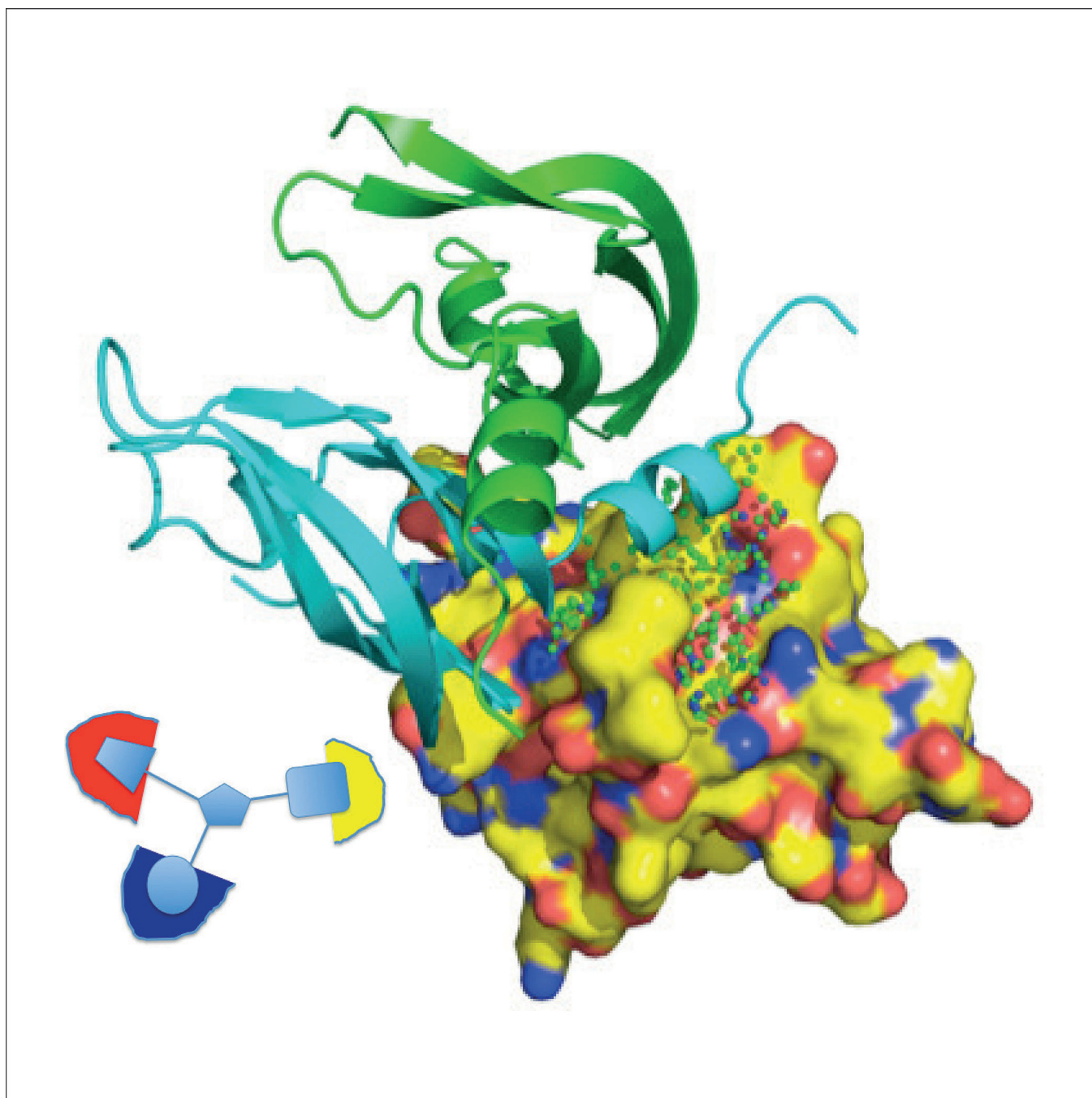# Drug-Like Protein−Protein Interaction Modulators: Challenges and Opportunities for Drug Discovery and Chemical Biology

Bruno O. Villoutreix,*[a, b, e] Melaine A. Kuenemann,[a, b] Jean-Luc Poyet,[a, b, c, e] Heriberto Bruzzoni-Giovanelli,[a, b, d] Céline Labbé,[a, b] David Lagorce,[a, b] Olivier Sperandio,[a, b, e] and Maria A. Miteva[a, b]

Abstract: Fundamental processes in living cells are largely controlled by macromolecular interactions and among them, protein–protein interactions (PPIs) have a critical role while their dysregulations can contribute to the pathogenesis of numerous diseases. Although PPIs were considered as attractive pharmaceutical targets already some years ago, they have been thus far largely unexploited for therapeutic interventions with low molecular weight compounds. Several limiting factors, from technological hurdles to conceptual barriers, are known, which, taken together, explain why research in this area has been relatively slow. However, this last decade, the scientific community has challenged the dogma and became more enthusiastic about the modulation of PPIs with small drug-like molecules. In fact, several success stories were reported both, at the preclinical and clinical stages. In this review article, written for the 2014 International Summer School in Chemoinformatics (Strasbourg, France), we discuss in silico tools (essentially post 2012) and databases that can assist the design of low molecular weight PPI modulators (these tools can be found at www.vls3d.com). We first introduce the field of protein–protein interaction research, discuss key challenges and comment recently reported in silico packages, protocols and databases dedicated to PPIs. Then, we illustrate how in silico methods can be used and combined with experimental work to identify PPI modulators.

Keywords: Protein–protein interaction modulators · Drug discovery · Drug-like molecules · In silico methods · PPI network

1 Introduction: PPIs, Past and Present

Proteins are polymers composed of amino acids that generally fold into a highly specific tertiary structure. Proteins can interact with basically all types of molecules, from small organic compounds, inorganic salts and metals, sugars, fatty acid, nucleotides, peptides to phospholipids of cell membranes and with other proteins. Short overviews of some key events in the field of protein science and protein–protein interaction (PPI) have been reported recently.[1–3] We here briefly report some key dates that have contributed to the field of PPIs (Figure 1).

The term "protein" seems to have been first mentioned in a scientific correspondence on the 10th of July 1838 between two scientists, Berzelius and Mulder (identification of a new unknown large molecule and two theories are associated with the term "protein": this substance could be conceived as a primordial substance and linked with the name Proteus, a Greek mythological character, possibly the first son of Poseidon, or the term "protein" could be linked to Proteus because this ancient god could assume a variety of shapes although this fact about protein was not known at that time). Around 1920, the interaction between the enzyme trypsin and one protein inhibitor was anticipated and so was the concept of antibody (*Antikörper*) binding some other substances (Ehrlich). Important advances also came around 1920 with the invention of ultracentrifugation (Svedberg, 1927), the realization that proteins could be purified and around 1950, it was possible to determine the amino acid sequence of insulin (Sanger) and alpha helix and beta sheet were pointed out by Pauling and Corey. Additional breakthroughs came from the determination of the 3D structure of proteins by X-ray crystallography: myoglobin (Kendrew) and hemoglobin (Perutz, Fersht, Simon, Roberts), in 1959 and 1960, respectively. Interestingly, the X-ray structure of hemoglobin is composed of four subunits noncovalently bound (i.e., tetramer, obligate complex, see below) and such work laid the groundwork for understanding quaternary structures (nomenclature of Linderstrom-Lang and Schellman, 1959) at the structural level and helped in gaining new insights about allostery (developed by Monod and collaborators around 1963). At about the same time, the concept of DNA and of mRNA for the synthesis of proteins was demonstrated by Monod, Jacob and Lwoff. Another major biophysical approach to investigate the 3D structure of proteins and in some cases of protein–protein interactions is NMR, first applied to proteins around 1982[1] while around 1978, Wodak and Janin implemented the first modeling algorithm for protein–protein docking. As our knowledge increased, it was more and more obvious that proteins were not acting alone. Clarification about networks of interactions yet required the development of large-scale tools together with global collective decisions

[a] B. O. Villoutreix, M. A. Kuenemann, J.-L. Poyet, H. Bruzzoni-Giovanelli, C. Labbé, D. Lagorce, O. Sperandio, M. A. Miteva
Université Paris Diderot, Sorbonne Paris Cité
UMRS 973 Inserm, Paris 75013, France
*e-mail: Bruno.Villoutreix@inserm.fr

[b] B. O. Villoutreix, M. A. Kuenemann, J.-L. Poyet, H. Bruzzoni-Giovanelli, C. Labbé, D. Lagorce, O. Sperandio, M. A. Miteva
Inserm
U973, Paris 75013, France

[c] J.-L. Poyet
IUH, Hôpital Saint-Louis
Paris, France

[d] H. Bruzzoni-Giovanelli
CIC, Clinical investigation center, Hôpital Saint-Louis
Paris, France

[e] B. O. Villoutreix, J.-L. Poyet, O. Sperandio
CDithem, Faculté de Pharmacie
1 rue du Prof Laguesse, 59000 Lille, France

Dr. *Bruno Villoutreix* is Research Director at INSERM (the French National Institute of Health and Medical Research), the French National Institute of Health and Medical Research, head of the research unit 973 also named MTi (http://www.mti.univ-paris-diderot.fr), co-director of a doctoral school focusing on Therapeutic Innovation (MTCI) and a co-founder of CDithem. He works on in silico approaches that assist the design and optimization of small molecule modulators of challenging targets with an emphasis on protein–protein interactions. Applications are carried out essentially in the area of cardiovascular diseases and cancer. He has been studying and working in France, Sweden, Finland and the United States, in the private and academic sectors (http://www.vls3d.com/). He has over 150 publications and a strong contribution in a compound that is now in clinical trial.

Mélaine Kuenemann is actually PhD student at the Inserm 973 research unit with Dr. B. Villoutreix and Dr. O. Sperandio as supervisors. She is specialized in chemistry, chemoinformatics and bioinformatics. She is presently working on the characterization of the chemical space of protein–protein interactions inhibitors using chemoinformatics approaches.

Dr. *Jean-Luc Poyet* is Research Director at INSERM (the French National Institute of Health and Medical Research) working at the Hospital St-Louis (Paris Diderot), in Paris, where he heads the research team Protein–Protein Interaction in the control of Apoptosis. He is a co-founder of CDithem. He is specialized in the identification of therapeutic targets through apoptotic and inflammatory pathways study and in the development of novel, validated pre-clinical leads in the field on cancer.

Dr. *Heriberto Bruzzoni-Giovanelli* is Associate Professor of Pharmacology at Denis Diderot University and Clinical Practitioner at the Clinical Researcher Center of the Saint Louis Hospital in Paris. Since 1995, his fundamental research has focused on apoptosis and protein–protein interaction and since 2006 he also works on pharmacogenomics. His clinical research is dedicated to early-phases clinical trials in oncology.

Céline Labbé is an engineer at the Inserm 973 research unit. She has a master degree on bioinformatics. She works on new in silico methods that assist the rational design of low molecular weight inhibitors of protein–protein interactions. Currently, she is developing a unique manually curated database of PPI modulators.

Dr. *David Lagorce* is Research Engineer at INSERM since 1995 and working at MTi unit 973. His research area is cheminformatics, small compound library design and in silico ADMET predictions.

Dr. *Olivier Sperandio* is research scientist at INSERM-University Paris Diderot U973. He has a background in chemistry and chemoinformatics. He is the head of the iPPI-DB initiative (http://www.ippidb.cdithem.fr) and coordinator of the Inserm CDithem platform. His project aims to the rationalization of the PPI chemical space in order to provide the chemist with specific guidelines in their attempt to design drug-like modulators of protein–protein interactions with therapeutic purposes.

Dr. *Maria Miteva* is Research Director at Inserm, France, currently working at the INSERM-University Paris Diderot unit 973. She leads the team "Virtual screening, PPI & ADMET in silico" of this research unit. She has over 20 years of expertise in the development of structural bioinformatics and drug design approaches. Her current work aims to develop strategies to identify small-molecule protein–protein interaction modulators and to predict their ADMET properties based on structural and mechanistic bases. She has published about 75 papers in reputable journals and edited an e-book entitled "In silico lead discovery".
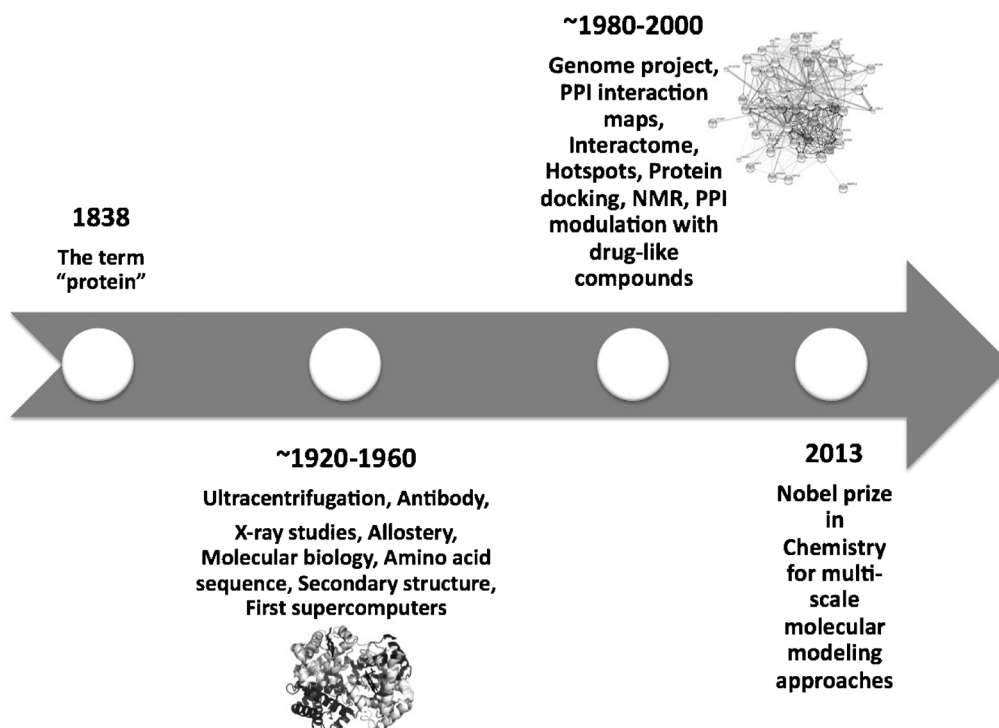
Figure 1. Timeline of Protein Science and PPI research.

of launching large-scale scientific projects such as the Human Genome project (1990) and various structural genomics initiatives (launched around 1998–2000). The yeast-2-hybrid (Y2H) method reported in 1989[4] is an example of a large-scale approach that greatly facilitated the identification of binary interactions. The first systematic PPI interaction maps were published in 2000, using Y2H while maps resulting from the use of another method, namely affinity purification-mass spectroscopy (AP-MS), started to be reported around 2002. Indeed, the term interactome (coined by a group of French scientists headed by B. Jacq) appeared in the literature in 1999.[5] During the 1990s, the 2000s, and up to now, an impressive amount of experimental efforts has been dedicated to PPIs, using known approaches tuned to PPIs or developed for the direct (measure the actual concentrations of the bound and free protein forms, eg., gel filtration, ultracentrifugation, etc) or indirect (imply the concentrations from an observed signal, e.g., many optical methods like fluorescence-based methods) analysis of interaction including PPIs. In addition, several methods were applied to investigate affinity such as isothermal titration calorimetry, surface Plasmon resonance, and fluorescence-based methods. At the same time, since 1990 and up to now, realizing that such complex system could not be assessed with experimental approaches alone, many in silico methods were developed. These approaches allow prediction of protein–protein complex by text mining, visualization of dynamic PPI networks, assessment of the PPI interfaces up to the screening of thousands of small molecules and the design of novel compound collections dedicated to PPIs (Figure 2).

Around the year 2000, as a tremendous amount of work on PPIs has already been carried out, as it was noticed that PPIs were playing a major role in many disease conditions[6] (e.g., in cancer[7]) and because new drug targets were needed, new projects aiming at identifying low molecular weight drug-like compound modulators of PPIs (in addition to the traditional ways of acting on PPIs such as with monoclonal antibodies and other types of proteins and peptides) got started in several academic and private laboratories. However, it is important to note that during many years up to around 2000–2005, it was essentially considered by the scientific community that PPIs could not be modulated (inhibitors or stabilizers) by drug-like compounds. Since then, the situation as changed and remarkable efforts are now being made to rationally design PPI modulators (see for instance the literature[8–29]). Many databases and in silico tools that assist drug discovery and chemical biology have been developed and most URLs for these services can be found at www.vls3d.com.[30] Of major importance for the research teams working on methodological developments and applications of in silico tools in the areas of Health and Biology, the 2013 Nobel prize in Chemistry was awarded to Karplus, Levitt and Warshel (see some recent reviews from these scientists[31–33]). It is indeed the first Nobel Prize given to work carried out in the field of computational biology and chemistry.
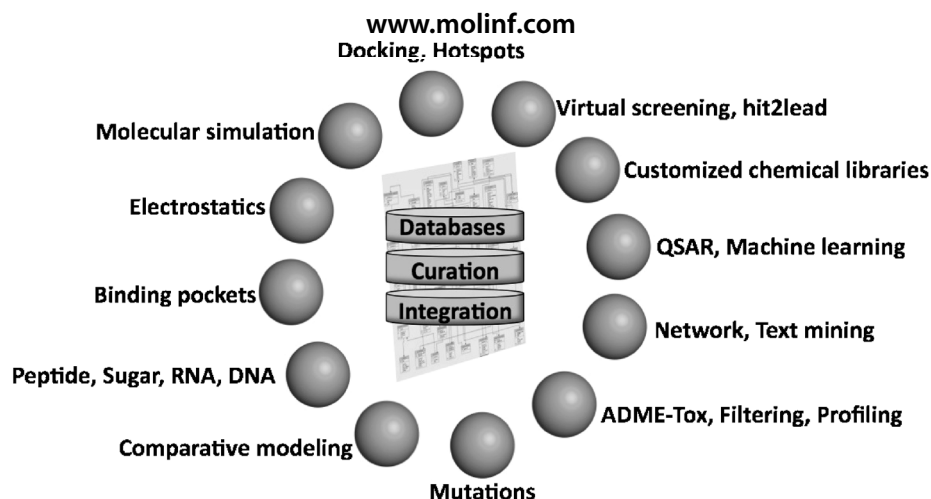
Figure 2. In silico tools and PPIs. Main in silico tools that assist the investigation of PPIs and the rational design of PPI modulators.

The present review will primarily focus on in silico approaches (focusing somewhat more on software packages and databases reported in 2013–2014) that can assist the rational design of "drug-like" orthosteric PPI inhibitors, while the readers can find recent reviews about other types of molecules able to modulate PPIs such as peptides, macrocycles and antibodies.[21,34–40] Fragment-based technologies are well-suited to target PPIs but will be only briefly commented upon here as recent reviews on the topic have been reported.[37,41,42] We will discuss some aspects of PPIs, from networks to structural analysis of the interface with notes on diseases and target selection. Key in silico methods that assist the rational design of PPI modulators are then introduced with a special emphasis on PPI inhibitors. We illustrate PPI hit discovery on two recently investigated biological systems, the VEGF-VEGFR complex[43] and the anticoagulant activated protein C.[44]

2 PPI Studies Combining In Vitro and In Silico Approaches: from Network to Structural Features and Mutations of the Interfaces, a Short Overview

It is important first to select the right protein–protein complex among several hundred thousands of known or anticipated interactions. In order to perform this step in a rational manner, knowledge about PPI networks can be critical. Yet, to gain additional knowledge about the selected complexes, structural analysis and predictions are usually needed. Several of these aspects can be investigated experimentally but in silico strategies can greatly assist the process.

2.1 PPI Network

The explosive growth of PPI data derived from small-scale to genome-scale studies implied the development of over 100 databases and in silico services dedicated to PPIs.

These many public PPI databases are important because they help the scientific community to gain new insights about PPIs (i.e., data have to be collected, integrated, curated and translated into knowledge). At present, some databases focus on some specific species and can be very specialized, others may contain data coming from large-scale studies. Overall and at present, many databases contain redundant information. Some databases contain data about "experimentally" identified protein–protein complexes (e.g., with Y2H, gene co-expression, split ubiquitin, protein complementation assays, AP-MS…each method has strengths and weakness and there are known artifacts[45]) while others are built using interactions collected from literature searches; some databases contain the 3D structures of protein complexes (please see for example[45–47]). One major difficulty with some large-scale data is that PPIs detected using the same methods or with different methods by different research groups but on the same organism can display very limited overlap (false positives: detectable interactions but functionally irrelevant; false negatives: miss interactions that do occur) calling for major curation efforts to filter out unreliable interactions (remove the noise) and quantification of errors. These differences also suggest that the techniques could be providing complementary descriptions.[48] There are in fact many reasons for such discrepancies including the obvious differences between the experimental methods used (i.e., some methods can capture transient interactions, others are geared towards identification of stable interactions, etc), but yet, at present, these differences represent a serious concern for the PPI field as downstream analyses of the resulting networks can preclude meaningful estimates of the size of the functional interactome or of the importance of some interactions in a given disease condition. Adoption of the Proteomics Standards Initiative Molecular Interaction (PSI-MI) format[49] and the related directives of the IMEx consortium should help improve the quality of the data,[50] but much work is needed in this area. It has been assumed until recently that literature curated PPI data were of higher accuracy than those

produced by high throughput studies because they were derived from focused investigations, but recent analyses are suggesting that this is not longer the case.[51]

Some well-known databases include, the Database of Interacting Proteins (DIP),[52] the Biomolecular Interaction Network (BIND),[53] the Molecular Interaction (MINT),[54] the Mammalian Protein—Protein Interaction (MIPS),[55] the host-pathogen interaction database (HPIDB),[56] IntAct,[57] Bio-GRID,[58] STRING.[59] Structural information can be found at the Protein Data Bank (PDB) and for instance PISite collects interface data from the PDB.[60] Structures of domain-domain interactions are available from 3did,[61] and iPfam,[62] while a spatial classification of 3D protein domain interaction database, KBDOCK, has recently been reported.[63] The database Instruct contains high-quality 3D structurally resolved protein interactome networks.[64] Homology models can also be used to study further PPIs and increase coverage. The Interactome3D database[46] provides 12 000 structurally resolved PPIs in 8 organisms while Instruct contains over 6500 human PPIs.[64] Another resource, the Protein Interaction and Molecule Search (PRIMOS) platform, represents a novel web portal that unifies six primary PPI databases (BIND; DIP; HPRD (Human Protein Reference Database); IntAct; MINT and MIPS, Munich Information Center for Protein Sequences) into a single consistent repository.[65] Along the same line, iRefWeb is a bioinformatics resource that offers access to a large collection of data on protein—protein interactions in over a thousand organisms. This collection is consolidated from 14 major public databases.[51] Similarly, curated PPI data can also be found via the PSIC-QUIC Web service, which provide access on the fly to files made available by over 28 source databases.[66] Also, the DASMIweb service currently has access to 36 distributed data sources. Ten of these provide interaction data that have been experimentally determined or curated from the scientific literature, 24 data sources contain computational predictions, and 2 data sources can be used for scoring the quality of the interactions.[67] Another example is UniHI 7 (Unified Human Interactome). The online tool integrates about 350 000 molecular interactions for more than 30 000 human proteins. Besides protein—protein interactions from 12 different resources (including HPRD, BioGrid, IntAct, DIP, BIND and Reactome databases) as well as four interaction maps produced by computational predictions and two high-throughput yeast-2-hybrid screens, UniHI 7 also comprises curated transcriptional regulatory interactions from three complementary databases TRANSFAC, miRTarBase and HTRIdb. In addition to these interactions, the service also integrated drug target information from DrugBank that can be mapped and visualized online without having to download, manually process and load the data into an external standalone application.[68] The data can be filtered by the users (e.g., number of PubMed references, small-scale or large-scale experiments, direct or indirect connection, binary or complex interaction).

PPI networks can be derived from data collected by the above-mentioned methods, namely, methods that probe binary interactions (e.g, Y2H), and approaches that detect multi-protein complexes (e.g., AP-MS). Both, AP-MS and binary detection methods probe non-native protein constructs, where tags are appended to the native polypeptides, potentially altering their properties.[45] As mentioned above, networks can also be built from curated PPI data collected from the literature. All these data can be visualized using, for example, the Cytoscape package.[69] Such PPI networks, comprising in human approximately 130 000 to 650 000 protein interactions (only a small subset has been fully experimentally identified)[70,71] can shed new light on human diseases.[72,73] Further, monitoring portions of the network that change when cellular states and conditions are altered could also give new insights about the health and disease states.

Analysis of PPI networks using computational and statistical tools help to understand how networks mediate genotype to phenotype relationships. If we take as example binary interactome maps, structurally, these maps were found to have a so-called scale-free topology with hierarchical modularity.[74] In networks of this topology, proteins are depicted as nodes and interactions as edges and in general, only some proteins, so called hubs, have a very large number of interaction partners (see the literature[75,76] for in depth discussion of PPI networks and network visualization) (Figure 3). This also means that such networks are resilient
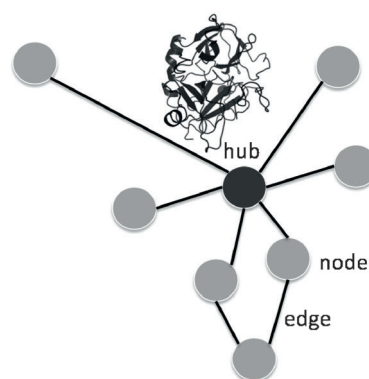

Figure 3. PPI network representation. A simple illustration of part of a PPI network with the serine protease thrombin at the center.

against failure of random nodes (e.g., by mutation) but sensitive to targeted attack of the hubs. Fascinatingly, in both plants and human, proteins of viral, bacterial, and fungal pathogens were all found to target such hub proteins.[2] Essential proteins tend to be more interconnected than non-essential proteins. It would seem that human disease-associated proteins too, are more interconnected than non-disease proteins.[74]

Finally, as whole genome and transcriptome sequencing gets cheaper and faster, gene expression profile analyses of

normal and pathological conditions should also contribute to the initial identification of clinically relevant PPIs that will then require further in depth investigation. For instance, the pro-survival IAP and BCL-2 proteins represent highly attractive PPI targets since their over-expression is associated with tumor progression and maintenance. Many comparative gene expression databases have therefore been developed, that allow the retrieval, analysis and comparison of gene expression patterns within or among species (see for instance the literature[77–79]). Although it is not exactly known how many PPIs would have a therapeutic potential, such picture strongly suggests that the design of small molecular weight compounds targeting PPIs could have a major impact in the near future. With regard to drug discovery endeavors, even if at present PPI networks do not fully reveal the topologies of networks truly operating in the cells because of the limitations mentioned above, such analysis can still help to propose a list of potentially interesting PPI targets that then need to be explored further and/or could pinpoint several proteins that would need to be targeted at the same time following, for example, the concept of rational polypharmacology design.[80–83]

2.2 Structural Analysis of Protein−Protein Interfaces: from Experimental Structures to Point Mutations Involved in the Disease State

By introducing atomic resolution knowledge (e.g., using X-ray, NMR, high resolution electron miscroscopy) in PPI networks and by small-scale in depth analysis of protein−protein interfaces, new insights can be gained with regard to the rational design of PPI modulators. General principles about PPIs at the atomic levels have been proposed for instance by Janin and Chothia in 1990[84] or by Jones and Thornton in 1996.[85] The range of K_d values observed in biologically relevant processes that rely on PPIs is wide and extends over about 12 orders of magnitude from 10^{-4} to 10^{-16} M (overall, the binding energy ΔG between protomers does not appear to be correlated with the size of the interface or other interface parameters such as the planarity and polarity for most PPIs[86]). A key fundamental distinction between PPIs is by their duration (that is whether the interaction is permanent or transient and this one can be further divided into weak and strong transient interactions). A slightly different definition expresses the duration as well as the functional aspect, dividing protein−protein interactions in terms of obligate (the protomers are not stable on their own in vivo) and non-obligate complexes[1,25,86–89] (Figure 4).

It is important to note that many PPIs do not fall into distinct types, rather, a continuum usually exists. Depending on the types of complexes (permanent, transient…), the nature of the interface usually differs. For instance, non obligate protein−protein complexes have been analyzed and the interface size measured by the buried surface area approach has a mean value of 1910 Å^2, with an average of 204 atoms contributing to this region belonging to 57 amino acids, that is about 28 residues per protein.[3] Analysis of a novel PPI dataset suggest that the minimum protein surface that must be buried to form a functional complex is in the order of 900 Å^2 (about 500 Å^2 provided by each partner) and involves about 12 residues on each partner (of course these values can differ slightly depending on the datasets and the way the computations are carried out). A large majority of atoms in non-obligate interfaces are usually still accessible to the solvent. Relative to the accessible protein surface, the interfaces of such complexes are depleted in Glu, Asp and Lys and enriched in Met, Tyr and Trp. The rim made of residues in which none of the interface atoms are fully buried has a composition close to the protein accessible surface. The core comprises buried atoms and about 55 % of all interface residues. This core region is enriched in aromatic residues and to a lesser extent, in aliphatic residues (but not Val, Ala and Pro). Arg residues can be present in both the core and the rim regions. Another region was also recently described, the so-called support zone that seems similar in composition to the protein interior.[90] By comparisons with other types of interfaces, like for instance homodimers, these complexes tend to have on average a buried surface area twice that of the non-obligate complexes.[3] The interface is here more hydrophobic and tends to be enriched in aliphatic and aromatic residues, on average, by a factor of 2 as compared to non-obligate interfaces. Analysis of interfaces can also be carried out in term of proteins involved in a given disease, and/or in term of hub versus non-hub proteins. For instance, it was shown that protein−protein complexes and hub proteins in cancer have smaller, more planar, less tightly packed binding sites compared to non-cancer proteins (and non-hub proteins), indicating low affinity and high specificity of the cancer-related interactions.[91,92]

Further, within interface regions, in general, not all residues are equally important and it is possible to use the concept of hotspots (the binding energy is not equally distributed among all amino acids participating in the interaction, some residues are directly responsible for the stabilization of the complex, these residues confer most of the binding energy to the interaction, typically they are defined as those residues contributing to at least 2 kcal mol^{-1} to the total binding energy of the complex).[93] These hotspots (hotspots tend to occur in clusters and can belong to the different protein partners, these ones are in contact with each other and form a network of interactions often called hot regions[25]) can be identified experimentally but a number of computational approaches can also be used.[94] It should be remembered that hotspot residues are not easy to identify experimentally (e.g., alanine-scanning experiments) or in silico (see for instance discussions about possible misconceptions of alanine-scanning results[89,95]). Hotspot residues (among the most conserved amino acids) are generally located around the center of the interface, and are protected from bulk solvent by energetically less
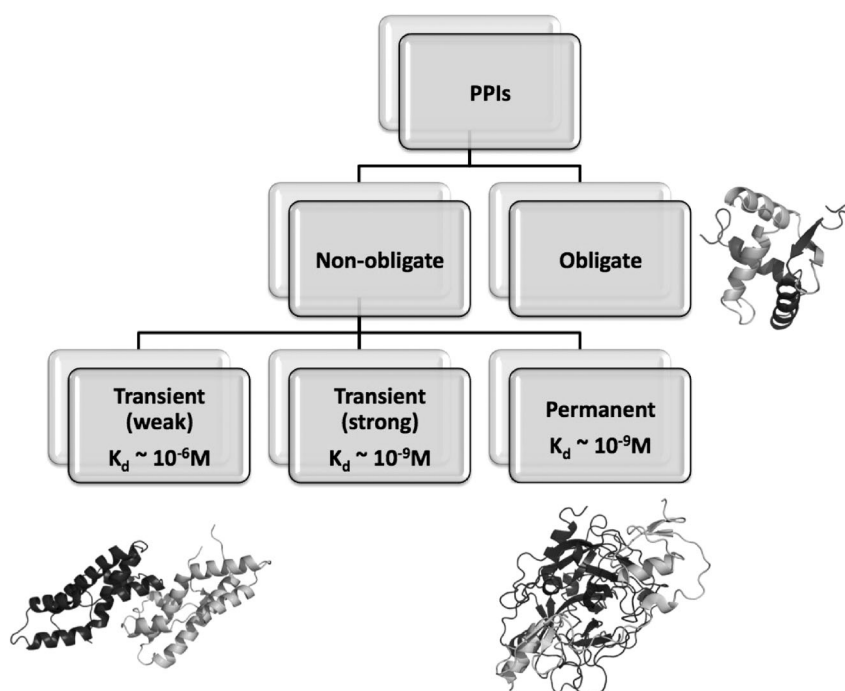
Figure 4. Transient and permanent interactions. A permanent interaction is usually very stable and thus generally exists only in its complexed form.[86,269] A transient interaction associates and dissociates in vivo. In an obligate PPI, the protomers are not found as stable structures on their own in vivo. Structurally or functionally obligate interactions are usually permanent, whereas non-obligate interactions may be transient or permanent (often antibody-antigen and enzyme-inhibitor systems). It is important to note that many PPIs do not fall into distinct types. Rather, a continuum exists between non-obligate and obligate interactions. The strong transient category of interactions illustrates the continuum that exists between the weak and the more permanent interactions. This category includes interactions that are triggered/stabilized by an effector molecule or conformational change. Some examples are given here to illustrate these concepts. Obligate: the Arc repressor dimer (PDB file: 1ARQ) (the Arc repressor of Salmonella bacteriophage P22 is a dimeric sequence-specific DNA-binding protein, one chain is shown in dark and the other is painted in grey); Non-obligate permanent heterodimer: insect derived double domain Kazal inhibitor Rhodniin in complex with thrombin (PDB file: 1TBQ) (thrombin (black) is a key serine protease of the coagulation system, the inhibitor is painted in grey); transient (weak): red abalone lysin dimer (PDB file: 2lyn) (Abalone sperm uses lysin to make a hole in the egg's protective vitelline envelope, one chain is in dark, the other is in grey).

important residues forming a hydrophobic O-ring. Tryptophan (21%), arginine (13.3%) and tyrosine (12.3%) are often hotspot residues whereas leucine, serine, threonine and valine tend to be disfavored.[96–98] The surface area of a region containing some hotspot residues is around 600 Å^2, a size that is compatible with a small molecule (NB: traditional protein-small ligand interaction ∼300–1000 Å^2 and the solvent accessible surface of many small molecule drugs usually ranges from 150–500 Å^2), and much smaller than a typical protein–protein interface (e.g., 1200 to 2000 to well over 3000 Å^2).[3] In addition, molecular dynamics studies have shown that hotspots are relatively rigid as compared to the surrounding interface residues.[99] Of importance also is the recent estimation of the number of possible protein interaction types, estimated to be around 4000.[100] By looking at the structure of the interface area and through investigations of the Protein Data Bank,[101] it seems that the interface space is limited and even chains with different folds often have similar interfaces. Possibly, the interface space is close to complete at present, suggesting that templates for interfaces are probably available

in the current version of the Protein Data Bank (about 100 000 protein structures in 2014).[102–104] Further, it is important to note that many protein complexes seem to be dominated by a hot segment where the interaction is dominated by a continuous epitope and as such hot segments could be good predictors of PPI druggability.[34] Still along this line of attempting to predict PPI druggability in term of a region capable of binding a small molecule, a recently reported study attempts to define classes of PPIs that could be more easily modulated by low molecular weight compounds and suggests that the "tight and narrow" and "weak and narrow" protein–protein complex categories are good candidates.[105] PPIs can also be classified from a secondary structure-centric approach but the links with druggability of the interface are not fully understood at present,[106] yet, the interactions involving one helix with a binding groove might be easier to modulate with a small compound than other types of interfaces.[107]

Additional information about the importance of a protein–protein complex and about interfaces could come from non-synonymous single nucleotide polymorphism

(nsSNP) studies (i.e., single base changes leading to a change to the amino acid sequence of the encoded protein) data because many of these variants are associated with disease. Clearly, the development of affordable techniques for sequencing genomes and the application of these approaches will generate vast amount of new data, including SNPs. Thus far, studies looking at the effects of nsSNPs were performed on individual proteins but now, the impact of nsSNPs on protein–protein interactions starts to be investigated.[108,109] It seems that when a disease-causing nsSNPs do not occur in a protein core region, they are preferentially located at a protein–protein interface rather than on non-interface regions.[110] These studies could help find rules that assist the selection of a target. Along this line, the manually curated SKEMPI database has been developed and contains the effects of mutation on binding energies for about 2792 mutations across 85 protein–protein complexes.[111] New insights should definitely come from the analysis of such repository.

2.3 Protein–Protein Interfaces and Zones that Could be Drugged

Many different in silico tools can be used to probe PPIs at a structural level. There are tools that predict hotspot residues and methods that suggest regions of protein most likely to be in an interface region. Other algorithms attempt to predict the structure of a protein–protein complex, either by docking (guided or not by in silico prediction methods of protein–protein interface regions) or by comparative modeling. As interfaces can be flexible, simulation tools such as molecular dynamics and normal mode analysis are of major importance. Further, as small ligands tend to bind in cavities, tools to predict binding pockets, to predict the druggability of a binding pocket and simulation tools (that can unravel transient binding pockets) are also needed. Further, it can be of interest to compare and cluster PPI interfaces to facilitate the design of a compound that could bind to several protein complexes or to gain knowledge about druggable PPIs. Several of these methods will be briefly presented below and it is important to note that some tools can be used for several purposes, for instance, predict interface residues and hotspots or define most likely binding pockets for a small compound and hotspots.

2.3.1 In Silico Predictions of Hotspots and Residues Present at the Protein–Protein Interfaces

Diverse protein–protein binding site prediction methods have been reported (see discussions about these tools in[97,112]) (Figure 5), mostly based on sequence conservation, residue propensities, surface topology (planarity and protrusion), electrostatics, hydrophobicity and solvent accessibility.[16,21,23,97,112–114] Some protein–protein binding site prediction approaches are based on the protein sequence
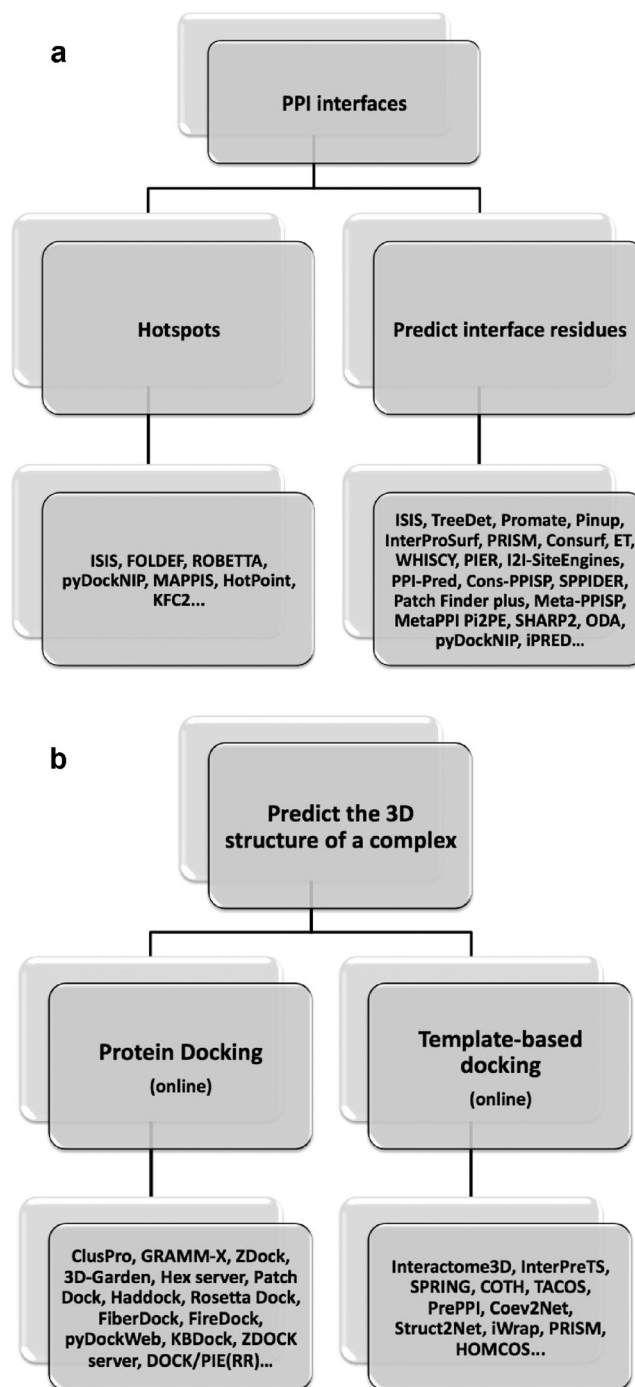

Figure 5. In silico tools to study protein interfaces. a) Examples of software packages that predict interface residues and hotspots. b) Some tools to predict the 3D structure of protein–protein complexes.

alone like the ISIS (interaction sites identified from sequence, neural network approach) approach[115] and PPI-cons (even if the tool used some structural information during training)[116] or SPPIDER (it runs with or without information about the 3D structure).[117] It has however been noted that methods that use structural information tend to

be more accurate than sequence-based approaches (see for instance the literature[46]). Other tools aiming at predicting interface regions require the structure of the protein(s). Some methods use empirical scoring functions like ODA (see for review the literature[97]), other approaches use sequence conservation among other parameters like Promate,[118] while others make use of machine learning techniques like SPPIDER.[117] Meta-servers, combining different tools have also been developed such as meta-PPISP.[119] Protein interface can also be probed by docking.[113]

Other tools developed to predict specifically hotspots (the experimental approach commonly used is alanine-scanning) require in general the 3D structure of the protein complex and can use, for instance, an empirical scoring function to assess the interface (e.g., HotPoint[120]) while others are energy-based like ROBETTA (or ROSETTA),[121] or FoldX,[122] or iPred.[123] Other methods for hotspot prediction use the unbound protein structures of each partner of a complex and docking e.g., the module pyDockNIP of the pyDock software package[97,124]). Hotspots can also be investigated by molecular dynamics in water and in isopropanol/water cosolvent environment (see for instance the literature[125]). Very few methods based on the sequence alone have been reported to predict hotspots, yet the method ISIS mentioned above has been also applied to predict hotspot residues.

2.3.2 Protein−Protein Docking and Template-Based (and Threading) Structural Predictions of Protein Complexes

A major difficulty in the field of PPI modulation by small molecules has been the lack of structural knowledge about the individual proteins forming the complex or about the macromolecular complex itself and the fact that some PPIs involve at least one partner (or one region) that is intrinsically disordered.[126,127] At present, about 100 000 experimental structures are reported at the PDB (the % of protein−protein complex is still very low) and it is possible to build homology models for numerous individual proteins while the second generation structural genomics initiatives together with advances in in silico protein−protein interaction predictions should improve the situation with regard to getting structural information about protein complexes in the coming years.[46,112,113,128] Protein−protein docking approaches and template-based structure modeling of PPI tools can indeed be used to propose reliable (at least some possible solutions that will need to be validated experimentally) models of the complex[129–131] (Figure 5). Yet, because of the complexity of the problem, these tools usually benefit from the knowledge of predicted interacting residues (e.g., such as to perform docking under restraints), site directed mutagenesis and other experimental information such as SAXS or electron microscopy. Many protein−protein docking engines have been reviewed like for instance in[23,128] while some new protein−protein docking tools released (or optimized) in 2013 include DockTrina (for docking triangular protein trimers),[132] ATTRACT,[133] MEGADOCK,[134] pyDockWEB,[135] F(2)Dock 2.0 and GB-rerank,[136] and SwarmDock (incorporating flexibility).[137] These approaches can also benefit from new scoring functions as illustrated by the combination of DockRank and the protein−protein docking tool ClusPro.[138] The other approach to build a protein complex is to use template-based modeling which constructs a complex by copying and refining the structural framework of related protein−protein complexes known experimentally. A list of in silico methods has been recently reported by the literature[130,131] and include for instance TACOS (Template-based Assembly of Complex Structures)[139] or the Struct2Net server.[140]

2.3.3 Binding Pocket Prediction for Small Molecules and Hotspots, Druggability and Clustering of PPIs

Most therapeutic targets (e.g., enzymes, GPCR, ion channels) usually display a clear concave binding pocket that can bind a small molecule. While having at hand the 3D structure of a protein−protein complex is very useful, it is still possible to design small PPI modulators even if one has only the 3D structure of one partner (experimental or homology model) of the complex. Several tools have been developed to predict binding pockets and to access the druggability (here defined as the likelihood of finding high-affinity low molecular weight binders (i.e., also called ligandability[87] first coined by Edfeldt et al.,[141] yet the term bindability can be used[142]) of these pockets. The tools were essentially developed for regular targets but such methods can still be applied (with cautions) on PPIs. In general, PPIs have not evolved to bind a low molecular weight chemical compound; interfaces tend to be flat, relatively large, often lacking a clear ligand-binding cavity but protein−protein interfaces that bind small molecules are often found to possess regions with 3 to 5 subpockets (see below)[143] and it has also been found that binding pockets may not be directly at the interface but within 6 Å of the interface[102] (of interest we also found that small ligand binding pockets can be found near the amino acids of a protein domain interacting transiently with a membrane surface,[144] could small ligand binding pockets be present next to most macromolecular interfaces ?). Also, protein−protein interfaces tend to dynamically adapt to upcoming ligands (small- or large macro-molecules), and transient cavities not visible in some experimental structures can appear on the molecular surface during (or prior to) the binding event.[126] In such cases, while the flexibility at the interface poses a significant challenge for structure-based drug design approaches, molecular simulation tools can assist and complement X-ray or NMR studies.[145–148]

Binding pocket detection algorithms are essentially subdivided into two major classes, geometry-based and energy-based tools.[149] In addition to predict binding pockets, some tools also provide a druggability score, that is, they give a score and rank the pockets for their likelihood

to bind a low molecular weight drug-like compounds (which can be different from reporting a list of cavities). In general, methods work on a static protein structure but some also take into account protein flexibility. Several recent articles or reviews describe these tools and the underlying concepts.[112,149–156] Of importance is the observation that protein-ligand binding hotspots in PPIs seems to correlate with protein–protein hotspots.[88]

Because, the identification of small molecular binding sites within or nearby protein–protein interfaces can be difficult with conventional methods, other tools geared toward PPIs were developed. For example, methods based on probing a protein surface (that can be tuned for PPIs) with organic fragments and predicting the locations of likely binding regions based on where fragments interact with high affinity were developed. Tools like FTMAP[157] (computational mapping with 16 different chemical probes) and the FTFLEX extension of FTMAP (which takes into account side chain flexibility on the fly)[158] can be used to pinpoint a region that can be explored by structure-based virtual screening approaches and for hotspot prediction while if multiple structures are available (or obtained via molecular dynamics), FTProd[159] could be applied. In fact, it has been shown that when a main hotspot region at a protein–protein interface has a concave topology, with one or two additional hot-spots close enough to be reached from the first main hotspot site by a drug-sized molecule, then the region is likely to be druggable.[95,157]

Related to hotspots and about tools that could help target PPIs with small organic molecules is the concept of 'anchor' sites, which contrary to hotspots have explicit concave/convex geometries appealing for pharmaceutical intervention (i.e., anchors can also be hotspots). The ANCHOR tool was developed along this concept to assist the design of small molecule modulators of PPIs.[160] Another related concept is the notion of druggable interface. The 2P2I scoring function has been specifically designed to investigate interfaces and suggests interfaces that could be drugged.[161]

Another tool dedicated to PPI identifies and ranks clusters of interface residues in a PPI that are most suitable as starting points for rational small-molecule design. These clusters are called Small-Molecule Inhibitor Starting Points (SMISPs) and the approach is complementary to methods that identify binding sites through an analysis of the receptor surface (either through shape descriptors or chemical probes). The PocketQuery web service has been developed around this concept to predict hotspots, anchor residues and hot regions.[162] In the same study, the authors expect after a PDB-wide analysis, that about 48% of the protein complexes could be modulated with a low molecular weight molecule. A related concept involves the investigation of overlaps between small-molecules and protein binding sites within families of protein structures (i.e., bi-functional sites, so far about 8000 proteins from the human proteome have been annotated with bi-functional resi-

dues[163]). Davis et al.[164] reported the HOMOLOBIND software,[163] a tool that identifies residues in protein sequences with significant similarity to structurally characterized binding sites.

These tools tend to work on a static structure (although one can generate alternative conformations prior to the computations) while some others combine identification of hotspots by MM-PBSA free energy decomposition on the basis of the structural ensemble generated by molecular dynamics (MD) and generation of transient pockets using molecular dynamics and FRODA simulations.[165]

Tools to compare traditional binding pockets have been developed[149] and some examples of recently reported approaches include PocketAnnotate,[166] APoc[167] and Site-Comp.[168] All these pockets have been stored in databases like for instance the pocketome.[169] Somewhat related, in a recent study, interfaces were defined and clustered leading to the identification of 22 604 unique interface structures in the PDB.[170]

3 PPIs are Challenging but Should be Tractable Molecular Targets: Supports from In Silico Methods

Despite their therapeutic relevance, most small molecule drugs do not in general hit PPIs but rather enzymes, ion channels, nuclear hormone receptors and G-protein coupled receptors. In fact, these last 50 years, PPIs have been essentially modulated by therapeutic antibodies, therapeutic proteins and peptides (or modified peptides or more recently stapled peptides).[7,13,20,35,106,171] However, while biologics can possess outstanding qualities and be valuable in some pathological conditions,[38], some of these molecules tend to be problematic for at least three reasons:[40,172] (a) most of them are difficult or impossible to administrate orally with our present knowledge and can be unstable, (b) adverse immune reactions can occur,[173] and (c) biologics are usually expensive to develop, and/or produce, and/or store with treatment for one patient easily reaching over $100 000 per year[174] (a cost that most healthcare systems are not able to afford, and the associated problem of aligning the cost of small chemical compounds to the cost of biologics; the cost of drugs is a very controversial issue and it should be mentioned here that now, small chemical compounds can also be extremely expensive such as the recently approved prodrug Sofosbuvir for the treatment of hepatitis C infection, thus the debate about cost is far from being closed). Although significant advances have been made and will take place in the coming years,[175] several obstacles will have to be overcome, from cost to delivery issues.[21] It is here interesting to note that small-molecules and biologics can be combined (e.g., a small molecule can be given with a monoclonal antibody (mAb), or the grafting of a small molecule to a protein including mAb can be valuable in some cases). Along the same line, small-mole-

cules could be given together with biologics, not to gain a synergistic effect but rather to allow proper functioning of the biologics like for instance to avoid aggregation of a monoclonal antibody by using a small molecule PPI inhibitor (e.g., proof of concept study with the mAb bevacizumab or Avastin).[176]

There are many reasons that have led to the widespread perception that modulation of PPIs could not be addressed by small drug-like molecules, as mentioned earlier in this review: PPI interfaces are often considered as large, flat, lacking a well-defined ligand-binding cavity characteristic of conventional targets[12,22,23,87,177] (these obviously do not apply to all interfaces); the lack of appropriate high-throughput screening (HTS) technologies and/or the very low hit rates observed upon running HTS experiments[26,28] (in our hand, several PPI targets were screened experimentally with a "traditional compound collection" and the hit rates were around 0.01% compared to 0.1 to about 5% for the screening of a traditional target, yet it can be higher, see some examples in the recent review of Nero et al.[107]). Another type of problem concerns the compound collections used to screen PPIs. Indeed, the paucity of small-molecules in traditional screening libraries dedicated to PPIs reinforced by the observations that small PPI hit molecules often have physicochemical characteristics slightly outside of what would be expected for a typical oral drug starting point suggest that novel screening collections have to be designed for PPIs.[15,21,22,128,178,179] Thus, many in silico methods have been developed to assist the design of novel compound collections.

If we take as example PPI inhibitors, there are 3 main ways a small compound can block an interaction, the direct orthosteric inhibition where the compound binds at sites that overlap with the area of the protein that interacts with the other protein of a complex, allosteric inhibition where the small compound binds away from the interface area (it can be close to the interface or at a very different site) and interfacial inhibition, where the ligand binds to a transient pocket appearing for example during a conformational change and locks the protein complex in a nonproductive conformation (the targeted complex is a transient kinetic intermediate that is characterized by unbalanced energetic and structural conditions that create the binding site for the drug).[180,181] All approaches have pros and cons and have to be considered, but definitively, some types of allosteric inhibitions and in general interfacial inhibitions are difficult to predict in silico.[18,24,180,182,183]

3.1 PPI Compound Collections and ADMET

The earliest efforts to develop small protein–protein modulators were based on the mimicry of secondary structure elements of the interacting partners and thus at compounds able to mimic beta-turns, alpha helices and beta strands.[13,20] Such approaches are still valuable today.[171] Indeed, the greatest successes for HTS have been with PPIs

in which a helix of one protein binds into a groove of the interacting partner (e.g., the Bcl family)[107] illustrating the potential of compounds mimicking such secondary structure elements. As low molecular weight molecules addressing PPI were identified and collected, it became possible to characterize the key properties and structural features of these compounds (a principal component analysis of chemical vendor collections versus PPI inhibitors and allosteric inhibitors is reported in Figure 6, it should be mentioned that at least 4 chemical vendors now provide collections dedicated to PPIs, Asinex, Chemdiv, Life Chemicals and Otava, but in our hands, a preliminary PCA analysis indicates relatively similar trends even with these specialized collections). At present, at least 3 databases are dedicated to modulators of PPIs, the 2P2Idb (manually curated),[161] TIMBAL,[184] and iPPI-DB (manually curated).[185] Exploring and navigating these collections should help gaining insights into privileged scaffolds or substructures particularly well suited to bind at the PPI interface as well as required physicochemical thresholds and could help to derive new rules to design ADMET-friendly collections dedicated to PPIs.[12,15,18,22,177–179,184,186] Two research groups provide in silico filters that help to design PPI-focused libraries enriched in PPI inhibitors starting from large traditional compound collections. A decision tree approach was used by Reynes et al.[187] (PPI-HitProfiler which is now available online via the ADME-Tox filtering tool FAF-Drugs2,[188]) while support vector machines were used by Hamon et al.[189] (2P2I$_{HUNTER}$). Two studies reporting the rational design of compound collections dedicated to PPI that contain alpha helical binding epitopes integrating the concept of increasing the three-dimensionality of the compounds have been reported recently.[27,171] Along the same line but combining 5 physicochemical properties, a filter to select potential allosteric inhibitors was published together with an online server.[190]

With regard to the main physicochemical properties of known PPI inhibitors and potential ADME-Tox problems, investigation of known PPI binders showed that the molecules tend to have a higher molecular weight (average MW of 421 Da for protein–protein inhibitors versus 341 Da for regular drugs), higher log P (a mean value of ~5.1 for protein–protein inhibitors was found while it is around 3.5 for enzyme inhibitors) and a more complex three-dimensional structure than typical drugs, underlining further the need of rationally designing the screening collection. Yet, this general view does not apply to all PPI modulators.[18,22,177,178,185,191] For example, many compounds that are known to inhibit PPIs (because of some known physicochemical properties e.g., high lipophilicity) tend to violate several rules of thumb commonly used to select compounds after screening, or to prepare compound collections or to predict bioavailability or toxicity.[128,192–195] Such rules can be, the Lipinski rule of five (initially related to oral administration)[196] or the 3/75 rule (related to in vivo toxicity) which states that compounds with high lipophilicity
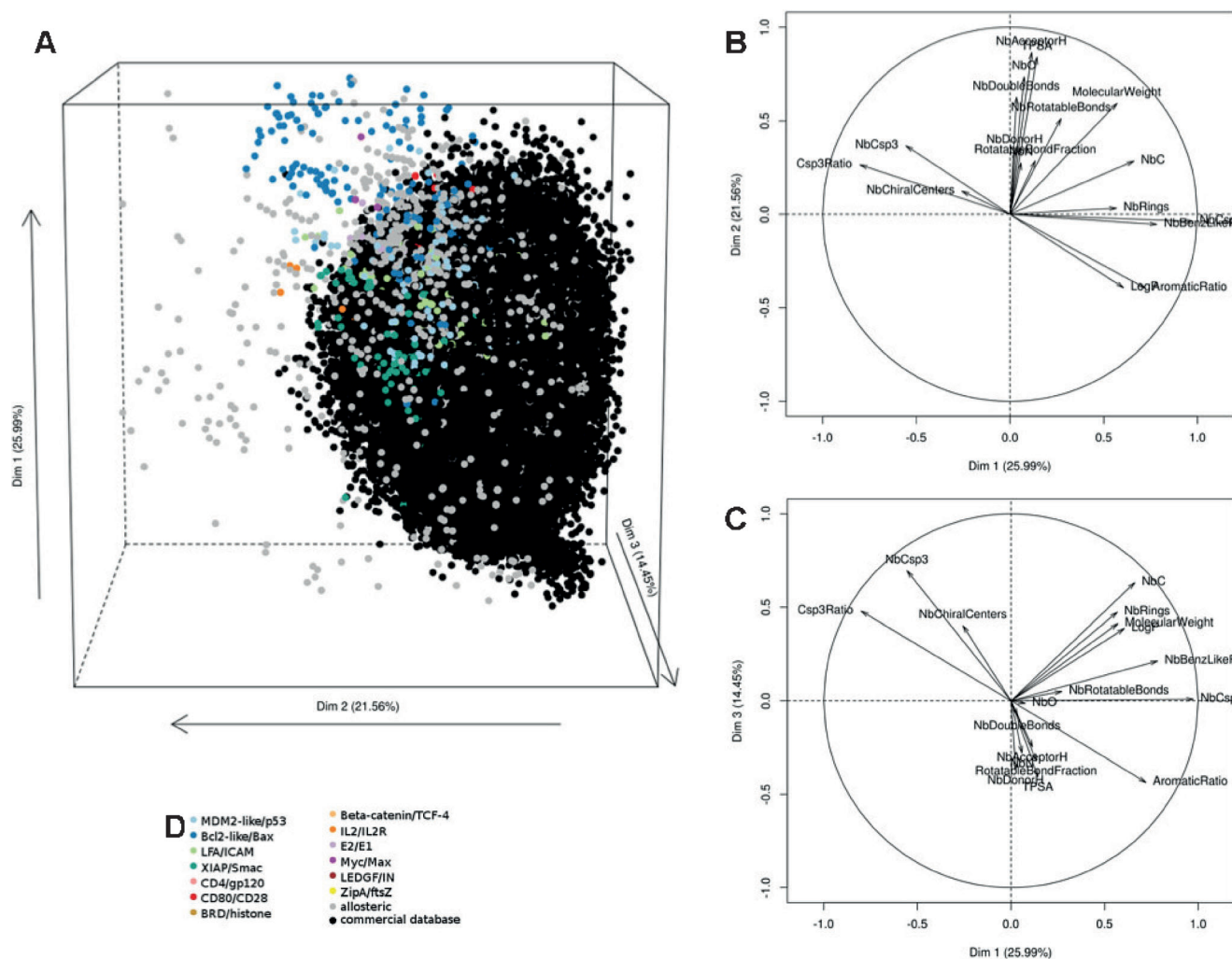
Figure 6. Principal Component Analysis (PCA) on a commercial and allosteric databases and a database of inhibitors of protein–protein interactions. A chemical vendor database, an allosteric database (version 2013)[270] and the protein–protein interaction inhibitor iPPI-DB (version 2013)[185] were used in this analysis. Redundant compounds, amino acids, salts, compounds with less than 10 or more than 140 heavy atoms, compounds with a molecular weight less than 150 g/mol, or more than 1000 g/mol, compounds with a logP less than −10 or more than 10, carbocation, compounds with phosphorus, and non-organic compounds were removed with PipelinePilot v.8.5. A diversity criterion was applied on our dataset using Bemis and Murcko assemblies on PipelinePilot v.8.5. The dataset contained 421 522 compounds for the commercial database, 5141 allosteric compounds and 685 protein–protein inhibitor compounds. The PCA calculations were run using 18 physico-chemical properties. (a) PCA of commercial database, allosteric database and iPPI-DB: individual map of compounds, PPI inhibitors are represented as color dots, commercial compounds are represented as black dots and allosteric compounds are represented in grey. (b) PCA of commercial database, allosteric database and iPPI-DB: variable map of axes 1 and 2. (c) PCA of commercial database, allosteric database and iPPI-DB: variable map of axes 1 and 3. (d) IPPI family colors used for the PCA. The three first axes of the PCA represent 62 % of the total variance. The first axis is represented by the compound's aromaticity (number of benzene-like rings and number of sp2 carbon) and the compound's complexity (number of sp3 carbon and Csp3 Ratio). The second axis is characterized by the number of hydrogen bond donors/acceptors. The third axis is characterized by the compound's polarity (here evaluated by TPSA, topological surface area). The global position of the protein–protein interaction inhibitors population is the upper part of the individual map showing that these compounds seems to be more hydrophobic and more aromatic. According to the global position of the commercial and the allosteric compounds, the two dataset seems to share the same chemical space, even if, a small number of allosteric compounds presents an unusual profile. This tendency was confirmed on 5 similar analyses with different commercial datasets from different chemical vendors.

(computed log $P > 3$) and low topological polar surface area ($TPSA < 75$) can have an increased risk of generalized toxicities (about 6 times more likely to be toxic in short-term animal studies)[197] In fact, because of some of these physicochemical properties, some PPI modulators may fit the so-called class II (low solubility, high permeability) or class IV (low solubility, low permeability) category of the Biopharmaceutics Classification System (see for review the literature[198]). Along the physicochemical properties line of reasoning and rules of thumb, a GSK team showed that increasing lipophilicity usually contributes to lower drug efficiency and consequently such molecules tend to require

higher doses which in turn can increase the risk of adverse drug reactions (e.g., increased promiscuity leading to increase binding to anti-targets).[199] Further, knowing that starting hits usually have to grow in size during the compound optimization phase with in general a further increase in log P (in fact compounds that have more chance to succeed tend to preserve a relatively low lipophilicity during the optimization program),[200] a clever design of PPI screening collection is required such as to obtain compounds with balanced ADME-Tox properties still compatible with the binding site features present at or near the protein–protein interfaces. Other difficulties could occur if a PPI compound has to hit a CNS target since, in general, the CNS physicochemical property ranges for molecular weight are around 300 and for log P, they are around 2.8.[201] However, although the ADME-Tox properties of PPI modulators are a legitimate concern, we have recently reviewed several PPI interfaces that can be modulated by hits that match the generally accepted rule of five-like guidelines.[18,185] For the remaining molecules, it is true that some PPI modulators have a higher log P (but many have an appropriate log P to start drug discovery) and molecular weight than many drug candidates that hit traditional targets, yet for the PPI modulators that have reached clinical trials, several seem to be orally available.[202] In addition, many studies highlight that some of these physical chemistry rules might be too restrictive (e.g.,[203,204]) as, for instance, larger compounds could reduce their effective size and lipophilicity through hydrophobic collapse or by forming internal hydrogen bonds, thereby enhancing membrane permeability and possibly impacting the overall bioavailability.[204–207] Furthermore, it has been stated that small PPI inhibitors tend to use more aromatic interactions than the corresponding protein partners that utilize also several charged residues, suggesting that new PPI modulators could possess more charged groups (this could improve some ADME-Tox properties while deteriorate others like permeability, but this information is important to explore).[208] Tremendous amount of work is needed in this area while inspiration from natural compounds could help understanding how to rationally go beyond several ADME-Tox rules of thumb.[209,210] Overall, it is likely that such obstacles will be overcome in the coming years.[18,183] In silico approaches that make use of multi-parameter optimization protocols should indeed facilitate the design of molecules with balanced ADME-Tox properties, adequate potency and relevant selectivity tuned to the disease type.[211] Definitively, gain in knowledge will come from the analysis of compounds that are in advanced preclinical stages and in clinical trials. At present only around 30–50 compounds[185] are at these stages but the increase research activity in this area should rapidly bring new insights that will favor rational and quality by design approaches.

3.2 Virtual Screening of PPIs

PPIs can be probed using HTS or virtual screening experiments followed by in vitro assays of a small selected list of molecules resulting from these computations. The term "virtual screening" (or in silico screening) was first reported in the scientific literature in 1997;[212] it can be defined as a set of computer methods that analyzes large databases or collections of compounds in order to identify and prioritize likely hit candidates.[128,213–223] In silico screening search can be performed on libraries that contain physically existing compounds, on PPI enriched focused collections or on virtual libraries, and thus on compounds that are not yet synthesized. Noteworthy is the fact that virtual screening can be used on very large databases that no experimental approaches can tackle. It is important to remember that the easily accessible drug-like space contains about 10^{33} molecules[224] and that with 17 atoms and simple chemistry rules, it is already possible to generate 166 billion compounds.[225] Yet, it should be noted that in silico screening goes much beyond number crunching, it helps to generate ideas, to reduce the cost and to gain knowledge. In silico screening experiments can be performed to complement HTS (and are indeed often integrated in screening campaigns), prior to experimental screening, or after HTS to rescue some compounds potentially missed by the in vitro readouts (see latent hits by[226]).[227] The complementarity between HTS and virtual screening has been shown in many studies, like for instance by screening both in silico and experimentally the same 198 000-compound collection against cruzain, a cysteine protease target for Chagas disease.[228] Along the same line, a computer screening experiment performed on a subset of the ChemBridge compound collection (about 500 000 molecules) and a study making use of HTS (50 000 molecules using also molecules from ChemBridge) found quasi-identical hit molecules for the proteasome cancer target.[229,230]

Virtual screening approaches have been traditionally subdivided into two main methods[231–236] (Figure 7): first, ligand-based screening, in which 2D or 3D chemical structures or molecular descriptors of known actives (and sometimes inactive molecules) are used to retrieve other compounds of interest from a database using some types of similarity measure or by seeking a common substructure or pharmacophore between the query molecule and the compounds in the database; and second, structure-based (or 3D receptor-based) screening in which compounds from the database are docked into a binding site (or over the entire surface) and are ranked using one or several scoring functions. Structure-based virtual screening also includes tools to perform binding site-derived pharmacophore search. There are some slight differences on how the methods are classified but the nomenclature used here is generally well accepted.[217] Of importance for PPIs is that structure-based screening can be carried out on homology models or on low-resolution structures.[237–239]
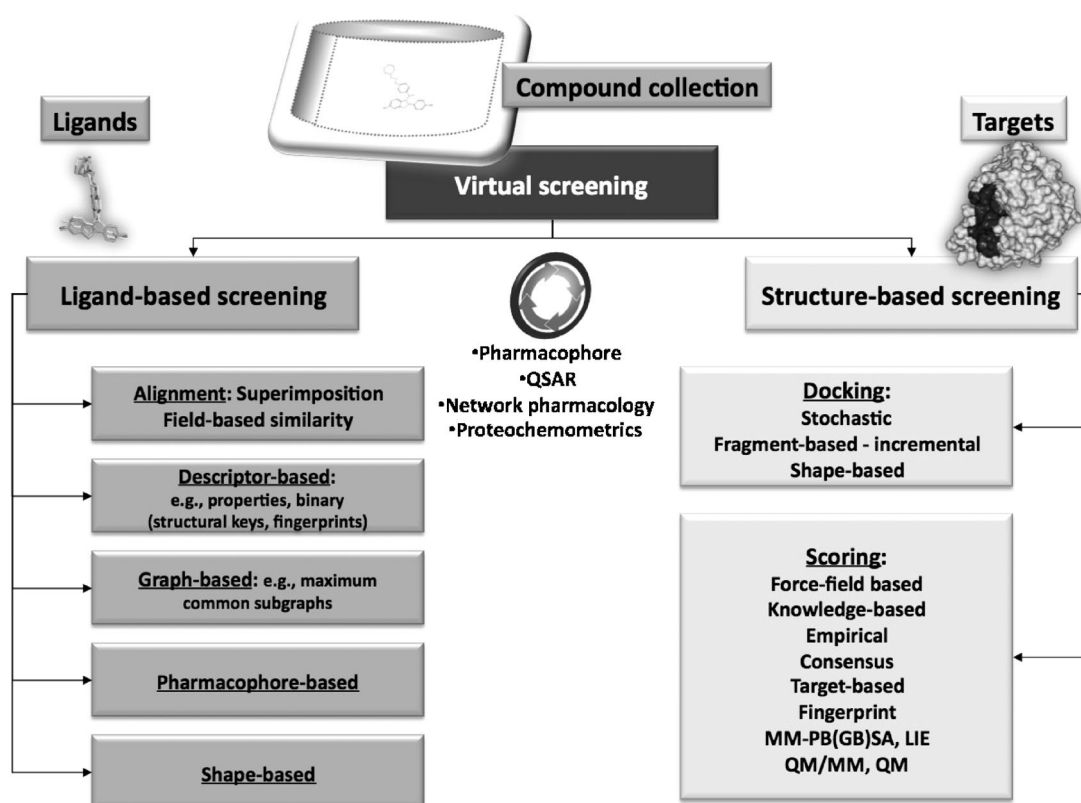
Figure 7. Main virtual screening methods. The two main virtual screening methods are ligand-based methods and structure-based methods. Some approaches can be considered to be at the interface between the two main screening concepts, such as some types of pharmacophore modeling that use information derived from co-crystallized target-ligand complexes or in the case of proteochemometric modeling, QSAR and systems pharmacology. Abbreviations: LIE: linear energy interaction; MM-PB(GB)SA, molecular mechanics-Poisson Boltzmann (Generalized Born) Solvent accessibility; QM/MM, quantum mechanics/molecular mechanics; QM, molecular mechanics. Additional information can be found in some recent reviews about virtual screening,[217] fragment-based approaches,[271] structure-based tools for screening and compound optimization,[272] or systems pharmacology,[128] or pharmacophore.[246]

The structure-based virtual screening process can then be continued if deemed appropriate using different types of post-processing approaches (see for instance the literature[214,215,236,240–242]). Ligand- and structure-based methods can be combined if the necessary information is available.[243] Virtual screening methods are relatively well-established, and numerous success stories in terms of hit identification, contribution to the development of drug candidates, or marketed products have been recently reviewed.[233,244] This does not mean that the methods have no flaws but yet they contribute significantly to the identification of interesting molecules.[231] Over 100 commercial and free tools are available to carry out virtual screening, many of these approaches have been discussed in several recent reviews.[30,128,232,242,245,246]

A compound collection is required to perform virtual screening and its preparation is as mentioned above critical in the case of PPI screening.[22,178] Physicochemical properties, structural alerts and flags for promiscuity should in general be considered. This is also important because molecules have to be optimized[234] and as it has been noticed that artifact compounds (e.g., PAINS, pan assay interference

compounds) are reported at a growing rate[128,247,248] (warning, some authors do not find that some PAINS molecules are that problematic[249]). In silico tools such as the FAF-Drugs online server[188] can assist in the preparation of a compound collection and, for instance, evaluate physicochemical properties, search for the presence of PAINS and toxicophores as well as assess the potential of a compound (the molecule has to be in 3D) to be a protein–protein interaction inhibitor according to the rules defined in the literature.[187] It is important to note that when searching for PPI modulators it might be necessary to apply soft in silico ADME-Tox filters to prepare the collection. For example, chemical groups that could react with a protein to form a covalent bond are usually not welcome in a drug discovery program, yet this could be useful when probing a PPI like in the case of inhibitors of the thyroid hormone receptor and co-regulator proteins.[250]

While all virtual screening approaches can be used for PPIs, some like pharmacophore derived from protein–protein interfaces[246] seems well-suited. Other tools like docking-scoring can be used although they have not been designed to target PPI pockets (the docking step can be af-

fected by the lack of well defined binding cavity and the scoring step is always sensitive). In a recent study performed over several ligand-docking engines, it was found that good docking solutions could be obtained using conventional docking-scoring tools (yet a drop of about 10% has been noticed as compared to regular pockets[251]), suggesting that structure-based screening can already assist the design of PPI modulators although additional methodological developments are required.[252]

4 PPI Low Molecular Weight Hit Discovery by Combining In Silico and In Vitro Approaches

Modulating PPIs with a small molecule could be beneficial in many cases, even more so if the small molecule can be administered orally.[12,19,253] First, this could open new avenue for therapeutic intervention (see some recent reviews discussing the combination of in silico, in vitro and in some case in vivo studies to assist the design of hits and or clinical candidates or even PPI drugs such as[18,22,23,29,171,178,181,183]). Further, as discussed in recent reviews, designing catalytic site inhibitors can be limited by high structural similarity among enzymes of the same family whereas the greater structural variability of protein–protein interfaces may provide a real opportunity for selectivity. Also, PPI modulators could be less prone to drug resistance than catalytic site inhibitors. In addition, even if two proteins bind with high affinity, they could be successfully out-competed by a weak small molecule binder. Indeed, in some cases, the mere alteration of a binding equilibrium will be sufficient to produce a significant biological effect without the need to completely inhibit the selected PPI. Several drugs (which target PPIs) are already in clinical use such as tirofiban (mainly orthosteric) targets integrins (glycoprotein (GP) IIb/IIIa receptors on platelets) to treat cardiovascular disease; maraviroc (mainly allosteric) interferes with the HIV gp120 interaction with the CCR5 receptor and blocks HIV viral entry (see for instance the literature[23]). At least 30 molecules are in preclinical or clinical stages[185] and target systems such as the IL2 and IL2R, Bcl2 and Bcl-XL, HDM2 and p53, the E2 viral transcription factor and the viral helicase E1, the ZipA membrane anchored protein and the FtsZ tubulin, and the TNF (tumor necrosis factor) trimers. These pioneered PPI systems have been discussed extensively in numerous recent reviews and we will thus focus here on two new complexes.

4.1 Inhibition of the VEGF-VEGFR Interaction

Modulation of PPI by small molecules has been applied in several therapeutic areas and given the pivotal roles of PPIs in many processes relevant to malignant development, the concept has been used very actively in the field of cancer.[107,253,254] Among the many PPIs important in cancer is the vascular endothelial growth factor (VEGF)-VEGF re-

ceptor (VEGFR) signaling pathway. Vascular endothelial growth factor (VEGF) plays a key role in angiogenesis, one of the hallmarks of cancer.[255] VEGF binds to several receptors including two major tyrosine kinase receptors (TKR), VEGFR-1 and VEGFR-2, on the surface of endothelial cells, thereby activating signal transduction and regulating both physiological and pathological angiogenesis. Whereas VEGFR-1 has been shown to stimulate endothelial cells migration.[256], VEGFR-2 is known to be a main initiator of signaling pathways in endothelial cells.[257] The VEGF-VEGFR system is a validated and promising target for anti-angiogenic treatments. Although the VEGF-VEGFR interface was found to be one of the flattest protein–protein interfaces available in the investigated dataset (the interface is relatively large > 800 Å² with a planarity value of 1.7 Å[43]), well below those of most transient protein–protein complexes (mean planarity value = 2.7 Å,[258] successful structure-based in silico screening was performed by targeting the VEGF-binding zone of the extracellular domain D2 of VEGFR-1. As flexibility is known to be important at protein–protein interfaces, the DFprot server was used to investigate the possible plasticity of the D2 domain of VEGFR-1.[259] Analysis of the X-ray structure (with the probe mapping algorithm ProtoMol implemented in the screening package Surflex and with LigBuilder) and of the simulation suggested that this region of the D2 domain was essentially rigid and as such docking experiments were performed on only one 3D structure of the D2 domain. Then, 8000 proprietary drug-like molecules (a subset of the French National Compound Collection) were docked with Surflex[260] onto the predicted binding pockets of the target (Figure 8).

After the in silico analysis, 206 compounds were selected for in vitro assays. Twenty compounds inhibiting the formation of the VEGF-VEGFR complex in the micromolar range were identified. The bioactive molecules contained a (3-carboxy-2-ureido)thiophen unit and the best IC_{50} was ~10 μM. Moreover, the most potent compound (compound ID 4321) decreased the auto-phosphorylation of VEGFR-1 induced by VEGF, inhibited HUVE cells capillary formation and disrupted the actin and tubulin networks. These findings suggest that the best hit could be a promising scaffold to probe this macromolecular complex and used as a starting point to develop new treatments of diseases linked to VEGFR-1.

4.2 Protein–Protein Interaction Inhibition Involving the Anticoagulant Protein C

The blood coagulation pathway comprises a series of efficient enzyme-cofactor complexes assembled on the surface of negatively charged phospholipids that are exposed on activated cells at sites of vascular damage. Activation of the pathway results in the generation of high concentrations of thrombin, which clots the blood. Several anticoagulant mechanisms control the coagulation pathway and under normal conditions the systems are balanced and bleeding
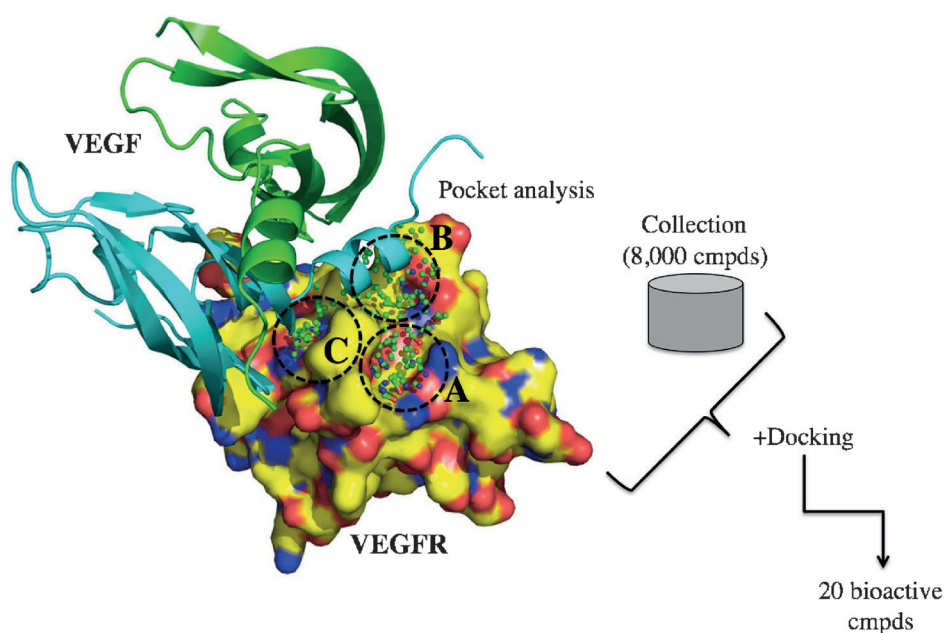
Figure 8. In silico-in vitro screening of the VEGF-VEGFR complex. The crystal structure of residues 8 to 109 of VEGF (cartoon diagram) in complex with VEGFR-1 D2 domain (shown here in solid surface representation; yellow, hydrophobic/aromatic; red, oxygen atom and/or negatively charged; blue, nitrogen atom and/or positively charged) is shown. A probe-mapping algorithm was used to analyze the interface area (green sphere highlights regions where carbon atoms can bind with reasonable affinity, blue spheres represent nitrogen atoms and red spheres, carbonyl groups). Three pockets (A, B, C) could be identified (dashed circles) and are surround by dashed circles. Structure-based virtual screening was carried out over this entire zone and 20 molecules were identified after in vitro studies. The best compound binds directly to the VEGFR-1 D2 domain and inhibits protein—protein interaction.

and thrombosis are avoided.[261–264] Inherited and acquired conditions can tip the pro- anti-coagulant balance resulting in bleeding or thrombosis. The therapeutic principle used for treatment of bleeding disorders such as hemophilia is to supplement the missing coagulation factor, whereas inhibition of coagulation factors is the dominating approach for treatment of thrombosis. An alternative approach for treatment of hemophilia could be to inhibit the anticoagulant pathways. This could decrease the demand for recombinant factor concentrates (e.g. FVIII) and be also beneficial for the treatment of hemophilia patients with inhibitory antibodies. Protein C circulates as a vitamin K-dependent zymogen serine protease that is activated to an active form, activated protein C (APC), by the thrombin-thrombomodulin complex on the surface of endothelial cells. APC has multiple substrates cleaving at several positions in both coagulation cofactors FVa and FVIIIa and in addition APC cleaves also the membrane-bound PAR1 receptor. The cleavage of PAR1 on endothelium results in cytoprotective effects. Thus, APC, is a key component of the protein C anticoagulant pathway and a key regulator of the coagulation cascade. A point mutation in the FV gene (FV Leiden) resulting in the APC-resistance phenotype due to the replacement of Arg506 with Gln is a highly prevalent thrombophilic risk factor.[261–263] The observation that hemophilia patients carrying this mutation have a milder bleeding tendency suggests that inhibition of APC could potentially alleviate the bleeding tendency in hemophilia patients. We

have used a structure-based virtual screening approach to discover drug-like molecules that bind to an exosite of APC (the catalytic site should remain functional as much as possible to carry out the cytoprotective effect) and inhibit the interaction between APC and its substrate FVa.[44] Such molecules could potentially be developed into drugs to treat bleeding disorders. The experimentally determined 3D structure of APC was used and druggable binding pockets were search using several different in silico tools (FTsite, DoGSiteScorer and MetaPocket which combines 8 predictors: LIGSITEcs, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity and POCASA). Potentially interesting sites on APC were identified in one exosite located next to the active site (Figure 9).

Structure-based screening (with the package Surflex) of 50 000 compounds (ChemBridge Diversity set) resulted in the identification of 624 compounds that were then experimentally tested. The ability of these compounds to inhibit the degradation of FVa by APC was used as mean to further select the most potent compounds. After several repeated rounds of testing, the best 20 compounds were tested for direct binding to APC using surface plasmon resonance (SPR). To verify that the compounds specifically bound to the targeted exosites, we took advantage of available recombinant APC variants (i.e., mutations in the exosite) in the SPR analysis. The majority of compounds influenced cleavages in FVa. It remains to investigate whether the compounds affected the degradation of FVIIIa and the
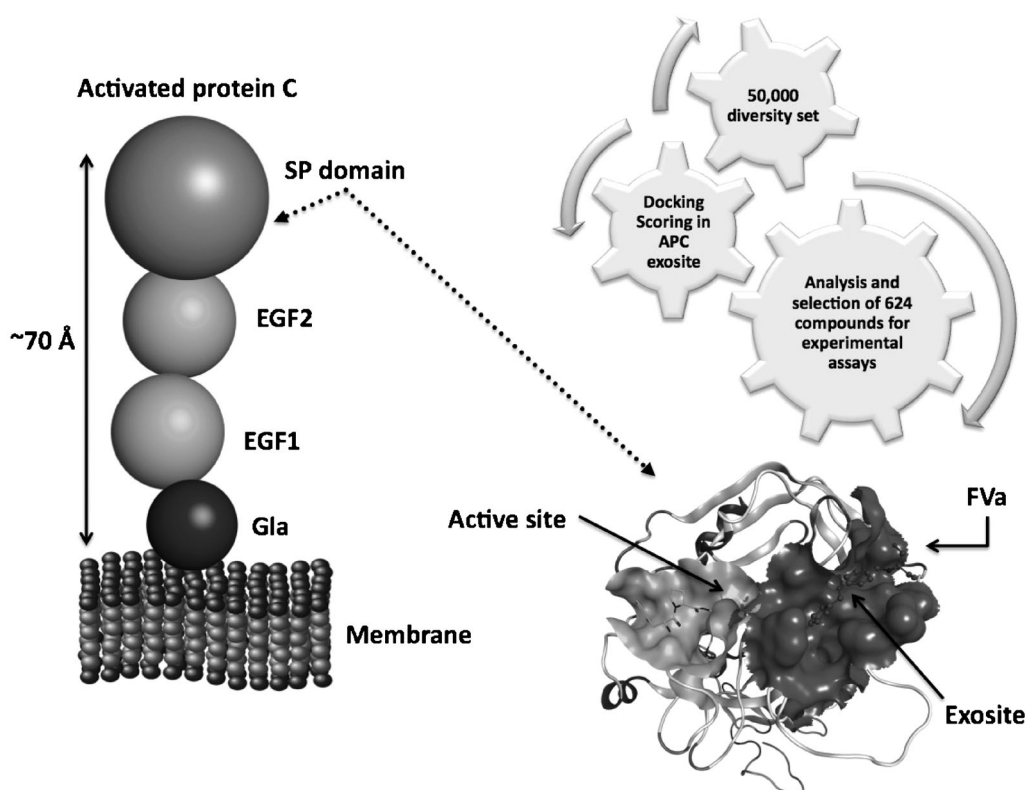
Figure 9. Inhibition of the APC-FVa interaction. A schematic diagram represents the anticoagulant activated protein C (APC) (left). APC is composed of a Gla domain allowing interactions with the appropriate cell membranes, two EGF-like modules and a serine protease (SP) domain. Such organization positions the active site far above the membrane, in the right location to interact with its substrates and partner proteins. A diversity set containing 50 000 molecules was docked in an exosite of APC (right, top) and a possible pose for one active compound is shown (right, bottom). The position of this docked molecule seems reasonable as the compound was not binding properly a mutant protein C that had mutations in the exosite area. This exosite region is known to be important for interacting with the blood coagulation cofactor Va.

cleavage of PAR1. At this stage, these molecules bind to APC with K_d values in the range of 10^{-3}–10^{-5} M and will clearly require optimization. Yet, this work provides a first proof of principle that it may be possible to rationally design small molecules targeting the exosites of APC to achieve inhibition of the anticoagulant protein C system. The future will tell whether this strategy will be a useful approach for the treatment of bleeding disorders.

5 Summary and Outlook

Currently, most small molecule drugs on the market (about 70 % of the about 3500 available drugs) hit only about 400 to 500 targets (most of them being proteins) while functional genomics predicts about 3000 to 10 000 disease modifying "traditional" proteins as potential targets (GPCRs, enzymes and ion channels… many traditional proteins are druggable but unfortunately many are of poor quality with respect to disease rationale).[265,266] As discussed in this review, ongoing investigations of PPI research teams should contribute to expand considerably the number of potential drug targets, much beyond the traditional

ones.[87,267] Small molecules can be used as chemical probes to explore biology and definitively, small compounds dedicated to PPIs will be of great interest to get new insights in the health and disease states. Indeed, most disease-modifying proteins exert their functions through interactions with other proteins. Although PPIs are essential for cellular functions, targeting such interactions with low molecular weight compounds (and if possible orally available molecules) was considered impossible during many years, but, fortunately, several research groups have challenged the dogma. As we gain knowledge about macromolecular complexes, about PPI networks, about the chemistry required to hit such targets, we expect to see more and more modulators of PPIs entering clinical trials and most likely, new drugs acting on this target class will get approved in the coming years. We have also discussed in this review several in silico tools that can be used to assist the rational design of PPI modulators (a simple flowchart is provided Figure 10) and combined with in vitro-in vivo experiments. These in silico methods include PPI network analysis, structural analysis and prediction of the interfaces, druggability predictions, rational design of focused compound collections and various virtual screening computations. Drug re-
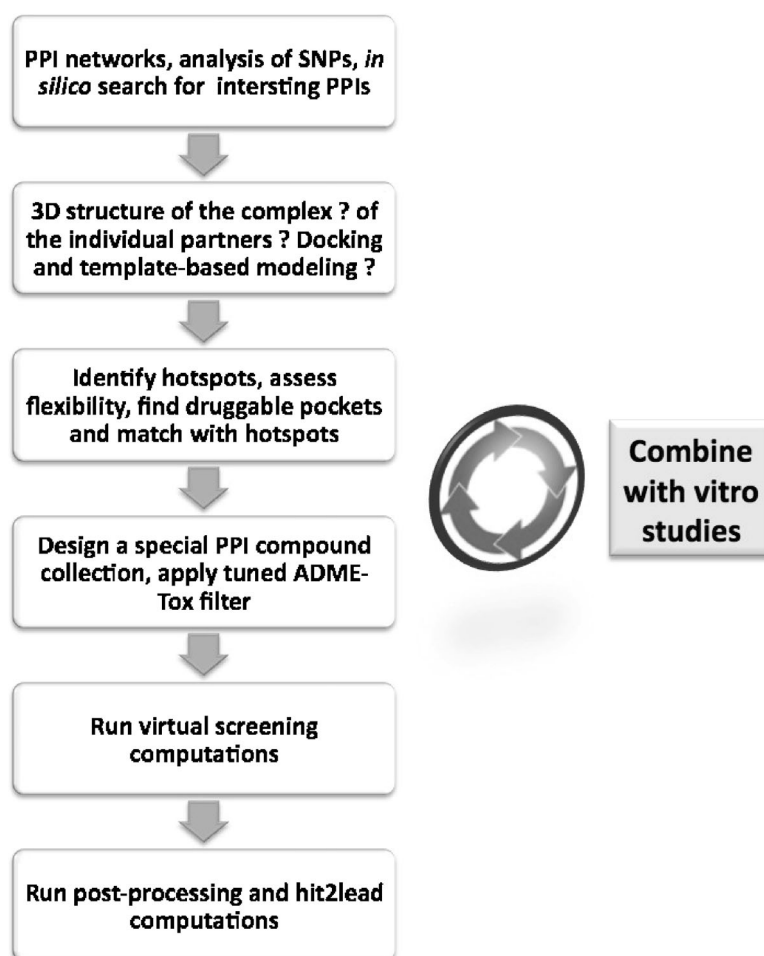
Figure 10. Structure-based design flowchart of orthosteric PPI inhibitors. One possible flowchart (among many) making use of in silico methods to assist the rational design of PPI modulators.

positioning could also be applied to PPIs as illustrated by the discovery of raloxifene and bazedoxifene as novel inhibitors of the IL-6-GP130 interface.[268] With regard to drug discovery, clearly, some biological systems are going to be easier to address with low molecular weight compounds than others just like in the case of enzymes or of other targets in general. The many ongoing in silico developments worldwide combined with the right in vitro-in vivo experiments, and many ongoing clinical studies should definitively contribute to a more efficient and rational discovery of new types of PPI modulators against an ever-increasing number of protein–protein complexes in all therapeutic areas.

The authors declare no conflict of interest.

Acknowledgements

References

[1] P. L. Kastritis, A. M. Bonvin, *J. Roy. Soc. Interf.* **2013**, *10*, 20120835.
[2] P. Braun, A. C. Gingras, *Proteomics* **2012**, *12*, 1478–1498.
[3] J. Janin, R. P. Bahadur, P. Chakrabarti, *Q. Rev. Biophys.* **2008**, *41*, 133–180.
[4] S. Fields, O. Song, *Nature* **1989**, *340*, 245–6.
[5] C. Sanchez, C. Lachaize, F. Janody, B. Bellon, L. Roder, J. Euzenat, F. Rechenmann, B. Jacq, *Nucleic Acids Res.* **1999**, *27*, 89–94.
[6] M. W. Gonzalez, M. G. Kann, *PLoS Comput. Biol.* **2012**, *8*, e1002819.
[7] A. W. White, A. D. Westwell, G. Brahemi, *Expert. Rev. Mol. Med.* **2008**, *10*, e8.
[8] P. L. Toogood, *J. Med. Chem.* **2002**, *45*, 1543–1558.
[9] T. Berg, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 2462–2481.
[10] M. R. Arkin, J. A. Wells, *Nat. Rev. Drug Discov.* **2004**, *3*, 301–317.

[11] P. Chene, *ChemMedChem* **2006**, *1*, 400–411.

[12] J. A. Wells, C. L. McClendon, *Nature* **2007**, *450*, 1001–1009.

[13] A. J. Wilson, *Chem. Soc Rev.* **2009**, *38*, 3289–3300.

[14] A. Domling, *Curr. Opin. Chem. Biol.* **2008**, *12*, 281–291.

[15] B. O. Villoutreix, K. Bastard, O. Sperandio, R. Fahraeus, J. L. Poyet, F. Calvo, B. Deprez, M. A. Miteva, *Curr. Pharm. Biotechnol.* **2008**, *9*, 103–222.

[16] N. Tuncbag, G. Kar, O. Keskin, A. Gursoy, R. Nussinov, *Brief Bioinform.* **2009**, *10*, 217–232.

[17] M. R. Arkin, A. Whitty, *Curr. Opin Chem. Biol.* **2009**, *13*, 284–290.

[18] B. O. Villoutreix, C. M. Labbe, D. Lagorce, G. Laconde, O. Sperandio, *Curr. Pharm. Des.* **2012**, *18*, 4648–4667.

[19] P. Thiel, M. Kaiser, C. Ottmann, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 2012–2018.

[20] C. G. Cummings, A. D. Hamilton, *Curr. Opin. Chem. Biol.* **2010**, *14*, 341–346.

[21] J. Wanner, D. C. Fry, Z. Peng, J. Roberts, *Future Med. Chem.* **2011**, *3*, 2021–2038.

[22] X. Morelli, R. Bourgeas, P. Roche, *Curr. Opin. Chem. Biol.* **2011**, *15*, 475–481.

[23] R. J. Bienstock, *Curr. Pharm. Des.* **2012**, *18*, 1240–1254.

[24] B. Ma, R. Nussinov, *Curr. Pharm. Des.* **2014**, *20*, 1293–1301.

[25] O. Keskin, A. Gursoy, B. Ma, R. Nussinov, *Chem. Rev.* **2008**, *108*, 1225–1244.

[26] L. N. Makley, J. E. Gestwicki, *Chem. Biol. Drug Des.* **2013**, *81*, 22–32.

[27] D. Fry, K. S. Huang, P. Di Lello, P. Mohr, K. Muller, S. S. So, T. Harada, M. Stahl, B. Vu, H. Mauser, *ChemMedChem* **2013**, *8*, 726–732.

[28] A. G. Ngounou Wetie, I. Sokolowska, A. G. Woods, U. Roy, K. Deinhardt, C. C. Darie, *Cell Mol. Life Sci.* **2014**, *71*, 205–228.

[29] F. Falchi, F. Caporuscio, M. Recanatini, *Future Med. Chem.* **2014**, *6*, 343–257.

[30] B. O. Villoutreix, D. Lagorce, C. M. Labbe, O. Sperandio, M. A. Miteva, *Drug Discov. Today* **2013**, *18*, 1081–1089.

[31] P. Koehl, M. Levitt, *Curr. Opin. Struct. Biol.* **1999**, *9*, 155–156.

[32] A. Warshel, P. K. Sharma, M. Kato, W. W. Parson, *Biochim. Biophys. Acta* **2006**, *1764*, 1647–1676.

[33] T. Lazaridis, M. Karplus, *Biophys. Chem.* **2003**, *100*, 367–395.

[34] N. London, B. Raveh, O. Schueler-Furman, *Curr. Opin. Chem. Biol.* **2013**, *17*, 952–959.

[35] A. Loregian, G. Palu, *J. Cell Physiol.* **2005**, *204*, 750–762.

[36] J. Mallinson, I. Collins, *Future Med. Chem.* **2012**, *4*, 1409–1438.

[37] A. P. Higueruelo, H. Jubb, T. L. Blundell, *Curr. Opin. Pharmacol.* **2013**, *13*, 791–796.

[38] D. J. Craik, D. P. Fairlie, S. Liras, D. Price, *Chem. Biol. Drug Des.* **2013**, *81*, 136–147.

[39] K. Imai, A. Takaoka, *Nat. Rev. Cancer* **2006**, *6*, 714–727.

[40] B. Meunier, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 8702–8706.

[41] A. Kumar, A. Voet, K. Y. Zhang, *Curr. Med. Chem.* **2012**, *19*, 5128–5147.

[42] A. Winter, A. P. Higueruelo, M. Marsh, A. Sigurdardottir, W. R. Pitt, T. L. Blundell, *Q. Rev. Biophys.* **2012**, *45*, 383–426.

[43] B. Gautier, M. A. Miteva, V. Goncalves, F. Huguenot, P. Coric, S. Bouaziz, B. Seijo, J. F. Gaucher, I. Broutin, C. Garbay, A. Lesnard, S. Rault, N. Inguimbert, B. O. Villoutreix, M. Vidal, *Chem. Biol.* **2011**, *18*, 1631–1639.

[44] O. Sperandio, K. C. Wildhagen, R. Schrijver, S. Wielders, B. O. Villoutreix, G. A. Nicolaes, *Thromb. Res.* **2014**, *133*, 1105–1114.

[45] S. J. Wodak, J. Vlasblom, A. L. Turinsky, S. Pu, *Curr. Opin. Struct. Biol.* **2013**, *23*, 941–53.

[46] R. Mosca, T. Pons, A. Ceol, A. Valencia, P. Aloy, *Curr. Opin. Struct. Biol.* **2013**, *23*, 929–940.

[47] S. Orchard, *Proteomics* **2012**, *12*, 1656–1662.

[48] L. J. Jensen, P. Bork, *Science* **2008**, *322*, 56–57.

[49] H. Hermjakob, L. Montecchi-Palazzi, G. Bader, J. Wojcik, L. Salwinski, A. Ceol, S. Moore, S. Orchard, U. Sarkans, C. von Mering, B. Roechert, S. Poux, E. Jung, H. Mersch, P. Kersey, M. Lappe, Y. Li, R. Zeng, D. Rana, M. Nikolski, H. Husi, C. Brun, K. Shanker, S. G. Grant, C. Sander, P. Bork, W. Zhu, A. Pandey, A. Brazma, B. Jacq, M. Vidal, D. Sherman, P. Legrain, G. Cesareni, I. Xenarios, D. Eisenberg, B. Steipe, C. Hogue, R. Apweiler, *Nat. Biotechnol.* **2004**, *22*, 177–183.

[50] S. Orchard, S. Kerrien, S. Abbani, B. Aranda, J. Bhate, S. Bidwell, A. Bridge, L. Briganti, F. S. Brinkman, G. Cesareni, A. Chatr-aryamontri, E. Chautard, C. Chen, M. Dumousseau, J. Goll, R. E. Hancock, L. I. Hannick, I. Jurisica, J. Khadake, D. J. Lynn, U. Mahadevan, L. Perfetto, A. Raghunath, S. Ricard-Blum, B. Roechert, L. Salwinski, V. Stumpflen, M. Tyers, P. Uetz, I. Xenarios, H. Hermjakob, *Nat. Meth.* **2012**, *9*, 345–350.

[51] A. L. Turinsky, S. Razick, B. Turner, I. M. Donaldson, S. J. Wodak, *Meth. Mol. Biol.* **2014**, *1091*, 315–331.

[52] L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, D. Eisenberg, *Nucleic Acids Res.* **2004**, *32*, D449–D451.

[53] R. Isserlin, R. A. El-Badrawi, G. D. Bader, *Database (Oxford)*, **2011**, *2011*, baq037.

[54] 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.* **2012**, *40*, D857–D861.

[55] P. Pagel, S. Kovac, M. Oesterheld, B. Brauner, I. Dunger-Kaltenbach, G. Frishman, C. Montrone, P. Mark, V. Stumpflen, H. W. Mewes, A. Ruepp, D. Frishman, *Bioinformatics* **2005**, *21*, 832–834.

[56] R. Kumar, B. Nanduri, *BMC Bioinformatics* **2010**, *11 Suppl 6*, S16.

[57] S. Kerrien, B. Aranda, L. Breuza, A. Bridge, F. Broackes-Carter, C. Chen, M. Duesbury, M. Dumousseau, M. Feuermann, U. Hinz, C. Jandrasits, R. C. Jimenez, J. Khadake, U. Mahadevan, P. Masson, I. Pedruzzi, E. Pfeiffenberger, P. Porras, A. Raghunath, B. Roechert, S. Orchard, H. Hermjakob, *Nucleic Acids Res.* **2012**, *40*, D841–D846.

[58] C. Stark, B. J. Breitkreutz, A. Chatr-Aryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Reguly, J. M. Rust, A. Winter, K. Dolinski, M. Tyers, *Nucleic Acids Res.* **2011**, *39*, D698–D704.

[59] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, L. J. Jensen, *Nucleic Acids Res.* **2013**, *41*, D808–D815.

[60] M. Higurashi, T. Ishida, K. Kinoshita, *Nucleic Acids Res.* **2009**, *37*, D360–D364.

[61] A. Stein, A. Ceol, P. Aloy, *Nucleic Acids Res.* **2011**, *39*, D718–23.

[62] R. D. Finn, M. Marshall, A. Bateman, *Bioinformatics* **2005**, *21*, 410–412.

[63] A. W. Ghoorah, M. D. Devignes, M. Smail-Tabbone, D. W. Ritchie, *Nucleic Acids Res.* **2014**, *42*, D389–D395.

[64] M. J. Meyer, J. Das, X. Wang, H. Yu, *Bioinformatics* **2013**, *29*, 1577–1579.

[65] R. Rid, W. Strasser, D. Siegl, C. Frech, M. Kommenda, T. Kern, H. Hintner, J. W. Bauer, K. Onder, *Assay Drug Dev. Technol.* **2013**, *11*, 333–346.

[66] N. del-Toro, M. Dumousseau, S. Orchard, R. C. Jimenez, E. Galeota, G. Launay, J. Goll, K. Breuer, K. Ono, L. Salwinski, H. Hermjakob, *Nucleic Acids Res.* **2013**, *41*, W601–W606.

[67] H. Blankenburg, R. D. Finn, A. Prlic, A. M. Jenkinson, F. Ramirez, D. Emig, S. E. Schelhorn, J. Buch, T. Lengauer, M. Albrecht, *Bioinformatics*, **2009**, *25*, 1321–1328.

[68] R. K. Kalathur, J. P. Pinto, M. A. Hernandez-Prieto, R. S. Machado, D. Almeida, G. Chaurasia, M. E. Futschik, *Nucleic Acids Res.* **2014**, *42*, D408–D414.

[69] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res.* **2003**, *13*, 2498–2504.

[70] K. Venkatesan, J. F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K. I. Goh, M. A. Yildirim, N. Simonis, K. Heinzmann, F. Gebreab, J. M. Sahalie, S. Cevik, C. Simon, A. S. de Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R. R. Murray, C. Lin, M. Lalowski, J. Timm, K. Rau, C. Boone, P. Braun, M. E. Cusick, F. P. Roth, D. E. Hill, J. Tavernier, E. E. Wanker, A. L. Barabasi, M. Vidal, *Nat. Meth.* **2009**, *6*, 83–90.

[71] M. P. Stumpf, T. Thorne, E. de Silva, R. Stewart, H. J. An, M. Lappe, C. Wiuf, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6959–6964.

[72] M. Vidal, M. E. Cusick, A. L. Barabasi, *Cell* **2011**, *144*, 986–998.

[73] L. I. Furlong, *Trends Genet.* **2013**, *29*, 150–159.

[74] H. Ruffner, A. Bauer, T. Bouwmeester, *Drug Discov. Today* **2007**, *12*, 709–716.

[75] P. Csermely, T. Korcsmaros, H. J. Kiss, G. London, R. Nussinov, *Pharmacol. Ther.* **2013**, *138*, 333–408.

[76] D. C. Fung, S. S. Li, A. Goel, S. H. Hong, M. R. Wilkins, *Proteomics* **2012**, *12*, 1669–1686.

[77] G. L. Zhang, D. S. DeLuca, V. Brusic, *Meth. Mol. Biol.* **2011**, *723*, 349–364.

[78] L. Geffers, B. Herrmann, G. Eichele, *Mamm. Genome* **2012**, *23*, 525–538.

[79] J. C. Huling, T. Pisitkun, J. H. Song, M. J. Yu, J. D. Hoffert, M. A. Knepper, *Am. J. Physiol. Renal. Physiol.* **2012**, *302*, F401–F407.

[80] A. Koutsoukas, B. Simms, J. Kirchmair, P. J. Bond, A. V. Whitmore, S. Zimmer, M. P. Young, J. L. Jenkins, M. Glick, R. C. Glen, A. Bender, *J. Proteomics* **2011**, *74*, 2554–74.

[81] J. U. Peters, *J. Med. Chem.* **2013**, *56*, 8955–8971.

[82] J. Mestres, E. Gregori-Puigjane, *Trends Pharmacol. Sci.* **2009**, *30*, 470–474.

[83] U. Koch, M. Hamacher, P. Nussbaumer, *Biochim. Biophys. Acta* **2014**, *1844*, 156–161.

[84] J. Janin, C. Chothia, *J. Biol. Chem*, **1990**, *265*, 16027–1630.

[85] S. Jones, J. M. Thornton, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13–20.

[86] I. M. Nooren, J. M. Thornton, *EMBO J.* **2003**, *22*, 3486–3492.

[87] S. Surade, T. L. Blundell, *Chem. Biol.* **2012**, *19*, 42–50.

[88] H. Jubb, A. P. Higueruelo, A. Winter, T. L. Blundell, *Trends Pharmacol. Sci.* **2012**, *33*, 241–248.

[89] A. Metz, E. Ciglia, H. Gohlke, *Curr. Pharm. Des.* **2012**, *18*, 4630–4647.

[90] E. D. Levy, *J. Mol. Biol.* **2010**, *403*, 660–670.

[91] E. Cukuroglu, A. Gursoy, O. Keskin, *Ann. Biomed. Eng.* **2010**, *38*, 2068–2078.

[92] G. Kar, A. Gursoy, O. Keskin, *PLoS Comput. Biol.* **2009**, *5*, e1000601.

[93] T. Clackson, J. A. Wells, *Science* **1995**, *267*, 383–386.

[94] R. R. Thangudu, S. H. Bryant, A. R. Panchenko, T. Madej, *J. Mol. Biol.* **2012**, *415*, 443–453.

[95] B. S. Zerbe, D. R. Hall, S. Vajda, A. Whitty, D. Kozakov, *J. Chem. Inf. Model.* **2012**, *52*, 2236–2244.

[96] A. A. Bogan, K. S. Thorn, *J. Mol. Biol.* **1998**, *280*, 1–9.

[97] J. Fernandez-Recio, *WIREs Comput. Mol. Sci.* **2011**, *1*, 680–698.

[98] I. S. Moreira, J. M. Martins, R. M. Ramos, P. A. Fernandes, M. J. Ramos, *Biochim. Biophys. Acta* **2013**, *1834*, 404–414.

[99] D. Rajamani, S. Thiel, S. Vajda, C. J. Camacho, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11287–11292.

[100] L. Garma, S. Mukherjee, P. Mitra, Y. Zhang, *PLoS One* **2012**, *7*, e38913.

[101] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.

[102] M. Gao, J. Skolnick, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 22517–22522.

[103] Q. C. Zhang, D. Petrey, R. Norel, B. H. Honig, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10896–10901.

[104] P. J. Kundrotas, Z. Zhu, J. Janin, I. A. Vakser, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 9438–41.

[105] M. C. Smith, J. E. Gestwicki, *Expert Rev. Mol. Med.* **2012**, *14*, e16.

[106] M. Raj, B. N. Bullock, P. S. Arora, *Bioorg. Med. Chem.* **2012**, *21*, 4051–4057.

[107] T. L. Nero, C. J. Morton, J. K. Holien, J. Wielens, M. W. Parker, *Nat. Rev. Cancer* **2014**, *14*, 248–62.

[108] C. M. Yates, M. J. Sternberg, *J. Mol. Biol.* **2013**, *425*, 3949–3963.

[109] Z. Zhang, S. Witham, M. Petukh, G. Moroy, M. Miteva, Y. Ikeguchi, E. Alexov, *J. Am. Med. Inform. Assoc.* **2013**, *20*, 643–651.

[110] A. David, R. Razali, M. N. Wass, M. J. Sternberg, *Hum. Mutat.* **2012**, *33*, 359–363.

[111] I. H. Moal, J. Fernandez-Recio, *Bioinformatics* **2012**, *28*, 2600–7.

[112] S. Leis, S. Schneider, M. Zacharias, *Curr. Med. Chem*, **2010**, *17*, 1550–1562.

[113] S. Grosdidier, J. Fernandez-Recio, *Curr. Pharm. Des.* **2012**, *18*, 4607–4618.

[114] Q. C. Zhang, D. Petrey, L. Deng, L. Qiang, Y. Shi, C. A. Thu, B. Bisikirska, C. Lefebvre, D. Accili, T. Hunter, T. Maniatis, A. Califano, B. Honig, *Nature* **2012**, *490*, 556–560.

[115] Y. Ofran, B. Rost, *Bioinformatics* **2007**, *23*, e13–e16.

[116] B. K. Sriwastava, S. Basu, U. Maulik, D. Plewczynski, *J. Mol. Model.* **2013**, *19*, 4059–4070.

[117] A. Porollo, J. Meller, *Proteins* **2007**, *66*, 630–645.

[118] H. Neuvirth, R. Raz, G. Schreiber, *J. Mol. Biol.* **2004**, *338*, 181–199.

[119] S. Qin, H. X. Zhou, *Bioinformatics* **2007**, *23*, 3386–3387.

[120] N. Tuncbag, O. Keskin, A. Gursoy, *Nucleic Acids Res.* **2010**, *38*, W402–W406.

[121] D. E. Kim, D. Chivian, D. Baker, *Nucleic Acids Res.* **2004**, *32*, W526–W531.

[122] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, *Nucleic Acids Res.* **2005**, *33*, W382–W388.

[123] T. Geppert, B. Hoy, S. Wessler, G. Schneider, *Chem. Biol.* **2011**, *18*, 344–353.

[124] S. Grosdidier, J. Fernandez-Recio, *BMC Bioinformatics*, **2008**, *9*, 447.

[125] C.-Y. Yang, S. Wang, *ACS Med. Chem. Lett.* **2011**, *2*, 280–284.

[126] S. Eyrisch, V. Helms, *J. Med. Chem.* **2007**, *50*, 3457–3464.

[127] C. M. Baker, R. B. Best, *WIREs Computational Molecular Science*, **2014**, *4*, 182–198.

[128] O. Taboureau, J. B. Baell, J. Fernandez-Recio, B. O. Villoutreix, *Chem. Biol.* **2012**, *19*, 29–41.

[129] H. Hwang, T. Vreven, Z. Weng, *Proteins* **2014**, *82*, 57–66.

[130] A. Szilagyi, Y. Zhang, *Curr. Opin. Struct. Biol.* **2014**, *24*, 10–23.

[131] S. Y. Huang, *Drug Discov. Today* **2014**.

[132] P. Popov, D. W. Ritchie, S. Grudinin, *Proteins* **2014**, *82*, 34–44.

[133] S. de Vries, M. Zacharias, *Proteins*, **2013**, *81*, 2167–2174.

[134] Y. Matsuzaki, N. Uchikoga, M. Ohue, T. Shimoda, T. Sato, T. Ishida, Y. Akiyama, *Source Code Biol. Med.* **2013**, *8*, 18.

[135] B. Jimenez-Garcia, C. Pons, J. Fernandez-Recio, *Bioinformatics* **2013**, *29*, 1698–1699.

[136] R. Chowdhury, M. Rasheed, D. Keidel, M. Moussalem, A. Olson, M. Sanner, C. Bajaj, *PLoS One* **2013**, *8*, e51307.

[137] M. Torchala, I. H. Moal, R. A. Chaleil, J. Fernandez-Recio, P. A. Bates, *Bioinformatics* **2013**, *29*, 807–809.

[138] L. C. Xue, R. A. Jordan, Y. El-Manzalawy, D. Dobbs, V. Honavar, *Proteins* **2014**, *82*, 250–267.

[139] S. Mukherjee, Y. Zhang, *Structure* **2011**, *19*, 955–966.

[140] R. Singh, D. Park, J. Xu, R. Hosur, B. Berger, *Nucleic Acids Res.* **2010**, *38*, W508–W515.

[141] F. N. Edfeldt, R. H. Folmer, A. L. Breeze, *Drug Discov. Today*, **2011**, *16*, 284–7.

[142] R. P. Sheridan, V. N. Maiorov, M. K. Holloway, W. D. Cornell, Y. D. Gao, *J. Chem. Inf. Model.* **2010**, *50*, 2029–2040.

[143] J. C. Fuller, N. J. Burgoyne, R. M. Jackson, *Drug Discov. Today* **2009**, *14*, 155–161.

[144] K. Segers, O. Sperandio, M. Sack, R. Fischer, M. A. Miteva, J. Rosing, G. A. Nicolaes, B. O. Villoutreix, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12697–12702.

[145] M. Rueda, G. Bottegoni, R. Abagyan, *J. Chem. Inf. Model.* **2010**, *50*, 186–193.

[146] O. Sperandio, L. Mouawad, E. Pinto, B. O. Villoutreix, D. Perahia, M. A. Miteva, *Eur. Biophys. J.* **2010**, *39*, 1365–1372.

[147] O. Ulucan, S. Eyrisch, V. Helms, *Curr. Pharm. Des.* **2012**, *18*, 4599–4606.

[148] C. A. Sotriffer, *Curr. Top Med. Chem.* **2011**, *11*, 179–191.

[149] S. Perot, O. Sperandio, M. A. Miteva, A. C. Camproux, B. O. Villoutreix, *Drug Discov. Today* **2010**, *15*, 656–667.

[150] I. Kufareva, Y. C. Chen, A. V. Ilatovskiy, R. Abagyan, *Curr. Top Med. Chem.* **2012**, *12*, 1869–1882.

[151] X. Zheng, L. Gan, E. Wang, J. Wang, *AAPS J.* **2012**, *15*, 228–241.

[152] E. B. Fauman, B. K. Rai, E. S. Huang, *Curr. Opin. Chem. Biol.* **2011**, *15*, 463–468.

[153] J. Y. Trosset, N. Vodovar, *Meth. Mol. Biol.* **2013**, *986*, 141–164.

[154] B. H. Dessailly, N. L. Dawson, K. Mizuguchi, C. A. Orengo, *Biochim. Biophys. Acta* **2013**, *1834*, 874–889.

[155] M. Wirth, A. Volkamer, V. Zoete, F. Rippmann, O. Michielin, M. Rarey, W. H. Sauer, *J. Comput. Aided Mol. Des.* **2013**, *27*, 511–524.

[156] K. Chen, M. J. Mizianty, J. Gao, L. Kurgan, *Structure* **2011**, *19*, 613–621.

[157] D. Kozakov, D. R. Hall, G. Y. Chuang, R. Cencic, R. Brenke, L. E. Grove, D. Beglov, J. Pelletier, A. Whitty, S. Vajda, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13528–13533.

[158] L. E. Grove, D. R. Hall, D. Beglov, S. Vajda, D. Kozakov, *Bioinformatics* **2013**, *29*, 1218–1219.

[159] L. Votapka, R. E. Amaro, *Bioinformatics* **2013**, *29*, 393–394.

[160] L. M. Meireles, A. S. Domling, C. J. Camacho, *Nucleic Acids Res.* **2010**, *38*, W407–W411.

[161] M. J. Basse, S. Betzi, R. Bourgeas, S. Bouzidi, B. Chetrit, V. Hamon, X. Morelli, P. Roche, *Nucleic Acids Res.* **2013**, *41*, D824–7.

[162] D. R. Koes, C. J. Camacho, *Bioinformatics* **2012**, *28*, 784–791.

[163] F. P. Davis, *Mol. Biosyst.* **2011**, *7*, 545–557.

[164] F. P. Davis, A. Sali, *PLoS Comput. Biol.* **2010**, *6*, e1000668.

[165] A. Metz, C. Pfleger, H. Kopitz, S. Pfeiffer-Marek, K. H. Baringhaus, H. Gohlke, *J. Chem. Inf. Model.* **2012**, *52*, 120–133.

[166] P. Anand, K. Yeturu, N. Chandra, *Nucleic Acids Res.* **2012**, *40*, W400–W408.

[167] M. Gao, J. Skolnick, *Bioinformatics* **2013**, *29*, 597–604.

[168] Y. Lin, S. Yoo, R. Sanchez, *Bioinformatics* **2012**, *28*, 1172–1173.

[169] I. Kufareva, A. V. Ilatovskiy, R. Abagyan, *Nucleic Acids Res.* **2012**, *40*, D535–D540.

[170] E. Cukuroglu, A. Gursoy, R. Nussinov, O. Keskin, *PLoS One* **2014**, *9*, e86738.

[171] V. Azzarito, K. Long, N. S. Murphy, A. J. Wilson, *Nat. Chem.* **2013**, *5*, 161–173.

[172] Z. J. Lu, S. J. Deng, D. G. Huang, Y. He, M. Lei, L. Zhou, P. Jin, *World J. Biol. Chem.* **2012**, *3*, 187–196.

[173] P. Buchwald, *IUBMB Life* **2010**, *62*, 724–731.

[174] E. i. CML, *Blood* **2013**, *121*, 4439–4442.

[175] S. J. Teague, *Drug Discov Today* **2011**, *16*, 398–411.

[176] Y. Westermaier, M. Veurink, T. Riis-Johannessen, S. Guinchard, R. Gurny, L. Scapozza, *Eur. J. Pharm. Biopharm.* **2013**, *85*, 773–780.

[177] A. P. Higueruelo, A. Schreyer, G. R. Bickerton, W. R. Pitt, C. R. Groom, T. L. Blundell, *Chem. Biol Drug Des.* **2009**, *74*, 457–467.

[178] O. Sperandio, C. H. Reynes, A. C. Camproux, B. O. Villoutreix, *Drug Discov. Today* **2010**, *15*, 220–229.

[179] L. Pagliaro, J. Felding, K. Audouze, S. J. Nielsen, R. B. Terry, C. Krog-Jensen, S. Butcher, *Curr. Opin. Chem. Biol.* **2004**, *8*, 442–449.

[180] Y. Pommier, J. Cherfils, *Trends Pharmacol. Sci.* **2005**, *26*, 138–145.

[181] L. Jin, W. Wang, G. Fang, *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 435–456.

[182] A. Schon, S. Y. Lam, E. Freire, *Future Med. Chem.* **2011**, *3*, 1129–1137.

[183] G. Zinzalla, D. E. Thurston, *Future Med. Chem.* **2009**, *1*, 65–93.

[184] A. P. Higueruelo, H. Jubb, T. L. Blundell, *Database (Oxford)* **2013**, *2013*, bat039.

[185] C. M. Labbe, G. Laconde, M. A. Kuenemann, B. O. Villoutreix, O. Sperandio, *Drug Discov. Today* **2013**, *18*, 958–968.

[186] A. Neugebauer, R. W. Hartmann, C. D. Klein, *J. Med. Chem.* **2007**, *50*, 4665–4668.

[187] C. Reynes, H. Host, A. C. Camproux, G. Laconde, F. Leroux, A. Mazars, B. Deprez, R. Fahraeus, B. O. Villoutreix, O. Sperandio, *PLoS Comput. Biol.* **2010**, *6*, e1000695.

[188] D. Lagorce, J. Maupetit, J. Baell, O. Sperandio, P. Tuffery, M. A. Miteva, H. Galons, B. O. Villoutreix, *Bioinformatics* **2011**, *27*, 2018–2020.

[189] V. Hamon, R. Bourgeas, P. Ducrot, I. Theret, L. Xuereb, M. J. Basse, J. M. Brunel, S. Combes, X. Morelli, P. Roche, *J. Roy. Soc. Interf.* **2013**, *11*, 20130860.

[190] Q. Wang, M. Zheng, Z. Huang, X. Liu, H. Zhou, Y. Chen, T. Shi, J. Zhang, *J. Mol. Graph. Model.* **2012**, *38*, 324–333.

[191] C. G. Bologa, M. M. Olah, T. I. Oprea, *Meth. Mol. Biol.* **2006**, *316*, 375–388.

[192] M. P. Gleeson, *J. Med. Chem.* **2008**, *51*, 817–34.

[193] P. D. Leeson, S. A. St-Gallay, *Nat. Rev. Drug Discov.* **2011**, *10*, 749–765.

[194] M. M. Hann, G. M. Keseru, *Nat. Rev. Drug Discov.* **2012**, *11*, 355–365.

[195] T. Bohnert, L. S. Gan, *J. Pharm. Sci.* **2013**, *102*, 2953–2994.

[196] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.

[197] J. D. Hughes, J. Blagg, D. A. Price, S. Bailey, G. A. Decrescenzo, R. V. Devraj, E. Ellsworth, Y. M. Fobian, M. E. Gibbs, R. W. Gilles, N. Greene, E. Huang, T. Krieger-Burke, J. Loesel, T. Wager, L. Whiteley, Y. Zhang, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4872–4875.

[198] A. Dahan, J. M. Miller, G. L. Amidon, *AAPS J.* **2009**, *11*, 740–746.

[199] K. Valko, E. Chiarparin, S. Nunhuck, D. Montanari, *J. Pharm. Sci.* **2012**, *101*, 4155–4169.

[200] E. Perola, *J. Med. Chem.* **2010**, *53*, 2986–2997.

[201] T. T. Wager, R. Y. Chandrasekaran, X. Hou, M. D. Troutman, P. R. Verhoest, A. Villalobos, Y. Will, *ACS Chem. Neurosci.* **2010**, *1*, 420–434.

[202] M. M. Hann, *Med. Chem. Commun.* **2011**, *2*, 349–355.

[203] B. Faller, G. Ottaviani, P. Ertl, G. Berellini, A. Collis, *Drug Discov. Today* **2011**, *16*, 976–984.

[204] H. Zhao, *Drug Discov. Today* **2011**, *16*, 158–163.

[205] M. Bruncko, S. K. Tahir, X. Song, J. Chen, H. Ding, J. R. Huth, S. Jin, R. A. Judge, D. J. Madar, C. H. Park, C. M. Park, A. M. Petros, C. Tse, S. H. Rosenberg, S. W. Elmore, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 7503–7506.

[206] A. Ettorre, P. D'Andrea, S. Mauro, M. Porcelloni, C. Rossi, M. Altamura, R. M. Catalioto, S. Giuliani, C. A. Maggi, D. Fattori, *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1807–1809.

[207] A. Alex, D. S. Millan, M. Perez, F. Wakenhut, G. A. Whitlock, *Med. Chem. Commun.* **2011**, *2*, 669–674.

[208] A. P. Higueruelo, A. Schreyer, G. R. Bickerton, T. L. Blundell, W. R. Pitt, *PLoS One* **2012**, *7*, e51742.

[209] A. Ganesan, *Curr. Opin. Chem. Biol.* **2008**, *12*, 306–317.

[210] R. A. Bauer, J. M. Wurst, D. S. Tan, *Curr. Opin. Chem. Biol.* **2010**, *14*, 308–314.

[211] M. D. Segall, *Curr. Pharm. Des.* **2012**, *18*, 1292–1310.

[212] D. Horvath, *J. Med .Chem.* **1997**, *40*, 2412–2423.

[213] S. S. Phatak, C. C. Stephan, C. N. Cavasotto, *Expert Opin. Drug Discov.* **2009**, *4*, 947–059.

[214] B. K. Shoichet, *Nature*, **2004**, *432*, 862–5.

[215] M. Totrov, R. Abagyan, *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–84.

[216] C. McInnes, *Curr. Opin Chem. Biol.* **2007**, *11*, 494–502.

[217] D. Stumpfe, P. Ripphausen, J. Bajorath, *Future Med. Chem.* **2012**, *4*, 593–602.

[218] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, Baskin II, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, *J. Med. Chem.* **2014**, in press.

[219] G. Schneider, *Nat. Rev. Drug Discov.* **2010**, *9*, 273–276.

[220] S. Ekins, J. Mestres, B. Testa, *Br. J. Pharmacol.* **2007**, *152*, 9–20.

[221] S. Ekins, J. Mestres, B. Testa, *Br. J. Pharmacol.* **2007**, *152*, 21–37.

[222] K. Heikamp, J. Bajorath, *Chem. Biol. Drug Des.* **2013**, *81*, 33–40.

[223] S. Cross, G. Cruciani, *Drug Discov. Today* **2010**, *15*, 23–32.

[224] P. G. Polishchuk, T. I. Madzhidov, A. Varnek, *J. Comput. Aided Mol. Des.* **2013**, *27*, 675–679.

[225] L. Ruddigkeit, R. van Deursen, L. C. Blum, J. L. Reymond, *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875.

[226] T. Varin, M. C. Didiot, C. N. Parker, A. Schuffenhauer, *J. Med. Chem.* **2012**, *55*, 1161–1170.

[227] Y. Tanrikulu, B. Kruger, E. Proschak, *Drug Discov. Today* **2013**, *18*, 358–364.

[228] R. S. Ferreira, A. Simeonov, A. Jadhav, O. Eidam, B. T. Mott, M. J. Keiser, J. H. McKerrow, D. J. Maloney, J. J. Irwin, B. K. Shoichet, *J. Med. Chem.* **2010**, *53*, 4891–4905.

[229] X. Marechal, E. Genin, L. Qin, O. Sperandio, M. Montes, N. Basse, N. Richy, M. A. Miteva, M. Reboud-Ravaux, J. Vidal, B. O. Villoutreix, *Curr. Med. Chem.* **2013**, *20*, 2351–62.

[230] S. Ozcan, A. Kazi, F. Marsilio, B. Fang, W. C. Guida, J. Koomen, H. R. Lawrence, S. M. Sebti, *J. Med. Chem.* **2013**, *56*, 3783–3805.

[231] T. Scior, A. Bender, G. Tresadern, J. L. Medina-Franco, K. Martinez-Mayorga, T. Langer, K. Cuanalo-Contreras, D. K. Agrafiotis, *J. Chem. Inf. Model.* **2012**, *52*, 867–881.

[232] P. Ripphausen, D. Stumpfe, J. Bajorath, *Future Med. Chem.* **2012**, *4*, 603–613.

[233] D. E. Clark, *Expert Opin. Drug Discov.* **2008**, *3*, 841–851.

[234] T. Zhu, S. Cao, P. C. Su, R. Patel, D. Shah, H. B. Chokshi, R. Szukala, M. E. Johnson, K. E. Hevener, *J. Med. Chem.* **2013**, *56*, 6560–6572.

[235] U. Rester, *Curr. Opin. Drug Discov. Devel.* **2008**, *11*, 559–568.

[236] A. N. Jain, *Curr. Opin. Drug Discov. Devel.* **2004**, *7*, 396–403.

[237] C. N. Cavasotto, S. S. Phatak, *Drug Discov. Today* **2009**, *14*, 676–683.

[238] J. Skolnick, H. Zhou, M. Gao, *Curr. Opin. Struct. Biol.* **2013**, *23*, 191–197.

[239] H. Fan, J. J. Irwin, B. M. Webb, G. Klebe, B. K. Shoichet, A. Sali, *J. Chem. Inf. Model.* **2009**, *49*, 2512–2527.

[240] T. Pencheva, D. Jereva, M. A. Miteva, I. Pajeva, *Curr. Comput. Aided Drug Des.* **2013**, *9*, 83–94.

[241] T. Pencheva, D. Lagorce, I. Pajeva, B. O. Villoutreix, M. A. Miteva, *BMC Bioinformatics* **2008**, *9*, 438.

[242] E. Yuriev, P. A. Ramsland, *J. Mol. Recognit.* **2013**, *26*, 215–239.

[243] O. Sperandio, M. A. Miteva, B. O. Villoutreix, *Curr. Comput. Aided Drug Des.* **2008**, *4*, 250–258.

[244] B. O. Villoutreix, R. Eudes, M. A. Miteva, *Comb. Chem. High Throughput Screen.* **2009**, *12*, 1000–1016.

[245] C. Liao, M. Sitzmann, A. Pugliese, M. C. Nicklaus, *Future Med. Chem.* **2011**, *3*, 1057–1085.

[246] A. Voet, E. F. Banwell, K. K. Sahu, J. G. Heddle, K. Y. Zhang, *Curr. Top Med. Chem.* **2013**, *13*, 989–1001.

[247] J. B. Baell, G. A. Holloway, *J. Med Chem*, **2010**, *53*, 2719–2740.

[248] A. Whitty, *Future Med. Chem.* **2011**, *3*, 797–801.

[249] N. Y. Mok, S. Maxe, R. Brenk, *J. Chem. Inf. Model.* **2013**, *53*, 534–544.

[250] L. A. Arnold, E. Estebanez-Perpina, M. Togashi, N. Jouravel, A. Shelat, A. C. McReynolds, E. Mar, P. Nguyen, J. D. Baxter, R. J. Fletterick, P. Webb, R. K. Guy, *J. Biol. Chem.* **2005**, *280*, 43048–43055.

[251] D. M. Kruger, G. Jessen, H. Gohlke, *J. Chem. Inf. Model.* **2012**, *52*, 2807–2811.

[252] R. Gowthaman, E. J. Deeds, J. Karanicolas, *J. Chem. Inf. Model.* **2013**, *53*, 2073–2081.

[253] A. L. Garner, K. D. Janda, *Curr. Top Med. Chem.* **2011**, *11*, 258–280.

[254] A. A. Ivanov, F. R. Khuri, H. Fu, *Trends Pharmacol. Sci.* **2013**, *34*, 393–400.

[255] N. Ferrara, R. S. Kerbel, *Nature* **2005**, *438*, 967–974.

[256] S. Kanno, N. Oda, M. Abe, Y. Terai, M. Ito, K. Shitara, K. Tabayashi, M. Shibuya, Y. Sato, *Oncogene* **2000**, *19*, 2138–2146.

[257] K. Holmes, O. L. Roberts, A. M. Thomas, M. J. Cross, *Cell Signal.* **2007**, *19*, 2003–2012.

[258] C. Reynolds, D. Damerell, S. Jones, *Bioinformatics* **2009**, *25*, 413–414.

[259] J. I. Garzon, J. Kovacs, R. Abagyan, P. Chacon, *Bioinformatics* **2007**, *23*, 901–902.

[260] A. N. Jain, *J. Med. Chem.* **2003**, *46*, 499–511.

[261] B. Dahlback, B. O. Villoutreix, *J. Thromb. Haemost.* **2003**, *1*, 1525–1534.

[262] B. Dahlback, B. O. Villoutreix, *FEBS Lett.* **2005**, *579*, 3310–3316.

[263] B. Dahlback, B. O. Villoutreix, *Arterioscler. Thromb. Vasc. Biol.* **2005**, *25*, 1311–1320.

[264] B. O. Villoutreix, O. Sperandio, *Curr. Opin. Struct. Biol.* **2010**, *20*, 168–179.

[265] J. P. Overington, B. Al-Lazikani, A. L. Hopkins, *Nat. Rev. Drug Discov.* **2006**, *5*, 993–996.

[266] A. L. Hopkins, C. R. Groom, *Nat. Rev. Drug Discov.* **2002**, *1*, 727–730.

[267] M. E. Bunnage, *Nat. Chem. Biol.* **2011**, *7*, 335–339.

[268] H. Li, H. Xiao, L. Lin, D. Jou, V. Kumari, J. Lin, C. Li, *J. Med. Chem.* **2014**, *57*, 632–641.

[269] S. E. Ozbabacan, H. B. Engin, A. Gursoy, O. Keskin, *Protein Eng. Des. Sel.* **2011**, *24*, 635–648.

[270] Z. Huang, L. Mou, Q. Shen, S. Lu, C. Li, X. Liu, G. Wang, S. Li, L. Geng, Y. Liu, J. Wu, G. Chen, J. Zhang, *Nucleic Acids Res.* **2014**, *42*, D510–6.

[271] V. Zoete, A. Grosdidier, O. Michielin, *J. Cell Mol. Med.* **2009**, *13*, 238–248.

[272] D. Joseph-McCarthy, J. C. Baber, E. Feyfant, D. C. Thompson, C. Humblet, *Curr. Opin. Drug Discov. Devel.* **2007**, *10*, 264–274.