Chen et al. identified a cell-impermeable compound that selectively inhibited the JIP1 binding to JNK1 with an IC₅₀ value of 5 µM [55]. The other compound, BI-78D3, has more potential as it selectively inhibits the binding of a JIP-derived peptide to JNK1 with an IC₅₀ of 0.5 µM. More importantly, BI-78D3 was shown to abrogate ConA-induced liver failure as measured by blood levels of the liver enzyme alanine-aminotransferase. In addition, inhibition of JNK1-JIP1 could restore insulin sensitivity as evidenced by the reduction in blood glucose a few hours after insulin injection of BI-78D3 (25 mg/kg)-treated insulin-insensitive mice [56].

Cystic fibrosis is an inherited disease characterized by the secretion of abnormally thick mucus by several organs like lungs and pancreas due to aberrant activity of CFTR Cl-channels. Using a HT alpha screen assay, Zhang et al. identified a small molecule, CO-068, that disrupts the LPA2–NHERF2 interaction within CFTR macromolecular complexes with an IC₅₀ of 63 µM. The authors showed that this compound augments CFTR Cl-channel activity in Calu-3 cells and fluid secretion from pig tracheal submucosal glands with no apparent cellular toxicity or morphological changes, and encourage the development of more potent inhibitors for this PPI for the treatment of cystic fibrosis patients [57].

The underlying mechanism of neuropathic pain, characterized by hyperalgesia, alldynia and spontaneous pain, is the NMDA-induced excessive NO production by nNOS. Because PSD95 recruits nNOS to the NMDA receptor, it was reasoned that interfering with the nNOS-PSD95 binding would disrupt downstream pain hypersensitivity. Using an in vitro HTS, Florio et al. identified IC87201 as a small molecule disruptor of the nNOS-PSD95 interaction. Intrathecal administration potently reversed NMDA-induced thermal hyperalgesia in mice without any effect on acute pain thresholds or motor coordination, and abolished mechanical alldynia in the Chronic Constriction Injury-model of neuropathic pain [58].

3. Filling the pipeline of PPI modulatory drugs

As illustrated in the previous paragraphs, judging from the flow of reports identifying novel small molecule PPI modulators, there is a very strong interest in this emerging target class. As a number of these molecules are currently proceeding through clinical evaluation, making a case for the druggability of PPIs, this interest is expected to further grow in the coming years. Additional progress in the field will depend on a number of critical factors, including the development of better screening technologies, the design of improved PPI-focused chemical libraries and, most importantly, the availability of more comprehensive interactome maps for the identification of disease-related PPIs.

Several PPI inhibitor screening approaches are being pursued: small scale screening of target-specific compound sets rationally designed or selected in silico, fragment-based screening or HTS screening of generic PPI-focused libraries or large diversity-oriented small molecule collections [59]. To date, HTS mainly relied on the use of in vitro assays such as FPA, surface plasmon resonance or NMR. More recently also cellular assays based on the 2-hybrid concept became available that are
compatible with the scale required for HTS [60, 61]. In these assays, two interacting proteins of interest are fused to complementary fragments of a reporter protein (e.g. transcription factor, luciferase, β-lactamase), the reconstitution of which will be prevented when the cell is treated with a compound that disrupts the interaction between those chimeric proteins. Interestingly, these assays offer unique advantages associated with the fact that they operate in intact cells. Target protein folding and post-translational modifications have greater chance of being carried out properly in the close-to-normal physiological context of a (mammalian) cell. Additionally, the compounds that are being screened for PPI inhibition are at the same time evaluated for membrane permeability and toxicity, and the effect of any off-target interaction on the potency of the compound is being assayed as well. Our group has been active in this field through the development of the MAPPIT technology platform, a two-hybrid approach based on reconstitution of a cytokine receptor signaling pathway that operates in human cells. Recent application of our technology in HTS for PPI modulators has indeed confirmed the benefit of utilizing cellular screening approaches, as hits have been identified that could be functionally validated but which failed to score in in vitro PPI assays [61]. We anticipate that the importance of cell-based HTS technologies for the discovery of drugs targeting intracellular PPIs will increase considerably in the future.

A second major determining factor for successful PPI disruptor screening projects is the quality of the compound libraries being screened. Over the past decade, the molecular structure of the protein interfaces and of the small molecules that bind these interfaces have been analyzed in detail for a number of PPIs. As mentioned previously, this led to the acknowledgement of binding energy hot spots and adaptive binding pockets at the PPI interface, findings that restored confidence in PPIs as a druggable target class. Additionally, these studies suggested that interfering with PPIs might pose particular demands towards the physico-chemical properties of small molecules. This conclusion might explain the limited success of early PPI inhibitor screening campaigns which were often using compound collections that originated from other screening projects which required different chemistry. The fact that standard small molecule libraries are not fully adequate for targeting PPIs was also demonstrated by Pagliaro and co-workers, who found that only half of 19 validated PPI inhibitors was covered by the diversity space of three commercial databases [62]. Therefore, efforts are needed to enrich chemical libraries with compounds that meet the chemical requirements of PPI inhibitors. Target-specific compound libraries are known to efficiently improve hit rates while reducing the overall cost of experimental screening [63]; and they are believed to increase potency and/or specificity of the identified hits. Applied to PPI inhibitors, recent efforts to characterize the appropriate chemical space have resulted in a number of chemical descriptors for which drug-like PPI inhibitors seem to differ from other drugs [64, 65]. This work is reviewed elsewhere in this issue.

Since every PPI-targeted screening project starts with the identification, and in a later stage the validation as a therapeutic target, of a pair of interacting proteins, access to large resources of PPIs is of prime importance. However, capturing PPIs in protein networks or interactomes is a largely unfinished business. Whereas most progress has been made in yeast, the human interactome is obviously the most relevant network towards human medicine. Additionally, interspecies interactomes, such as those between human proteins and the proteins of human pathogens, harbor attractive targets for therapeutic
interacting proteins in which the interaction was identified (interaction experiment) initiative systemization of literature published data information than literature that therefore widely used to study intuitively databases have to be combined. Datasets (BioGRID) protein Interaction database (IntAct) the Database of Interacting (MIPS) protein interaction database (MINT) the Munich Information Center for Protein Sequence (MIPS) protein interaction database [73], the Biomolecular Interaction Network Database (BIND) [69], the Database of Interacting Proteins (DIP) [75], the Molecular INTeraction database (MINT) [70], the protein InterAction database (IntAct) [71, 85], the Biological General Repository for Interaction Datasets (BioGRID) [86, 87], and the Human Protein Reference Database (HPRD) [72, 88]. Although the latter was shown to be the most comprehensive database for human PPIs, interactions from all these databases have to be combined to obtain a more complete dataset [89]. Literature-curated datasets are intuitively assumed to be better than datasets generated with HT technologies [10, 74, 90-92] and are therefore widely used to study PPIs in human disease. However, Vidal and co-workers recently showed that, when generated using rigorous screening protocols, HT experimental PPIs can be of higher quality than literature-curated interactions [9, 93-95]. Although literature-curated maps benefit from information that is already available, this comes with certain limitations because the quality of published data varies enormously and lacks negative results [94, 96]. In an effort to improve on the systemization of literature-curated interactomes, the MIMIx (Minimal Information about a Molecular Interaction experiment) initiative was set up to obtain consistent information with regards to the species in which the interaction was identified, the methods applied and the identity and name of the interacting proteins [97]. Despite the significant contribution of curated small-scale experiments, the unbiased, systematic

4. Construction of the human interactome

Currently available human PPI maps have been generated through genetic methods such as Y2H for mapping binary interactions [66, 67], biochemical technologies to identify (in)directly interacting components of protein complexes [68], curation of published low-throughput experiments [69-75], or computational predictions using orthogonal information such as sequence similarity, gene-order conservation, phylogenetic profiling and expression data, which is exponentially growing since the advent of next generation sequencing (NGS) [76, 77]. A list of online available tools to explore interactomics maps can be found in Table 1.

<table>
<thead>
<tr>
<th>Tool</th>
<th>URL</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIANA</td>
<td><a href="http://sbi.imim.es/piana/">http://sbi.imim.es/piana/</a></td>
<td>[78]</td>
</tr>
<tr>
<td>NaViGaTOR</td>
<td><a href="http://ophid.utoronto.ca/navigator/">http://ophid.utoronto.ca/navigator/</a></td>
<td>[79]</td>
</tr>
<tr>
<td>VisANT</td>
<td><a href="http://visant.bu.edu">http://visant.bu.edu</a>.</td>
<td>[80]</td>
</tr>
<tr>
<td>Cytoscape:</td>
<td><a href="http://www.cytoscape.org/">http://www.cytoscape.org/</a></td>
<td>[81]</td>
</tr>
<tr>
<td></td>
<td><a href="http://wiki.thebiogrid.org/doku.php/biogridplugin2">http://wiki.thebiogrid.org/doku.php/biogridplugin2</a></td>
<td></td>
</tr>
<tr>
<td>OSPREY</td>
<td><a href="http://biodata.mshri.on.ca:80/osprey/servlet/Index">http://biodata.mshri.on.ca:80/osprey/servlet/Index</a></td>
<td>[82]</td>
</tr>
<tr>
<td>PINAT</td>
<td><a href="http://www.manit.ac.in/pinat">http://www.manit.ac.in/pinat</a></td>
<td>[84]</td>
</tr>
</tbody>
</table>

Table 1: online available tools to explore interactomics maps

Literature-curated PPI datasets include the Munich Information Center for Protein Sequence (MIPS) protein interaction database [73], the Biomolecular Interaction Network Database (BIND) [69], the Database of Interacting Proteins (DIP) [75], the Molecular INTeraction database (MINT) [70], the protein InterAction database (IntAct) [71, 85], the Biological General Repository for Interaction Datasets (BioGRID) [86, 87], and the Human Protein Reference Database (HPRD) [72, 88]. Although the latter was shown to be the most comprehensive database for human PPIs, interactions from all these databases have to be combined to obtain a more complete dataset [89]. Literature-curated datasets are intuitively assumed to be better than datasets generated with HT technologies [10, 74, 90-92] and are therefore widely used to study PPIs in human disease. However, Vidal and co-workers recently showed that, when generated using rigorous screening protocols, HT experimental PPIs can be of higher quality than literature-curated interactions [9, 93-95]. Although literature-curated maps benefit from information that is already available, this comes with certain limitations because the quality of published data varies enormously and lacks negative results [94, 96]. In an effort to improve on the systemization of literature-curated interactomes, the MIMIx (Minimal Information about a Molecular Interaction experiment) initiative was set up to obtain consistent information with regards to the species in which the interaction was identified, the methods applied and the identity and name of the interacting proteins [97]. Despite the significant contribution of curated small-scale experiments, the unbiased, systematic
and well-controlled nature of HTS makes them indispensable for the construction of the human interactome. While genetic HT assays like Y2H tend to detect transient binary protein interactions, biochemical mass spectrometry (MS)-based HTS rather identify stable, stoichiometric complexes [98]. Both approaches are thus complementary sources of physical connections within biological processes. Noteworthy is the renewed interest in capturing more transient and weak interactions by combining cross-linking and MS techniques due to recent technical advances in the field [99, 100]. Importantly, upon MS-based identification of the cross-linked peptides, affinity purification (AP)-MS can be used to identify the interacting components within larger protein complexes [101, 102]. Although less amenable to HT format for technical reasons, large-scale AP-MS studies have successfully been used to unravel human PPI (sub)networks. Ewing et al. mapped the interactome of functionally essential and known or suspected disease genes [68]. Later, MS-based human PPI networks were unveiled for chromatin remodeling[103], deubiquitinating enzymes [104], the autophagy system [105] and mitosis [106].

Many genetic interaction mapping technologies have been developed, including PCA (Protein-fragment Complementation Assay), RET (Resonance Energy Transfer), MAPPIT (Mammalian Protein-Protein Interaction Trap) and phage display (reviewed in [107, 108]). Still, the ‘classic’ Y2H system remains the leading technique for large-scale binary mapping, especially now that the long-lasting skepticism regarding the accuracy of HT-Y2H has been refuted [92]. Reinvestigation of a set of random protein interactions of the S. cerevisiae and human interactomes with orthogonal techniques confirmed the high quality of Y2H maps and indicated that the reported limited overlap between different Y2H maps is due to low sensitivity and not to low specificity [9]. To take into account the fact that every method has specific limitations, resulting in an intrinsic set of false positives and false negatives being detected, these studies made use of standardized species-specific positive (PRS) and negative reference sets (RRS). PRSs were selected from manually curated information taken from literature and databases, whereas RRSs contain randomly chosen protein pairs that are unlikely to interact [9, 109]. Based on experimental quality assessment through retesting, the group of Vidal developed a confidence score for binary protein interactions using a panel of orthogonal methods, including MAPPIT, PCA, LUMIER (Luminescence-based Mammalian IntERactome) and wNAPPA (Nucleic Acid Programmable Protein Array) [93]. Interestingly, any of these methods could detect between 20 and 35 % of the PRS when the number of false positives was kept at a minimum, and each technology identified different, only partially overlapping, subsets of PPIs [93]. Thus, in order to reach complete coverage of the human interactome, the parallel use of complementary screening methods will be required. The recently developed Stitch-Seq interactome mapping protocol that combines PCR stitching (placing a pair of interacting proteins on the same PCR amplicon) with NGS will drastically increase the throughput and efficiency of interactome mapping projects [110]. An accurate view on the extent and complexity of the human interactome will be vital for further progress in drug discovery, both as a supply of novel PPI targets and as a source of information to increase our insight in disease-related processes, as illustrated in the next paragraph.

5. The human interactome to study diseases
The currently available human interactome is still far from complete (coverage has been estimated at around 8% [9]) and increasing the coverage and quality of (disease-specific) interaction maps will greatly improve our knowledge on the molecular mechanisms of disease and, as a consequence, the identification of new drug targets. Even without a full coverage, the human interactome has already been successfully applied to unveil the molecular mechanisms of a wide range of pathologies. A multitude of studies have analyzed PPI networks to predict disease genes by considering information on disease loci [111], gene-disease phenotype associations [112-115] or disease-specific changes in gene expression [116]. Table 2 lists a number of disease genes and pathways predicted from PPI network analysis in a wide range of pathologies.

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>DISEASE GENE or PATHWAY</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe Combined Immuno-Deficiency Syndrome</td>
<td>JAK3</td>
<td>Oti et al., 2006 [111]</td>
</tr>
<tr>
<td>Asthma</td>
<td>GNB2L1, BRCA1, CBL, VAV1</td>
<td>Hwang et al., 2008 [117]</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>Insulin signaling</td>
<td>Liu et al., 2007 [118]</td>
</tr>
<tr>
<td></td>
<td>Nuclear receptors</td>
<td>Liu et al., 2007 [118]</td>
</tr>
<tr>
<td></td>
<td>Oxidative stress</td>
<td>Jesmin et al., 2010 [119]</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>IL-12/IL-23</td>
<td>Wang et al., 2009 [120]</td>
</tr>
<tr>
<td>Obesity</td>
<td>Oxidative stress</td>
<td>Jesmin et al., 2010 [119]</td>
</tr>
<tr>
<td></td>
<td>VIP</td>
<td>Liu et al., 2010 [118]</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>ADRB1, ADRA2A and GRB2</td>
<td>Jensen et al., 2011 [121]</td>
</tr>
</tbody>
</table>

Table 2 Examples of disease-related genes and pathways derived from PPI networks.

Integration with other types of genome-wide data, such as DNA sequence, expression data and transcription factor binding, with localization studies and by examining the effect of naturally occurring and disease-related mutations on the network will further increase the biological accuracy and value of PPI networks and holds further promise for obtaining novel insights into the causes of and potential cures for disease. Integrating PPI networks with expression data, for example, led to the identification of cancer susceptibility genes and oncogenes in breast carcinomas [122], B-cell acute myeloid lymphomas [123, 124]; the identification of markers for metastasis in breast [125, 126], colorectal [127, 128] and gastric [129] cancer; the prediction of disease outcome [130] and response to chemotherapy [131, 132] and the determination of therapy-resistance genes [133, 134]. By combining protein-protein and protein-DNA interaction studies Kim et al identified a Myc-centered regulatory network in embryonic stem cells, and showed that the Myc module is active in various cancers and predicts cancer outcome [135]. Another illustration of the strong potential of data integration is provided by Pujana et al. and aimed at identifying breast cancer susceptibility genes by combining gene-expression and protein-interaction data with large-scale phenotypic and genetic-interaction datasets in various species. Within this network, putative disease players could be pinpointed and subsequent genetic analysis in patients revealed a genetic link between breast cancer susceptibility and centrosome dysfunction [122]. Finally, Mani et al. used a systems biology approach to assemble disease-related signatures. By combining a predisposed B-cell interactome and a large microarray data
set, they identified dysregulated interactions instead of dysregulated expression levels in three different types of lymphomas, which resulted in the successful identification of both known and candidate oncogenes as well as secondary effectors [124].

Because the rate-limiting step of such studies is the small coverage of the available PPI interaction maps, additional experimental efforts will be required to chart the interaction network of specific human diseases. This has already been done for spinocerebellar ataxia [136], Huntington disease (HD) [137-139] and schizophrenia [140].

Lim et al. used Y2H assays to generate an interaction network around 23 known ataxia-causing genes, a subset of which was confirmed by co-AP from mammalian cells. As such, they were able to identify Puratrophin-1 as an indirect binding partner of the ataxia-causing protein, Ataxin-1. Mice that lack this protein were later shown to indeed cause ataxia-like phenotypes. Further analysis of the network revealed that many human orthologs of genetic modifiers of the disease previously identified in Drosophila and mouse genetic screens directly interact with the causative proteins of ataxia. This suggests that the network might contain other genetic modifiers of ataxia, potentially including novel therapeutic targets [136].

Several groups contributed to the construction of a HD interactome by means of Y2H and AP-MS. The biological relevance of this subnetwork became evident upon the identification of GIT1 that directly interacts with and promotes aggregation of huntingtin, as this aggregation is related to the pathogenesis of the disease [137]. A lot of proteins from this network were also shown to act as genetic modifiers of neurodegeneration in a HD fly model [138, 139]. More specifically, proteins with functions in synaptic vesicle fusion and neurotransmission were later proven to be pathogenic in follow-up studies in HD transgenic Drosophila [141].

Camargo et al. used Y2H screens to build an interaction map around DISC1 and dysbindin, two well-established risk factors for schizophrenia. This map contained many proteins that are located in schizophrenia risk loci which underscored its biological significance [140].

6. The human interactome and infectious diseases

As alluded to already earlier, an important source of PPI targets are those between human and pathogen proteins. Here, not the human interactome as such, but its extension towards the proteins of viruses and bacteria is used as a starting point for identifying novel therapeutic targets.

Viruses are highly evolved organisms that hijack the cellular machinery of the host cells they infect to overcome the limitations of their small genomes. By defining the interactions between viral and host proteins we will obtain a better knowledge on how viruses reproduce and cause disease, which may ultimately reveal novel targets for therapeutic intervention. Systematic maps capturing viral-host physical PPIs have been obtained using Y2H for Epstein-Barr virus [142], hepatitis C virus [143], several herpesviruses [144, 145], vaccinia virus [146], influenza virus [147], dengue virus [148] and T-cell leukemia virus [149], and by AP-MS methodologies for HIV1 [150] and measles virus [151]. These studies indicated that despite the small size of viral genomes, the interaction map with host proteins is remarkably complex. Dyer et al. integrated publicly available human-pathogen PPIs from...
190 strains of various, mainly viral, pathogens to identify commonalities within the disease-causing strategies of different pathogens [152]. Similarly, Navratil et al. build an integrated human-virus interactome for more than 100 viral species and used it to decipher the molecular basis and complexity of virus-induced diseases [153].

To identify physical virus-host interactions that have deleterious consequences, genetic information derived from global RNA interference (RNAi) screens can be integrated into the PPI network. Using HT-Y2H, Shapira et al. assembled a physical map of associations between human and influenza viral proteins. A whole battery of genes that might be crucial for influenza replication or host defense mechanisms were identified by combining known cellular human networks with micro-array data of primary lung cells exposed to wild type influenza virus, virus lacking NS1, viral RNA or type I interferon. Functional validation of these candidates through RNAi-mediated depletion in primary lung cells revealed a potential role for RNA binding proteins, components of the WNT signaling pathway and viral polymerase subunits in fighting viral infections [147]. Khadka et al. mapped the interactome between dengue virus (DENV) and the human host, again using a HT-Y2H assay. DENV mostly causes relatively mild illness with high fever, rash, headache, nausea and joint pain, but sometimes it can progress to severe disease characterized by hemorrhage and shock. Some of the host factors that interact with DENV proteins were shown by RNAi studies to have an essential role during infection. These host factors are involved in the complement and coagulation cascade, are part of the centrosome or of the cytoskeleton. For calreticulin, this was further substantiated by its colocalization with viral RNA and two viral proteins in DENV-infected cells [148]. Komavora et al. analyzed virus-host protein complexes by AP-MS directly from cells infected with a recombinant measles virus (MV) expressing amino-terminal tagged version of the MV-V protein. Prominent targets of MV-V are cellular proteins from mitochondria, ribosomes, ER, PP2A and HDAC complex, in addition to key components of the antiviral response. Importantly, a negative role of deacetylation on viral replication could be demonstrated by chemical HDAC inhibition and by RNAi-mediated silencing of MV-V-binding HDAC components [151].

Bacteria do not integrate into the host cell as viruses do. However, they have evolved a broad arsenal of virulence factors that hijack the host cellular machinery to bypass the physical barricades and to evade or undermine the subsequent immune response [154]. Therefore, interactions between host proteins and bacterial virulence factors represent important targets for the development of antibacterial drugs. This might especially be the case for Gram-negative bacteria where restricted drug uptake by the bacterial species is a major restraint of new potential drugs. Unfortunately, extensive PPI maps for bacterial pathogens with their human hosts are sparse. Using a HT-Y2H approach, Dyer and co-workers identified 3.073 human-B. anthracis, 1.383 human-F. tularensis and 4.059 human-Y. pestis PPIs, with on average around 50 conserved protein interaction modules between different pairs of pathogens. These conserved modules may reveal new targets for broad-spectrum immunotherapeutic approaches because they reveal which crucial host immune modulating pathways are hijacked by different pathogens. The major histocompatibility complex and the NF-κB pathway were shown to be putative targets for treating a wide range of infectious diseases [155]. In another study, Yang et al.
constructed a human-Y. pestis PPI network by combining their Y2H-generated PPI dataset with literature-curated interactions. As baits they choose a total of 153 potential virulence-associated proteins from the approximately 4,000 Y. pestis proteins. As such, the authors could show that during plague infection, Y. pestis virulence factors interact with host proteins involved in focal adhesion and cytoskeleton regulation pathway, which could account for resistance of Y. pestis to phagocytosis. VASP was identified as a novel interaction partner of YpkA, binding of which results in the inhibition of VASP-mediated in vitro actin assembly. Moreover, it was suggested that by interacting with proteins of the TLR and MAPK signaling pathways, Y. pestis might evade host innate immune response. Importantly, binding of YscL or SycE with the p65 subunit of NF-κB was shown to inhibit TNF-induced NF-κB activation [156]. Another HT-Y2H screen between 221 putative human phagosomal proteins and 3,091 predicted B. melitensis proteins was used by de Barsy et al. to identify Brucella spp. effectors that control the intracellular trafficking of their vacuole. A Brucella spp. protein named RicA was picked up as a specific interactor of the human small GTPase Rab2, which was previously shown to be essential for intracellular replication of B. abortus [157, 158].

Examination of these host-pathogen PPI networks has repeatedly shown that viral proteins as well as bacterial virulence factors are highly prone to interact with cellular proteins that occupy central position in the human protein interaction network [142, 143, 147, 148, 151-153, 155, 156]. Moreover, viral and bacterial pathogens seem to attack common cellular functions to facilitate infections [156].

7. Concluding remarks
The challenges associated with developing small molecule modulators of PPIs, once considered insurmountable, now seem to be manageable as evidenced by the growing number of candidate drugs targeting PPIs in the pipeline. Key factors that appeared to be essential to restore confidence in PPIs as a valid target class include detailed insights in the nature of PPIs and the application of this knowledge in the successful design of a number of PPI inhibitors with proven functionality that made it to the clinic. It will be extremely exciting to follow progression of these molecules through clinical evaluation, and eventually their market approval, which would represent a key milestone for drug discovery. In the meantime, the field of PPI drug discovery urgently needs to be further elaborated, a.o. through the development of new screening tools, by exploring novel areas of chemical space and by identifying additional disease-related PPI targets. With regard to the latter, over the past decade interactomics research has become a dynamic and rapidly growing discipline, with ever expanding PPI networks being mapped at higher accuracy with new tools and giving rise to important novel insights into the functional organization of protein networks. Merging these large-scale interactomes with other genome-wide datasets should allow the development of a systems view on cellular function, which will aid in selecting the proper targets for therapeutic intervention.
References


