

Hot-spot analysis for drug discovery targeting protein-protein interactions

Mireia Rosell¹, Juan Fernández-Recio^{1,2}

¹ Barcelona Supercomputing Center (BSC), Spain

² Institut de Biologia Molecular de Barcelona (IBMB), CSIC, Spain

ABSTRACT

Introduction

Protein-protein interactions are important for biological processes and pathological situations, and are attractive targets for drug discovery. However, rational drug design targeting protein-protein interactions is still highly challenging. Hot-spot residues are seen as the best option to target such interactions, but their identification requires detailed structural and energetic characterization, which is only available for a tiny fraction of protein interactions.

Areas covered

In this review, the authors cover a variety of computational methods that have been reported for the energetic analysis of protein-protein interfaces in search of hot-spots, and the structural modeling of protein-protein complexes by docking. This can help to rationalize the discovery of small-molecule inhibitors of protein-protein interfaces of therapeutic interest. Computational analysis and docking can help to locate the interface, molecular dynamics can be used to find suitable cavities, and hot-spot predictions can focus the search for inhibitors of protein-protein interactions.

Expert opinion

A major difficulty for applying rational drug design methods to protein-protein interactions is that in the majority of cases the complex structure is not available. Fortunately, computational docking can complement experimental data. An interesting aspect to explore in the future is the integration of these strategies for targeting PPIs with large-scale mutational analysis.

ARTICLE HIGHLIGHTS

- Protein-protein interactions are critical for essential biological processes and pathological situations

- The effect of pathological mutations on protein interaction networks and pathways are difficult to foresee
- Structural and energetic data on protein-protein interactions is essential to understand biomolecular processes, but their availability is currently very limited
- A variety of computational methods have been reported for protein-protein interface characterization and complex structure modeling by docking
- Computational approaches can help to improve rational drug discovery targeting protein-protein interactions
- Integration of computational modeling, drug discovery targeting protein interactions, and large-scale mutational analyses can help to progress towards precision medicine

1. Introduction

Rational drug design has been the object of increased attention by the pharma industry, mostly propelled by the accumulated expertise on biological systems, and the large-scale availability of sequence, structural and functional data, together with better computer-based models. One of the most successful examples was the series of HIV protease inhibitors that were computationally designed and developed in the early 1990s, which provided the first effective clinical treatments against AIDS [1]. However, despite all the investments and the increased availability of computational resources and big data, the number of new molecular entities approved by FDA per year is reaching a plateau [2]. The reasons for this are multiple, but one of the key factors is that the majority of available small-molecule drugs target only three major protein families: GPCRs, ion channels and nuclear receptors [3]. In order to expand the target space, it is important to realize that proteins do not act alone, but they rather interact with other proteins and biomolecules, forming intricate networks of interactions, which determine the behavior of the system. Indeed, the drug discovery field is beginning to focus on new concepts such as systems pharmacology [4,5] and network medicine [6]. In this review, we will discuss the important role of protein-protein interactions in the cell, and how the characterization and identification of interface hot-spot residues could help to discover small-molecule compounds that can modulate protein-protein interactions of pathological interest and have an effect on the system at phenotypic level.

2. Importance of protein-protein interactions for drug discovery

2.1. *Protein-protein interactions in biological processes*

Protein-protein interactions (PPIs) play essential roles in biological systems [7]. They can form more or less stable multi-molecular complexes involved in a variety of functions, such as the ribosome, nuclear pore, spliceosome, etc., or they can form specific and dynamic interaction networks as in the case of cell signaling or metabolic pathways [8]. The complexity of these PPI networks is highly related to the dynamics of the cell. The dynamic activity in a cell is distributed as functional modules, defined as groups of different proteins that interact but that are not necessarily present at the same time in the same space [9,10]. These functional modules can be detected by clustering the groups of nodes in the network according to their connectivity.

The complete set of interactions in a living organism defines the so-called interactome [11], whose description is key to understand the behavior of the entire biological system, especially when integrated with other data generated from the "omics" sciences (genomics,

proteomics, transcriptomics, etc.). Recent studies have tried to describe the complete interactome of different organisms, but this is a highly challenging task due to experimental limitations. Indeed, the number of PPIs in human is highly uncertain, with some estimations widely ranging between 130,000 and 650,000 [12,13]. In a recent study, a total of 14,000 human interactions were experimentally identified with high reproducibility [11]. Using information from several public databases, the number of reliable interactions (i.e. those found in more than one database and showing reliable evidence that they are binary) can increase up to 93,000 PPIs (<http://interactome3d.irbbarcelona.org/>)[14]. This shows that there is still a significant degree of uncertainty in the description of the human protein-protein interactome.

2.2. Protein-protein interactions and disease

Protein-protein interactions are involved in the majority of diseases, and thus are key to understand, prevent and correct pathological situations. In many cases, the perturbation of a given protein-protein interaction due to environmental, genetic or other reasons is the cause of the disease. In other cases, perturbed protein-protein interactions might not be the primary cause of a pathological situation, but they could affect the behavior of a given network or metabolic pathway and thus explain the phenotype. Functional genomics studies have helped to understand the involvement of disease-associated mutations in protein-protein interactions. For instance, some pathological mutations can disrupt protein interactions by causing misfolding of their components [15,16], or they can directly affect protein-protein interaction networks involved in specific cellular phenotypes [17].

A wealth of data is available on mutations leading to pathological phenotypes, most of which are publicly accessible in different databases: ClinVar (www.ncbi.nlm.nih.gov/clinvar/), gnomAD (Genome Aggregation Database; www.hgmd.cf.ac.uk/ac/index.php), Humsavar (<http://www.uniprot.org/docs/humsavar>), OMIM (Online Mendelian Inheritance in Man; www.omim.org/) and HGMD (Human Gene Mutation Database; www.hgmd.cf.ac.uk/ac/index.php). The information about the genes, disease disorders, and proteins altered by known pathological mutations can be combined to build human disease networks. These can be defined as graphs in which a disease node is linked with a gene node containing a mutation associated to the disease, which in turn is connected to another gene based on the interactions of the encoded proteins [18].

The ultimate objective is to build the complete set of disease networks occurring in human, or diseasome [19]. Computational tools can be used to analyze the properties of these

disease networks (hubs, connectivity, topology, etc.) [20] and help to highlight the key players that drive most of the characterized diseases [21]. This holds promises for helping to find treatments to revert altered disease networks back to the normal state.

The effect of pathological mutations on the entire disease network can differ depending on the aminoacid change and its location in the protein. A strong change in protein folding or stability will likely affect to all the interactions of the mutated protein (Figure 1). However, a mutation located at a specific protein-binding interface could affect only to specific interactions or “edges” of a given network (Figure 1). This edgetic effect can vary depending on the type of mutation and its location, producing unforeseen effects on the entire network [6,22,23]. Indeed, interaction perturbation profiling of missense mutations across a broad spectrum of human disorders suggests that around one third of disease mutations have edgetic effects, modifying the range of effect of specific interactions of a given protein [22]. Interestingly, mutated proteins with edgetic effects have been found to play central roles in the protein network [22]. Due to such edgetic effects, dissimilar mutations within the same gene may produce distinct interaction profiles and, as a consequence, different disease phenotypes [22].

Thus, characterizing such subtle networks effects for a given observed disorder is of paramount importance to understand its cause and to rationally develop new strategies for therapeutic intervention, within the context of precision medicine.

[FIGURE 1 HERE]

2.3. Structural and energetic characterization of protein-protein interactions

Current interaction network maps provide useful information on the relationships of genes or interactions between proteins, in the context of normal and pathological situations. However, the vast majority of network analyses are done at a level of resolution that makes it difficult to observe details of the interactions at atomic level. As mentioned above, in order to understand and model subtle but key aspects of the disease networks, such as the edgetic effects of pathological mutations and genetic variants in general, it would be necessary to include data about the three-dimensional (3D) structures of the protein-protein complexes forming the network [24]. The integration of such structural data within the human disease network can help to understand the effect of disease-related mutations at molecular level [25]. In addition,

the characterization of the energetic effects of specific mutations in binding affinity can provide a first general explanation for some of the differences in phenotype.

Existing structural data on protein-protein complexes at atomic resolution, mostly from x-ray or NMR experiments, have provided a wealth of information on protein-protein association and the role of disease mutations. Thanks to the available structural information, the edgetic effect of thousands of cancer mutations at protein-protein interfaces has been clearly described [26]. There are useful databases mapping pathological variants on protein-protein complex structures, such as Structure-PPi [27] or dSysMap [28]. The effect of mutations at local level can be better understood by analyzing the structural information of protein-protein interfaces in more detail. It is known that protein-binding sites involve around 28 residues on average, forming mostly flat interfaces of around 1000 Å² on average [29], which are much larger than traditional protein-ligand sites [30]. Protein-protein interfaces are organized in three major regions: core, rim and support, which are defined according to the relative accessible surface area (rASA) of the interface residues (Figure 2). The core region is formed by exposed residues at the unbound protein that are significantly buried upon binding. They contribute with a significant number of inter-protein contacts during interaction and tend to be more hydrophobic. The rim residues are exposed residues at protein surface that remain partially exposed upon binding. They provide a suitable micro-environment for the establishment of the interaction. These residues are more mobile than core residues and help to shield the core from the solvent. The support region is formed by partially exposed residues in the unbound protein that become buried when the complex is formed, and thus can help to establish the interaction [29]. All these structural considerations are important when analyzing the role of disease-related mutations. Recent studies have shown that disease-causing mutations are enriched at the core region of the interface [23,31]. When analyzing all the structural aspects in protein-protein interactions, we should not forget that proteins are not static entities. In addition to their intrinsic conformational variability in solution, they often show different conformational states (as well as isoforms generated by alternative splicing, or post-translational modifications, such as methylation, phosphorylation, glycosylation, etc). Therefore, the definition of whether a given protein is active or inactive, or is engaged in a protein interaction or not, will depend on which of their different possible molecular states are considered. Along this line, a recent work has redefined the concept of protein-protein interaction networks, by including conformational variability of the interacting proteins [32]. Obviously, the effect of a given pathological mutation on conformational variability, post-translational events or splicing will be important in order to fully understand its role on the protein-protein interaction networks.

[FIGURE 2 HERE]

For a given mutation located at a protein-protein interface, it is also important to know the effect of such mutation on binding affinity. We mentioned above that a residue located at the core region of the protein-protein interface is expected to affect the interaction when mutated. Consequently, many disease-related mutations are directly involved at protein-protein interfaces [23,33,34]. However, the degree of this effect would depend on the type of mutation. Indeed, it has been found that missense mutations described as disease-related in the database OMIM can cause changes in protein-protein binding energy [35]. Thus, the energetic characterization of protein-protein interactions is of paramount importance in order to understand in detail the effects of disease mutations in the disease networks. Obtaining this type of data requires building point mutations and analyzing them experimentally, which is too costly to be done in a high-throughput manner. The SKEMPI database contains available experimental data of changes in binding affinity for over 3,000 mutations [36]. However, this data set includes only a few of the increasingly high number of known pathological mutations. Despite the limited experimental data, it has been already shown that disease causing mutations located at protein-protein interfaces tend to debilitate protein binding, by introducing clashes, removing key interactions, or simply by increasing the flexibility of the interacting proteins. On the other hand, some disease mutations may increase protein rigidity and protein-protein complex stability [37,38].

Related to the energetic characterization of protein-protein interactions, it is important to discuss the concept of "hot-spot" residues. Based on experimental mutation to alanine, it has been found that usually only a few residues, so-called "hot-spots", are responsible for the majority of the binding affinity (Figure 3A). They are typically defined as those ones contributing more than 2 kcal.mol⁻¹ to the total binding energy of the complex [39,40]. Experimental measurements based on specific mutants are costly, so the available data on hot-spots is limited [40,41]. Mutation of these hot-spot residues are more likely to be related to disease. [42,43]. Recent studies show that hot-spot residues can also be relevant in homo-oligomeric interfaces [16]. Hot-spot residues at self-assembly patches in fibers and punctate foci play an essential role in binding, and their mutation could induce the formation of non-natural assemblies and cause disease [15,16,44–46].

[FIGURE 3 HERE]

2.4 . Computational analysis and modeling of protein-protein complexes

A large variety of statistical potentials, empirical functions and energy-based algorithms can be applied to computationally characterize a protein-protein complex structure. Many of these functions are compiled and publicly available in the database CCharPPI [47]. They can be used to analyze surface patches, estimate binding affinities, or identify hot-spot residues [53] [54]. Focusing on the identification of protein-protein hot-spot residues, many computational approaches first model the mutated residues based on the complex structure and then compare the computed binding affinity of the modelled mutant with that of the original complex structure. Some examples of this strategy are HSPred [50], Rosetta [51] and FoldX [52]. Machine learning is also used in different hot-spot prediction approaches, such as PCRPI, which combines sequence conservation, energy score and contact number information [53], PPI-Pred, based on surface shape and electrostatics [54], or Pocket-Query, which provides an assortment of metrics useful for predicting hot-spots [55]. Other methods are: HotSpot Wizard [56], based on the integration of structural, functional and evolutionary information provided by several databases; DrugScorePPI [57], derived from experimental alanine scanning results; iPred [58], using pairwise potential atom types and residue properties; or ECMIS [59], using a new algorithm combining energetic, evolutionary and structural features. These methods can be used to analyze interfaces at large-scale. In this context, it is interesting to mention the PCRPI database (PCRPI-DB), which contains computationally annotated hot-spot residues in all protein-protein complexes for which a high-resolution 3D structure is known [60].

The above hot-spot prediction methods are based on the 3D structure of a protein-protein complex. However, despite all the methodological advances and the number of structural genomics projects initiated in recent years, the structural determination of all protein-protein complexes in human (structural interactome) remains one of the biggest challenges in structural biology. Indeed, structural information is available for only a tiny fraction of all the protein-protein interactions that are estimated to occur in human [14]. In this context, a variety of computational approaches have been reported to help characterizing a protein-protein interaction for which there is no structural information.

Many of these computational methods aim to predict the protein-binding regions at the surface of the unbound protein structures. One popular tool is ProMate, which uses a combination of biophysical properties [61]. Other interface prediction methods are: PINUP [62], which is also based on a combination of empirical energy functions, or cons-PPISP [63], a neural network predictor which uses sequence profiles and solvent accessibilities of each residue and its spatial neighbors. It is important to remark that all these methods are applied to the structure of one of the unbound proteins and therefore can only predict general interface residues, which are not necessarily specific for one particular partner protein. In the case of a protein interacting with other different proteins using different interfaces, such predicted interface residues are not useful to distinguish these specific interfaces.

On the other side, different computational docking methods have been reported for predicting more specifically the structure of a complex formed by two interacting proteins. Methodologically, two major aspects are considered in virtually all docking protocols: a sampling procedure, to search for different binding orientations, and a scoring procedure, which evaluate these decoys in order to identify the correct binding mode. Major sampling strategies are based on shape matching, exhaustive global search, and stochastic sampling. Shape matching methods include DOCK [64] or Patchdock [65], where binding regions are represented by a graph generated from geometric features, such as convex, flat and concave regions. Examples of exhaustive global search methods are FTDock [66], ZDOCK [67] or MolFit [68] based on Fast Fourier Transform (FFT) calculations [69] on rigid-body proteins. Methods using stochastic search are ICM-DISCO [70], HADDOCK [71], RosettaDock [72], SwarmDock [73], or LightDock [74]. After generating many different docking decoys by the above described methods, it is important to evaluate all these docking poses and identify the correct binding modes. In this context, scoring plays an essential role, and different strategies have been reported, based on statistical potentials, empirical functions or energy-based description, the latter usually including van der Waals, desolvation or electrostatics terms. Several scoring algorithms that can be independently applied to previously generated docking models are available, such as pyDock [75], ZRANK [76], or SIPPER [77]. Many methods, in addition to scoring, introduce flexible refinement of the docking models, such as FireDock [78]. Available computational docking methods are being continuously evaluated at the Critical Assessment of PRedicted Interactions (CAPRI) community experiment [79]. CAPRI is an international scientific effort to boost the development of different approaches to solve the problem of protein-protein docking. After more than fifteen years since its first edition, the CAPRI experiment is now the source of standard protein-protein docking sets and quality measurements.

The use of energy-based functions in docking can help to describe the energetic aspects of protein-protein association [36]. One interesting application of this is the capability of predicting hot-spot residues by analyzing the regions that are energetically important in the docking simulations by pyDock [80] (Figure 3B). The main advantages of this method are that it does not require the structure of a complex and that hot-spot predictions are specific for two given interacting proteins (as opposed to general interface prediction on one unbound protein as described above).

2.5. Modulation of protein-protein interactions for drug discovery

Traditional drug discovery has focused on targeting individual proteins. However, the number of proteins that are druggable, i.e. potentially used for rational drug discovery, is limited. The majority of FDA-approved drugs target a few families of proteins, basically GPCRs, ion channels and nuclear receptors [3]. The traditional goal of targeting a single protein has strong limitations, such as unforeseen side-effects (because the protein altered by the drug may be involved in other different functions), or a limited efficacy at system level (because of homeostatic compensation through other pathways). For these reasons, the pharmaceutical industry has a strong need for expanding current target space, considering the proteins involved in a given disease (traditional targets) in the context of an interaction network [81]. Detailed understanding of the behavior of a protein interaction network in pathological conditions is especially critical for complex diseases such as cancer or rare conditions related to the malfunction of the Ras signaling pathway (RASopathies) [82]. In addition to understanding the systemic effect of targeting one protein, a detailed description of a pathological network can help to identify specific protein-protein interactions as potential targets to be analyzed and modulated according to the interaction profiling.

Modulating protein-protein interactions of therapeutic interest with small-molecules is a long-sought goal in drug discovery. Several examples of antibodies and peptides capable of inhibiting PPIs have been reported. Monoclonal antibodies can target the protein-binding surface of one of the proteins, and mimetic peptides can replace the interaction surface of one of the proteins. But the low bioavailability of antibodies and peptides does not make them very attractive for therapeutic purposes. Small-molecules show many more advantages over large protein competitors, especially from a kinetic perspective [30]. Success in modulating PPIs with small-molecules for therapeutic purposes clearly depends on the target type [83], and this is why the characterization of protein-protein interfaces is so important. In the last decades,

increasing knowledge on small-molecules capable of inhibiting protein-protein interactions is available in hand-curated databases such as TIMBAL [84], 2P2I [85] and iPPI-DB [86,87]. TIMBAL database holds a wide diversity of PPI inhibitors, including peptides, useful for helping to characterize which type of molecules could be involved in protein interfaces [84]. The small-molecule compounds annotated in TIMBAL tend to be large and lipophilic, engaged in hydrophobic contacts, and involved in fewer hydrogen bonds [84]. 2P2I is a hand-curated database combined with an automated extractor from ChEMBL, and holds all structural information involving small-molecule inhibitors of protein-protein complexes. Interestingly, the protein-protein interfaces of the complexes in 2P2I database are smaller, more hydrophobic, with less charged residues and more non-polar atoms than those of standard hetero-dimeric complexes [85]. Finally, iPPI-DB database aims to compile the diversity of profiles of inhibitors of PPI in order to enable a rational characterization of the PPI inhibitor chemical space [86,87]. This database includes physicochemical and pharmacological data in addition to the profile of the PPI target.

Thus, the identification of small-molecule inhibitors of PPIs is not an easy task. Known PPI inhibitors do not have high similarity to traditional inhibitors of enzymes and receptors [30], showing physicochemical properties that may violate traditional rules such as the Lipinski's Rule of Five [88]. The molecular size of known small-molecule inhibitors of PPIs is around 500-900 Da, with K_i values of less than $1\mu\text{M}$. In many cases (e.g. IL-2, HDM2, HPV E2), this value is in the mid-nanomolar to low-nanomolar range, comparable to the binding affinity of the protein-protein complex [30]. There are currently some small-molecule inhibitors of PPIs in clinical trials, and a few of them have been approved by the Food and Drug Administration (FDA). Table 1 shows small-molecule drugs that are targeting PPIs [89,90]. Considering the difficulties for a potential drug to reach the clinical trials stage, a lot of attention is focused onto drugs that have already passed clinical trials or have been accepted for medical use. The case of Gabapentin is an interesting example of drug repositioning. This drug was originally designed to mimic the chemical structure of neurotransmitter GABA, and was used as a treatment for epilepsy. But later it was found that this drug significantly reduced PKC ϵ translocation by competitively inhibiting the interaction with the pronociceptive peptides bradykinin and prokineticin 2, and it is now widely used to relieve neuropathic pain in patients with amyotrophic lateral sclerosis (ALS) [91].

[TABLE 1 HERE]

3. Rational discovery of new modulators of protein-protein interactions

3.1. Major challenges for rational drug design targeting protein-protein interactions

Targeting protein-protein interactions with small-molecules is highly challenging. Protein-protein interfaces are usually larger than traditional protein-ligand interactions, and they are mostly flat, showing a large variety of topologies [89,92]. For this reason, known inhibitors of protein-protein interactions do not show high similarity to traditional inhibitors of enzymes and receptors [30], and usually exhibit physicochemical properties that may violate traditional rules such as the Lipinski's Rule of Five [88].

Perhaps the first difficulty for rational design is to find a correct spot to be targeted by small-molecule. Contrary to traditional targeting of enzymes and membrane receptors, in which the known active site is used as starting point for drug discovery, identifying suitable cavities in protein-protein interfaces is extremely difficult since natural protein-protein complexes have not evolved to host cavities for small ligands. Fortunately, protein-protein contact surfaces are not rigid in solution, and conformational motions at local level (basically side-chains and loops) can induce transient openings in the protein surface [30,93]. These small perturbations have been detected by molecular dynamics simulations, making them a suitable target surface for binding of small-molecules [30].

Provided that transient cavities can be generated by molecular dynamics, the question is, how can we identify those potentially useful cavities in the different conformations generated during the dynamics? There are many freely-available computational tools for the identification of cavities in the protein surface, like Fpocket [94], PASS [95] or QsiteFinder [96]. Most of these tools were developed to locate surface pockets involving active sites or natural substrate binding

sites. Whether they can be also efficient for the identification of suitable pockets for PPI inhibitors within protein-protein interfaces is yet to be evaluated.

Once transient pockets can be generated by molecular dynamics simulations, and identified by cavity predictor methods, the next problem is how to select the most suitable transient cavity for small-molecule positioning. The starting position for further docking or virtual ligand screening studies is critical, since the small-molecule inhibitor needs to bind at the optimal site in order to efficiently compete with a large protein. In addition, many small-molecule inhibitors of protein-protein interactions are still larger than traditional enzyme inhibitors, so in order to identify a suitable cavity for these cases, more than one predicted pocket might need to be merged [97] or combined for a fragment-based drug design strategy [98]. This would involve different computational problems, such as efficiently clustering the many detected pockets, or selecting a suitable threshold size for the pocket (small pockets might not cover the entire inhibitor, while large pockets would not be adequate for a small-molecule inhibitor) (Figure 4).

[FIGURE 4 HERE]

3.2. Hot-spots can focus the search for inhibitors of protein-protein interactions

Protein-protein interfaces that have already been successfully targeted by small-molecule inhibitors tend to have extended binding grooves that may be split into different sub-pockets [99,100] (Figure 4). Very often, hot-spot residues are located nearby these pockets and are complementary on both sides of the interface. In addition, small conformational changes in the binding site makes a pocket deeper when bound to a small-molecule than to the partner protein [101]. Finally, PPIs with known small-molecule inhibitors tend to have small, high-affinity interfaces and include a hot segment that is essential for the binding to the partner protein [100]. In addition, druggable hot-spots at protein-protein interfaces with a general tendency to bind organic compounds have been reported. These druggable hot-spots show significant conformational flexibility to facilitate the accommodation of a ligand of drug-like dimensions [102].

Thus, targeting hot-spot residues at protein-protein interfaces with small-molecules seems to be a reasonable strategy to disrupt protein-protein interactions. By targeting hot-spots,

a small-molecule could break critical interactions and be able to compete with a large protein-protein interface. This has been extensively applied to the discovery of PPI inhibitors, in some cases yielding potent and selective compounds [42,103–105]. The knowledge of interface and hot-spot residues based on the complex structures of several inhibitors can be used to localize the binding pocket [106]. In one interesting example, predicted hot-spots are used for the rational design of small-molecule compounds capable of blocking the IFN- α / receptor interaction. Hot-spots were predicted with iPred, then pharmacophore search was performed with VirtualLigand, and final docking of candidate compounds was done with Gold [107]. Computational docking can help to model protein-protein complexes of therapeutic interest and predict hot-spot residues, which can be helpful for developing drug discovery programs aiming to target protein-protein interactions with no available structure (Figure 5).

The use of hot-spots can be complemented by other computational approaches based on the use of peptides and protein fragments [104,108], fragment docking and coevolutionary analysis [109], or ligand docking to locate protein-protein interfaces and potentially druggable sites [110].

[FIGURE 5 HERE]

3.3. Stabilizing protein-protein interactions

The stabilization of protein-protein interactions by small-molecule compounds is an alternative strategy to modulate interactions of pathological interest. Although it has received much less attention in drug discovery, some of the FDA-approved drugs modulating PPIs are actually stabilizers (Table1). Small molecules known to stabilize protein-protein interactions need to bind simultaneously both protein partners, and therefore they are usually involved in druggable pockets from the two interacting proteins [111]. Therefore, these molecules are highly selective against a specific protein-protein complex, while PPI inhibitors usually bind to only one of the protein partners.

Besides the above mentioned difficulties in designing PPI inhibitors, the identification of PPI stabilizers encounters additional difficulties, such as the lack of understanding of the mechanisms and underlying principles of PPI stabilization, and the poor representation of PPI stabilizer chemical space in current small-molecule libraries [111]. The large variation of molecular structures found for PPI stabilizers (from large macrocycles to small hydrophobic compounds) makes it difficult to establish consensus chemical principles for these compounds

[111]. Indeed, most of the reported cases of PPI stabilization have been serendipitously found, and only a few of them have been rationally designed [111]. Although interface hot-spot residues could be also useful to locate potential pockets for PPI stabilizers, as in the case of PPI inhibitors, the identification of druggable pockets that could be optimal for the stabilization of protein-protein interactions needs still further efforts from the drug discovery community.

4. Conclusions

Protein-protein interactions are essential in life processes, and their perturbation, very often by missense mutations, can lead to pathological situations. As a consequence, they are increasingly seen as attractive targets for drug discovery. However, rational design of small-molecule compounds targeting protein-protein interfaces presents important challenges. The strategy of choice is usually to target interface hot-spot residues, but their experimental identification needs exhaustive alanine-scanning and biophysical characterization. We have reviewed here different computational methods that are available to help in the identification of such hot-spot residues. When the complex structure is not available, additional computational interface analysis and docking approaches are needed. Finally, transient cavities for ligand binding can be modelled by molecular dynamics. All these computational methods can complement experimental efforts and help to rationalize the discovery of small compounds capable of modulating protein-protein interactions of therapeutic interest.

5. Expert opinion

A major problem is that all of the discussed strategies to identify binding cavities in protein-protein interfaces need the structure of the complex, which is not available in the majority of complexes. Thus, identifying binding sites in the unbound forms of the interacting proteins is a desirable goal, but much more challenging than on the holo structures, because in many cases these pockets may show large conformational rearrangement and thus remain mostly hidden in the apo state (so-called cryptic sites). Recent studies have found that these cryptic sites tend to be conserved in evolution and could be identified by using machine learning methods [112]. The combination of molecular dynamics and docking predictions could also help to identify these sites. A sensible strategy would be to apply molecular dynamics to the unbound form of the protein, identify all transient pockets generated in the different conformers with methods like

MDpocket [113] or Caver[114], then apply protein docking and hot-spot predictions [80] to locate the interface and the important residues to target, and finally select the most suitable transient pockets to start virtual screening and ligand docking procedures. One example is the interaction between TGF- β Type I receptor with FKBP12, for which several disease-related mutations are annotated, and several functional inhibitors have been described [115]. Interestingly, one of the "unclassified" mutations is involving a binding hot-spot residue predicted by docking, which could be used to redefine their pathogenic character (Figure 6). Further molecular dynamics could help to find transient pockets around this hot-spot residue and be the basis of future drug development programs.

[FIGURE 6 HERE]

An interesting aspect to explore in the future is the integration of these strategies for targeting PPIs with large-scale mutational analysis. Characterization of the effect of a pathological mutation on the network of protein-protein interactions in human could help to identify potential targets for therapeutic intervention. Candidate targets for that would be protein-protein interactions whose binding energy are directly affected by the mutation. The majority of mutations affecting protein-protein interfaces are expected to be destabilizing, but a significant percentage of them (20% according to SKEMPI database) makes the interaction stronger. In other situations, over-expressed proteins by a mutation in a regulatory gene or over-activated proteins by a mutation affecting its structure or dynamics could be involved in protein-protein interactions that would need to be blocked for therapeutic purposes. For instance, over-expression of interleukin-8 (IL8) has been associated with KRAS mutations in tumors, and thus the interaction of IL8 with their receptors CXCR1 or CXCR2 has been proposed as an attractive therapeutic target in cancer [116]. It could also happen that a protein-protein interaction acquires a new unwanted role due to homeostatic compensation of the network upon a given mutation or other pathological situations. As an example, IL-2 levels are raised in normal immune response, but unwanted high IL-2 levels are also found in pathological situations as in autoimmune disease or graft rejection in organ transplantation. In these situations, the otherwise normal interaction of IL-2 with IL-2R produces unwanted effects, and thus constitutes a known drug target for which several inhibitors have been identified (see Table 1) [117]. In all these situations, the modulation of a target protein-protein interaction by a small-molecule could restore normal function, and thus such compound could have the potential to be further

developed into a therapeutic drug. Large-scale annotation of disease-related mutations thanks to lower sequencing costs could bring a wealth of protein-protein interactions that could be potentially used as targets for drug discovery. Then, it would be essential to integrate structural characterization, dynamics and hot-spot analysis, before attempting to rational design small-molecule inhibitors that could target such interactions.

[ACKNOWLEDGEMENTS HERE]

REFERENCES

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Gershell LJ, Atkins JH. A brief history of novel drug discovery technologies. *Nat. Rev. Drug Discov.* 2003;2:321–327.
2. Morrison C. Fresh from the biotech pipeline—2016. *Nat. Biotechnol.* 2017;35:108–112.
3. Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* 2016;16:19–34.
4. Boezio B, Audouze K, Ducrot P, et al. Network-based Approaches in Pharmacology. *Mol. Inform.* 2017;36:1–10.
5. Taboureau O, Baell JB, Fernández-Recio J, et al. Established and emerging trends in computational drug discovery in the structural genomics era. *Chem. Biol.* 2012;19:29–41.
6. Zanzoni A, Soler-López M, Aloy P. A network medicine approach to human disease. *FEBS Lett.* 2009. p. 1759–1765.
7. Levy ED, Boeri Erba E, Robinson CV, et al. Assembly reflects evolution of protein complexes. *Nature.* 2008;453:1262–1265.
8. Wu X, Hasan MA, Chen JY. Pathway and network analysis in proteomics. *J. Theor. Biol.* 2014;362:44–52.
9. Hartwell LH, Hopfield JJ, Leibler S, et al. From molecular to modular cell biology. *Nature.* 1999;402:C47–C52.
10. Pizzuti C, Rombo SE. Algorithms and tools for protein-protein interaction networks clustering, with a special focus on population-based stochastic methods. *Bioinformatics.* 2014;30:1343–1352.
11. Rolland T, Taşan M, Charlotheaux B, et al. A Proteome-Scale Map of the Human Interactome Network. *Cell.* 2014;159:1212–1226.
12. Venkatesan K, Rual JF, Vazquez A, et al. An empirical framework for binary interactome mapping. *Nat. Methods.* 2009;6:83–90.
13. Rual JF, Venkatesan K, Hao T, et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature.* 2005;437:1173–1178.
14. Mosca R, Céol A, Aloy P. Interactome3D: adding structural details to protein networks. *Nat. Methods.* 2012;10:47–53.

** A useful compilation of structural data for the entire interactomes of several organisms, including human
15. Dobson CM. Protein folding and misfolding. *Nature.* 2003;426:884–890.
16. Garcia-Seisdedos H, Empereur-Mot C, Elad N, et al. Proteins evolve on the edge of supramolecular self-assembly. *Nature.* 2017;548:244–247.
17. AlQuraishi M, Koytiger G, Jenney A, et al. A multiscale statistical mechanical framework

- integrates biophysical and genomic data to assemble cancer networks. *Nat. Genet.* 2014;46:1363–1371.
18. Goh K-I, Cusick ME, Valle D, et al. The human disease network. *Proc. Natl. Acad. Sci.* 2007;104:8685–8690.
 19. Goh K-I, Choi I-G. Exploring the human diseasome: The human disease network. *Brief. Funct. Genomics.* 2012;11:533–542.
 20. Holland DO, Shapiro BH, Xue P, et al. Protein-protein binding selectivity and network topology constrain global and local properties of interface binding networks. *Sci. Rep.* 2017;7:5631.
 21. Janjić V, Pržulj N. The Core Diseasome. *Mol. Biosyst.* 2012;8:2614–2625.
 22. Sahni N, Yi S, Taipale M, et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell.* 2015;161:647–660.
- ** Experimental functional profiling of several thousand disease-associated missense mutations in the context of protein interaction networks. Different PPI profiles (edgotypes) were found: quasi-wild-type (unperturbed interactions), quasi-null (folding/stability effects perturbing all interactions), edgetic perturbations.
23. David A, Sternberg MJE. The Contribution of Missense Mutations in Core and Rim Residues of Protein-Protein Interfaces to Human Disease. *J. Mol. Biol.* 2015;427:2886–2898.
- * Large-scale computational analysis on the distribution of known pathological mutations in the different regions of protein-protein interfaces in a large set of protein-protein complex structures
24. Kiel C, Beltrao P, Serrano L. Analyzing protein interaction networks using structural information. *Annu. Rev. Biochem.* 2008;77:415–441.
 25. Fraser JS, Gross JD, Krogan NJ. From Systems to Structure: Bridging Networks and Mechanism. *Mol. Cell.* 2013;49:222–231.
 26. Engin HB, Kreisberg JF, Carter H. Structure-Based analysis reveals cancer missense mutations target protein interaction interfaces. *PLoS One.* 2016;11: e0152929.
 27. Vázquez M, Valencia A, Pons T. Structure-PPI: A module for the annotation of cancer-related single-nucleotide variants at protein-protein interfaces. *Bioinformatics.* 2015;31:2397–2399.
 28. Mosca R, Tenorio-Laranga J, Olivella R, et al. dSysMap: exploring the edgetic role of disease mutations. *Nat. Methods.* 2015;12:167–168.
 29. Levy ED. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. *J. Mol. Biol.* 2010;403:660–670.
 30. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature.* 2007;450:1001–1009.
 31. Barradas-Bautista D, Fernández-Recio J. Docking-based modeling of protein-protein interfaces for extensive structural and functional characterization of missense mutations. *PLoS One.* 2017;12:e0183643.

* A systematic study on the use of computational docking to characterize the involvement of pathological mutations at protein-protein interfaces and its application to predict edgetic effects

32. Halakou F, Kilic ES, Cukuroglu E, et al. Enriching Traditional Protein-protein Interaction Networks with Alternative Conformations of Proteins. *Sci. Rep.* 2017;7:7180.
 33. David A, Razali R, Wass MN, et al. Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. *Hum. Mutat.* 2012;33:359–363.
 34. Wang X, Wei X, Thijssen B, et al. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nat. Biotechnol.* 2012;30:159–164.
 35. Teng S, Madej T, Panchenko A, et al. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys. J.* 2009;96:2178–2188.
 36. Moal IH, Fernández-Recio J. SKEMPI: A Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. *Bioinformatics.* 2012;28:2600–2607.
- * A hand-curated database of binding free energy changes upon mutation extracted from literature for protein-protein heterodimeric complexes with experimentally determined structure
37. Kucukkal TG, Petukh M, Li L, et al. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr. Opin. Struct. Biol.* 2015;32:18–24.
 38. Yates CM, Sternberg MJE. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. *J. Mol. Biol.* 2013;425:3949–3963.
 39. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science.* 1995;267:383–386.
 40. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 1998;280:1–9.
 41. Fischer TB, Arunachalam KV, Bailey D, et al. The binding interference database (BID): A compilation of amino acid hot spots in protein interfaces. *Bioinformatics.* 2003;19:1453–1454.
 42. Cukuroglu E, Engin HB, GURSOY A, et al. Hot spots in protein-protein interfaces: Towards drug discovery. *Prog. Biophys. Mol. Biol.* 2014;116:165–173.
 43. Engin HB, Guney E, Keskin O, et al. Integrating structure to protein-protein interaction networks that drive metastasis to brain and lung in breast cancer. *PLoS One.* 2013;8:e81035.
 44. Li P, Banjade S, Cheng H-C, et al. Phase transitions in the assembly of multivalent signalling proteins. *Nature.* 2012;483:336–340.
 45. Noree C, Sato BK, Broyer RM, et al. Identification of novel filament-forming proteins in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. *J. Cell Biol.* 2010;190:541–551.
 46. Narayanaswamy R, Levy M, Tsechansky M, et al. Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *PNAS.* 2009;106:10147-

10152.

47. Moal IH, Jiménez-García B, Fernández-Recio J. CCharPPI web server: Computational characterization of protein-protein interactions from structure. *Bioinformatics*. 2015;31:123–125.
48. Morrow JK, Zhang S. Computational prediction of protein hot spot residues. *Curr. Pharm. Des.* 2012;18:1255–1265.
49. Fernández-Recio J. Prediction of protein binding sites and hot spots. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2011;1:680–698.
50. Lise S, Buchan D, Pontil M, et al. Predictions of hot spot residues at protein-protein interfaces using support vector machines. *PLoS One*. 2011;6:e16774.
51. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci.* 2002;99:14116–14121.
52. Schymkowitz J, Borg J, Stricher F, et al. The FoldX web server: An online force field. *Nucleic Acids Res.* 2005;33:W382–388.
53. Assi SA, Tanaka T, Rabbitts TH, et al. PCRPI: Presaging Critical Residues in Protein interfaces, a new computational tool to chart hot spots in protein interfaces. *Nucleic Acids Res.* 2010;38:e86.
54. Bradford JR, Westhead DR. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics*. 2005;21:1487–1494.
55. Koes DR, Camacho CJ. PocketQuery: Protein-protein interaction inhibitor starting points from protein-protein interaction structure. *Nucleic Acids Res.* 2012;40:387–392.
56. Pavelka A, Chovancova E, Damborsky J. HotSpot Wizard: A web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* 2009;37:376–383.
57. Krüger DM, Garzón JI, Montes PC, et al. Predicting protein-protein interactions with DrugScorePPI: Fully-flexible docking, scoring, and in silico alanine-scanning. *J. Cheminform.* 2011;3(Suppl 1):P36.
58. Geppert T, Hoy B, Wessler S, et al. Context-based identification of protein-protein interfaces and “hot-spot” residues. *Chem. Biol.* 2011;18:344–353.
59. Shingate P, Manoharan M, Sukhwai A, et al. ECMIS: computational approach for the identification of hotspots at protein-protein interfaces. *BMC Bioinformatics*. 2014;15:303.
60. Segura J, Fernandez-Fuentes N. PCRPI-DB: A database of computationally annotated hot spots in protein interfaces. *Nucleic Acids Res.* 2011;39:D755-60.
61. Neuvirth H, Raz R, Schreiber G. ProMate: A structure based prediction program to identify the location of protein-protein binding sites. *J. Mol. Biol.* 2004;338:181–199.
62. Liang S, Zhang C, Liu S, et al. Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res.* 2006;34:3698–3707.
63. Chen H, Zhou HX. Prediction of interface residues in protein-protein complexes by a consensus neural network method: Test against NMR data. *Proteins* 2005;61:21–35.

64. Kuntz ID, Blaney JM, Oatley SJ, et al. a Geometric Approach To Macromolecule-Ligand Interactions. *J. Mol. Biol.* 1982;161:269–288.
65. Schneidman-Duhovny D, Inbar Y, Nussinov R, et al. PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* 2005;33:W363–367.
66. Gabb HA, Jackson RM, Sternberg MJE. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* 1997;272:106–120.
67. Chen R, Li L, Weng Z. ZDOCK: An initial-stage protein-docking algorithm. *Proteins* 2003;52:80–87.
68. Redington PK. MOLFIT: A computer program for molecular superposition. *Comput. Chem.* 1992;16:217-222.
69. Katchalski-Katzir E, Shariv I, Eisenstein M, et al. Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci. U. S. A.* 1992;89:2195–2199.
70. Fernández-Recio J, Totrov M, Abagyan R. Identification of protein-protein interaction sites from docking energy landscapes. *J. Mol. Biol.* 2004;335:843–865.
71. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 2003;125:1731–1737.
72. Lyskov S, Gray JJ. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* 2008;36:W233–238.
73. Moal IH, Bates PA. SwarmDock and the use of normal modes in protein-protein Docking. *Int. J. Mol. Sci.* 2010;11:3623–3648.
74. Jiménez-García B, Roel-Touris J, Romero-Durana M, et al. LightDock: a new multi-scale approach to protein–protein docking. *Bioinformatics.* 2017 (in press).
75. Cheng TM, Blundell TL, Fernández-Recio J. pyDock: Electrostatics and Desolvation for Effective Scoring of Rigid-Body Protein–Protein Docking. *Proteins.* 2007;68:503–515.
76. Pierce, B. Weng Z. ZRANK: Reranking Protein Docking Predictions With an Optimized Energy Function. *Proteins.* 2007;67:1078–1086.
77. Pons C, Talavera D, de La Cruz X, et al. Scoring by intermolecular pairwise propensities of exposed residues (SIPPER): A new efficient potential for protein-protein docking. *J. Chem. Inf. Model.* 2011;51:370–377.
78. Andrusier N, Nussinov R, Wolfson HJ. FireDock: Fast interaction refinement in molecular docking. *Proteins* 2007;69:139–159.
79. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. *Proteins.* 2010;78:3073–3084.
80. Grosdidier S, Fernández-Recio J. Identification of hot-spot residues in protein-protein interactions by computational docking. *BMC Bioinformatics.* 2008;9:447.

* Application of protein-protein docking for predicting hot-spot residues in case of no complex structure

81. Vidal M, Cusick ME, Barabási A-L. Interactome networks and human disease. *Cell*. 2011;144:986–998.
82. Kiel C, Serrano L. Structure-energy-based predictions and network modelling of RASopathy and cancer missense mutations. *Mol. Syst. Biol.* 2014;10:727.
83. Makley LN, Gestwicki JE. Expanding the Number of “Druggable” Targets: Non-Enzymes and Protein-Protein Interactions. *Chem. Biol. Drug. Des.* 2014;81:22–32.
84. Higuero AP, Schreyer A, Bickerton GRJ, et al. Atomic interactions and profile of small molecules disrupting protein-protein interfaces: The TIMBAL database. *Chem. Biol. Drug Des.* 2009;74:457–467.
85. Basse MJ, Betzi S, Bourgeas R, et al. 2P2Idb: a structural database dedicated to orthosteric modulation of protein–protein interactions. *Nucleic Acids Res.* 2013;41:D824-7.
86. Labbé CM, Laconde G, Kuenemann MA, et al. IPPI-DB: A manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions. *Drug Discov. Today*. 2013;18:958–968.
87. Labbé CM, Kuenemann MA, Zarzycka B, et al. IPPI-DB: An online database of modulators of protein-protein interactions. *Nucleic Acids Res.* 2016;44:D542–D547.
88. Zhang MQ, Wilkinson B. Drug discovery beyond the “rule-of-five.” *Curr. Opin. Biotechnol.* 2007;18:478–488.
89. Scott DE, Bayly AR, Abell C, et al. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. *Nat. Rev. Drug Discov.* 2016;15:533–550.
- ** Highly relevant review of small molecules targeting protein-protein interactions.
90. Shin WH, Christoffer CW, Kihara D. In silico structure-based approaches to discover protein-protein interaction-targeting drugs. *Methods*. 2017;131:22-32.
91. Vellani V, Giacomoni C. Gabapentin Inhibits Protein Kinase C Epsilon Translocation in Cultured Sensory Neurons with Additive Effects When Coapplied with Paracetamol (Acetaminophen). *Sci. World J.* 2017;2017:3595903.
92. Pérot S, Regad L, Reynès C, et al. Insights into an Original Pocket-Ligand Pair Classification: A Promising Tool for Ligand Profile Prediction. *PLoS One*. 2013;8:e63730.
93. Eyrisch S, Helms V. Transient pockets on protein surfaces involved in protein-protein interaction. *J. Med. Chem.* 2007;50:3457–3464.
94. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*. 2009;10:168.
95. Brady GP Jr, Stouten PFW. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided. Mol. Des.* 2000;14:383–401.
96. Laurie ATR, Jackson RM. Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*. 2005;21:1908–1916.
97. Degac J, Winter U, Helms V. Graph-Based Clustering of Predicted Ligand-Binding Pockets on Protein Surfaces. *J. Chem. Inf. Model.* 2015;55:1944–1952.

98. Li H, Liu A, Zhao Z, et al. Fragment-based drug design and drug repositioning using multiple ligand simultaneous docking (MLSD): Identifying celecoxib and template compounds as novel inhibitors of signal transducer and activator of transcription 3 (STAT3). *J. Med. Chem.* 2011;54:5592–5596.
99. Fuller JC, Burgoyne NJ, Jackson RM. Predicting druggable binding sites at the protein-protein interface. *Drug Discov. Today.* 2009;14:155–161.
100. Arkin MR, Tang Y, Wells JA. Small-molecule inhibitors of protein-protein interactions: progressing towards the reality. *Chem. Biol.* 2014;18:1102–1114.
101. Johnson DK, Karanicolas J. Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation than the Rest of the Protein Surface. *PLoS Comput. Biol.* 2013;9: e1002951.
102. Kozakov D, Hall DR, Chuang G-Y, et al. Structural conservation of druggable hot spots in protein-protein interfaces. *Proc. Natl. Acad. Sci.* 2011;108:13528–13533.
103. Jubb H, Blundell TL, Ascher DB. Flexibility and small pockets at protein-protein interfaces: New insights into druggability. *Prog. Biophys. Mol. Biol.* 2015;119:2–9.
104. London N, Raveh B, Schueler-Furman O. Druggable protein-protein interactions - from hot spots to hot segments. *Curr. Opin. Chem. Biol.* 2013;17:952–959.
105. Guo W, Wisniewski JA, Ji H. Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorganic Med. Chem. Lett.* 2014;24:2546–2554.
106. Metz A, Pflieger C, Kopitz H, et al. Hot spots and transient pockets: Predicting the determinants of small-molecule binding to a protein-protein interface. *J. Chem. Inf. Model.* 2012;52:120–133.
107. Geppert T, Bauer S, Hiss JA, et al. Immunosuppressive small molecule discovered by structure-based virtual screening for inhibitors of protein-protein interactions. *Angew. Chem. Int. Ed.* 2012;51:258–261.
108. Winter A, Higuero AP, Marsh M, et al. Biophysical and computational fragment-based approaches to targeting protein-protein interactions: applications in structure-guided drug discovery. *Q. Rev. Biophys.* 2012;4:1–44.
109. Bai F, Morcos F, Cheng RR, et al. Elucidating the druggable interface of protein-protein interactions using fragment docking and coevolutionary analysis. *Proc. Natl. Acad. Sci. U. S. A.* 2016;113:E8051–E8058.
110. Li H, Kasam V, Tautermann CS, et al. Computational method to identify druggable binding sites that target protein-protein interactions. *J. Chem. Inf. Model.* 2014;54:1391–1400.
111. Andrei SA, Sijbesma E, Hann M, et al. Stabilization of protein-protein interactions in drug discovery. *Expert Opin. Drug Discov.* 2017;12:925–940.
112. Cimermancic P, Weinkam P, Rettenmaier TJ, et al. CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J. Mol. Biol.* 2016;428:709–719.
113. Schmidtke P, Bidon-chanal A, Luque FJ, et al. MDpocket: Open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics.* 2011;27:3276–

3285.

114. Petrek M, Otyepka M, Banás P, et al. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*. 2006;7:316.
115. Stockwell BR, Schreiber SL. TGF- β -signaling with small molecule FKBP12 antagonists that bind myristoylated FKBP12-TGF- β type I receptor fusion proteins. *Chem. Biol.* 1998;5:385-395.
116. Sunaga N, Kaira K, Tomizawa Y, et al. Clinicopathological and prognostic significance of interleukin-8 expression and its relationship to KRAS mutation in lung adenocarcinoma. *Br. J. Cancer*. 2014;110:2047–2053.
117. Wilson CGM, Arkin MR. Small-Molecule Inhibitors of IL-2/IL-2R: Lessons Learned and Applied. *Curr Top Microbiol Immunol*. 2011;348:25–59.

* Interesting example of a protein-protein interaction modulated by small molecules

FIGURE LEGENDS

Figure 1. Effects of pathological mutations on protein-protein interaction networks.

The network of protein interactions for ETFB according to Interactome3D (<https://interactome3d.irbbarcelona.org/>) is represented. The proteins are represented in nodes (ETFB in square, interacting partners in circles) and the interactions by edges (blue if there is a complex structure available; grey otherwise). The effect of mutations in ETFB can be disruptive (removal of all interactions), neutral (no interactions affected), or edgetic (only specific interactions are affected, with effects ranging from complete removal, weakening or strengthen them). All these no neutral effects can cause a modification or even rewiring on the entire network.

Figure 2. Type of interface residues.

The protein-protein complex structure between Xiap-BIR3 (surface) and Caspase (ribbon) (PDB 1NW9) is shown. Protein-protein interface of Xiap-BIR3 is represented with the three classes of interface regions: core (yellow), rim (blue) and support (green). Known hot-spot residues for this interaction are also shown (red sticks).

Figure 3. Binding hot-spot residues are critical for the interaction.

(A) The human TGF- β type II receptor extracellular domain is shown in white surface, with the most important residues (hot-spots) for binding to TGF- β 3 according to SKEMPI database in red. On the right, the complex structure is also shown, with ligand in white ribbon (PDB 1KTZ). (B) In red are shown the hot-spot residues predicted from pyDock docking calculations, with NIP > 0.2. They are consistent with the known experimental hot-spots, and are located at the interface in the complex structure.

Figure 4. Binding cavities are difficult to identify at protein-protein interfaces.

Bcl-XI / Bak complex is a clear example of the problems arising when trying to find binding pockets at protein-protein interfaces. Protein-protein interfaces may have large flat pockets that are difficult to identify with methods originally developed to detect deep cavities, such as Fpocket. (A) Bcl-XI (white surface) is shown in the complex (PDB 2YXJ) with one of the known

inhibitors (in green) for the interaction with Bak. The cavity predictor Fpocket identifies two different pockets at the inhibitor binding site, which corresponds with the two deep sub-pockets (orange and pink) in which the inhibitor anchors to Bcl-XL surface. (B) Bcl-xL is shown in complex (PDB 1YSG) with two small molecules that need to bind simultaneously to inhibit interaction with Bak. In this case, Fpocket detects the two binding sites (pink and orange), which could be merged into a single one and overlap with a longer inhibitor. These problems should be considered when devising virtual screening or fragment-based search of inhibitors.

Figure 5. Computational approaches for rational drug discovery targeting protein-protein interactions.

This shows a scheme of a general pipeline for drug discovery targeting protein-protein interactions, focusing on the different computational approaches that can help in each phase. An important part of target characterization is the identification of hot-spot residues in protein interfaces. Docking-based hot-spot prediction can help to locate binding cavities for hit identification, as well as in the process of lead optimization.

Figure 6. Hot-spot residues and pathological mutations in the interaction between Type I TGF- β receptor with FKBP12.

The x-ray structure of the complex between Type I TGF- β receptor (surface) and FKBP12 (ribbon) is shown (PDB 1B6C). (A) Residues involving mutations annotated in Humsavar (pathological: T200I, S241L, D266Y; unclassified: N267H) are shown in blue or magenta. Predicted hot-spot residues by docking (using pyDock NIP values) are shown in red or magenta. Interestingly, N267 residue (from N267H mutation, annotated as unclassified) is predicted as hot-spot (magenta). (B) Detail of protein-protein interface. Residues involving mutations T200I, D266Y, and N267H are shown before (left) and after modeling mutation (right).

Table 1. FDA-approved drugs that are targeting protein-protein interactions [89,90].

PPI target	Drug	Disease	PDB ID	Drug Bank ID
<i>PPI inhibitors</i>				
Bcl-2 family	Venetoclax	Cancer	-	DB11581
BIII	Gabapentin	Epilepsy 1 st ; Neurophatic pain	-	DB00996
c-Myc/Max	Nadroparin	Cardiovascular	-	DB08813
CCR5/gp120	Maraviroc	HIV	4MBS	DB04835
HIF-1a	Carvedilol	Cardiovascular	-	DB01136
IL-2/IL2-R	Apremilast	Psoriatic arthritis	-	DB05676
KEAP1/NRF2	Dimethyl fumarate	Multiple Sclerosis	-	DB08908
LFA1/CAM1	Lifitegrast	Dry eye	-	DB11611
PPAR-gamma/NCOA	Rosiglitazone	Diabetes	4EMA	DB00412
Rac1	Azathioprine	Asthma	-	DB00993
S100B/p53	Olopatadine	Itching eyes	-	DB00768
STAT5	Dasatinib	Cancer	-	DB01254
Tubulin	Griseofulvin	Tinea infections	-	DB00400
α Ib β 3	Tirofiban	Cardiovascular	2VDM	DB00775
<i>PPI Stabilizers</i>				
Cyclophilins	Cyclosporine	Gaft rejection	1CWA	DB00091
Immunoglobulin FKBP1A	Tacrolimus	Imunosuppressor (after transplant)	1FKJ	DB00864
Transthyretin/RBP	Diflunisal	Reumathoid arthritis and osteoarthritis	3D2T	DB00861

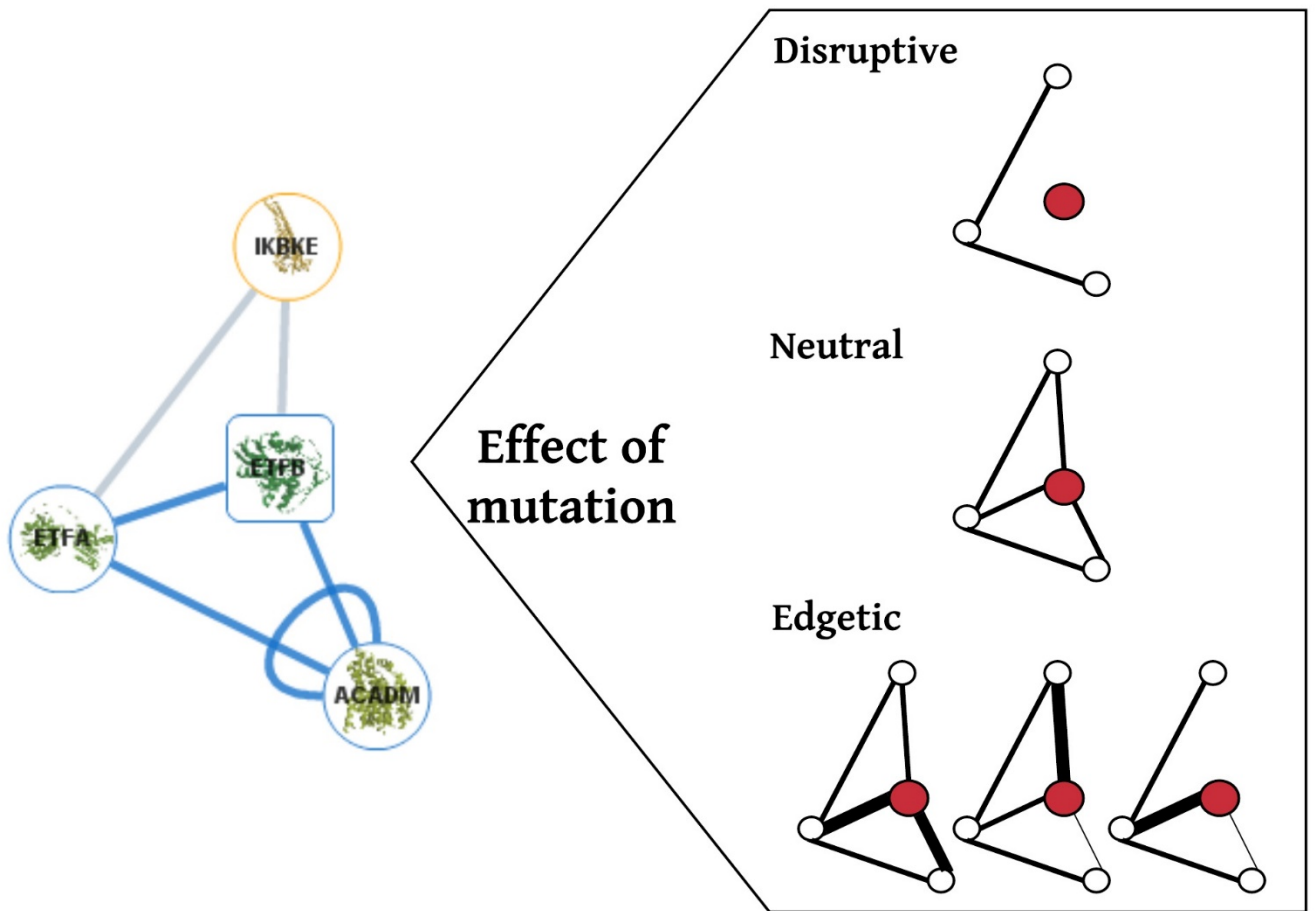


FIGURE 1

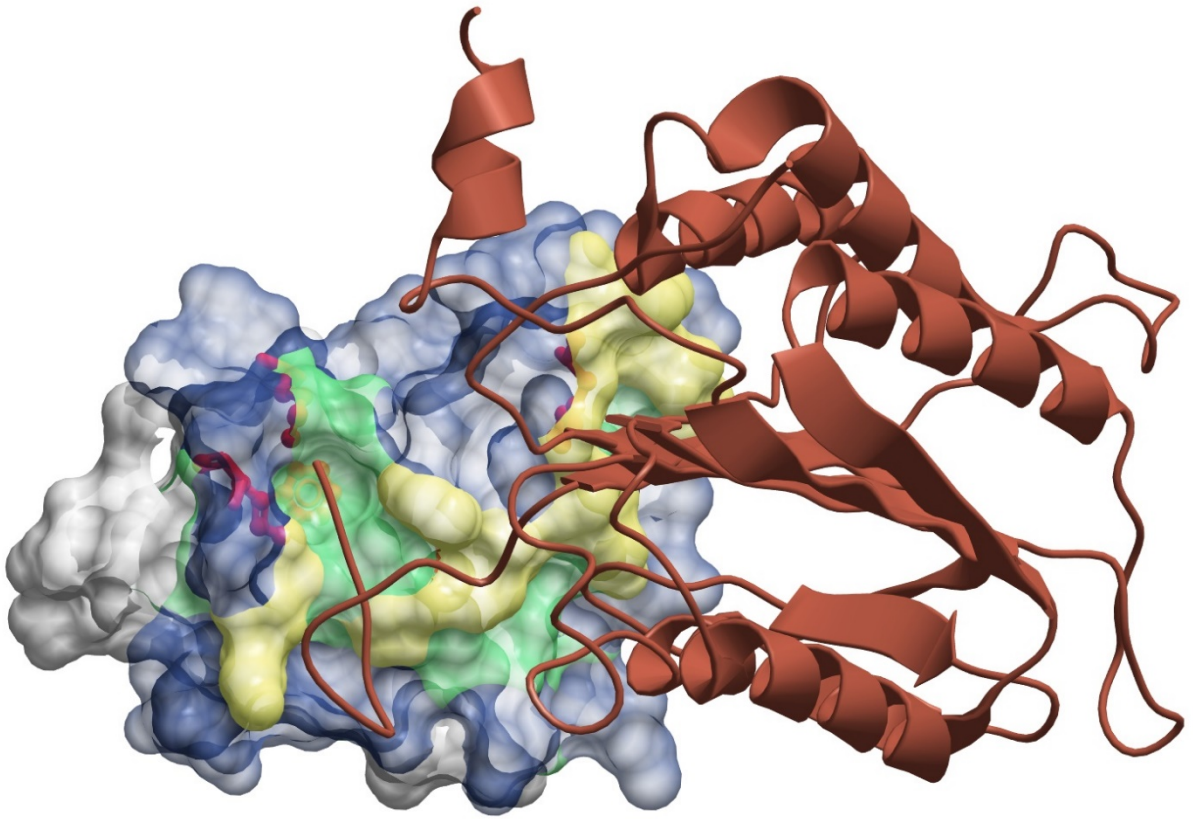


FIGURE 2

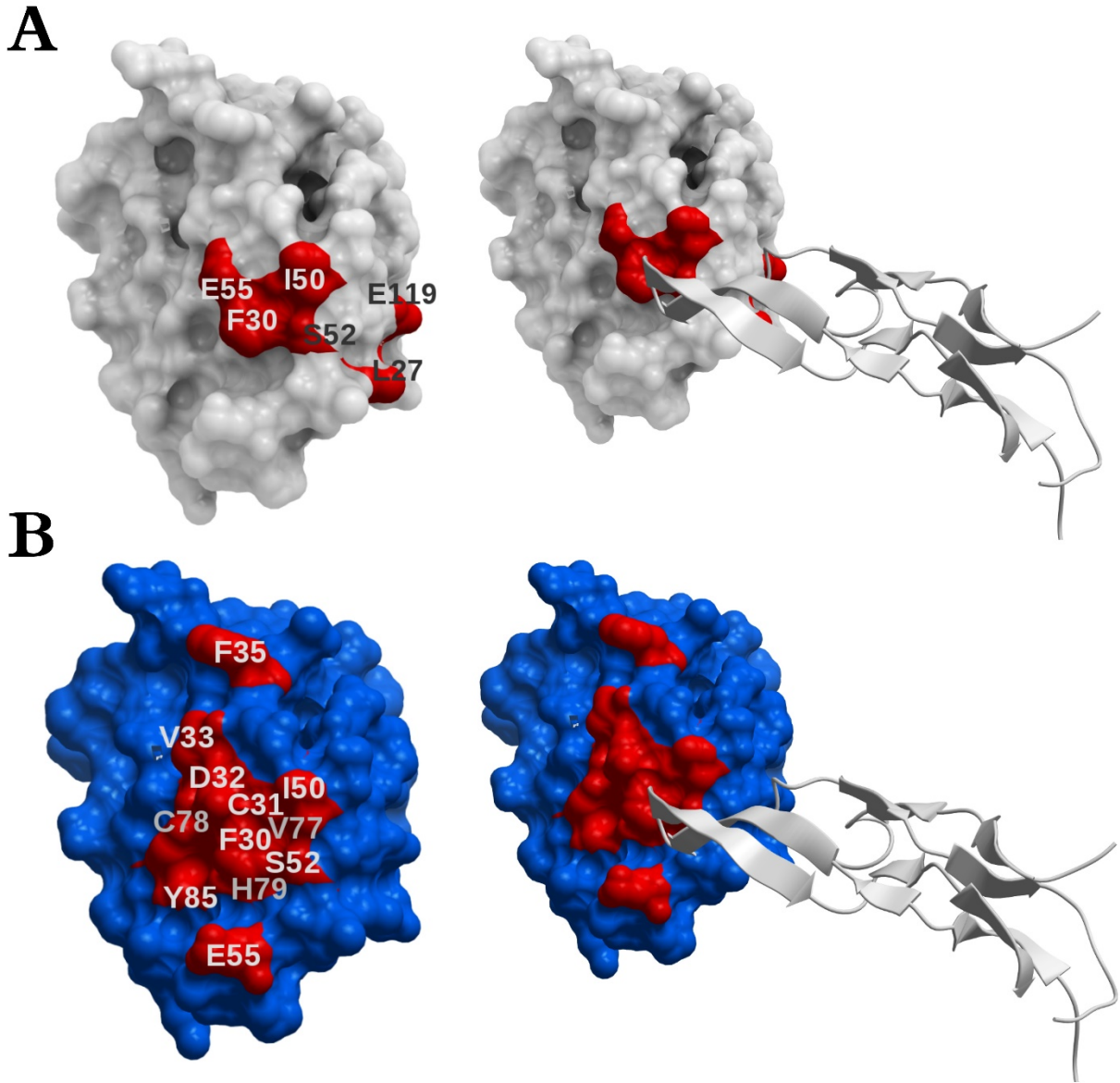
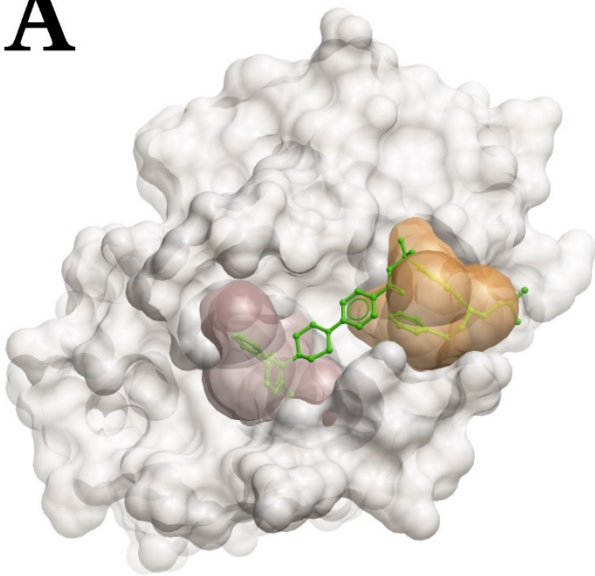


FIGURE 3

A



B

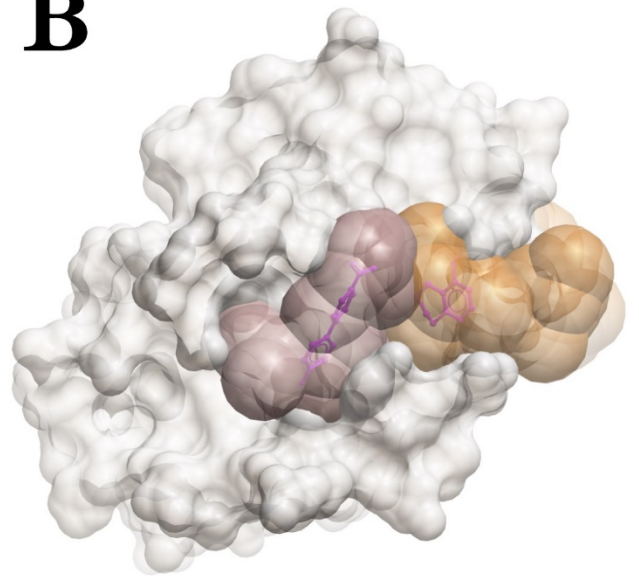


FIGURE 4

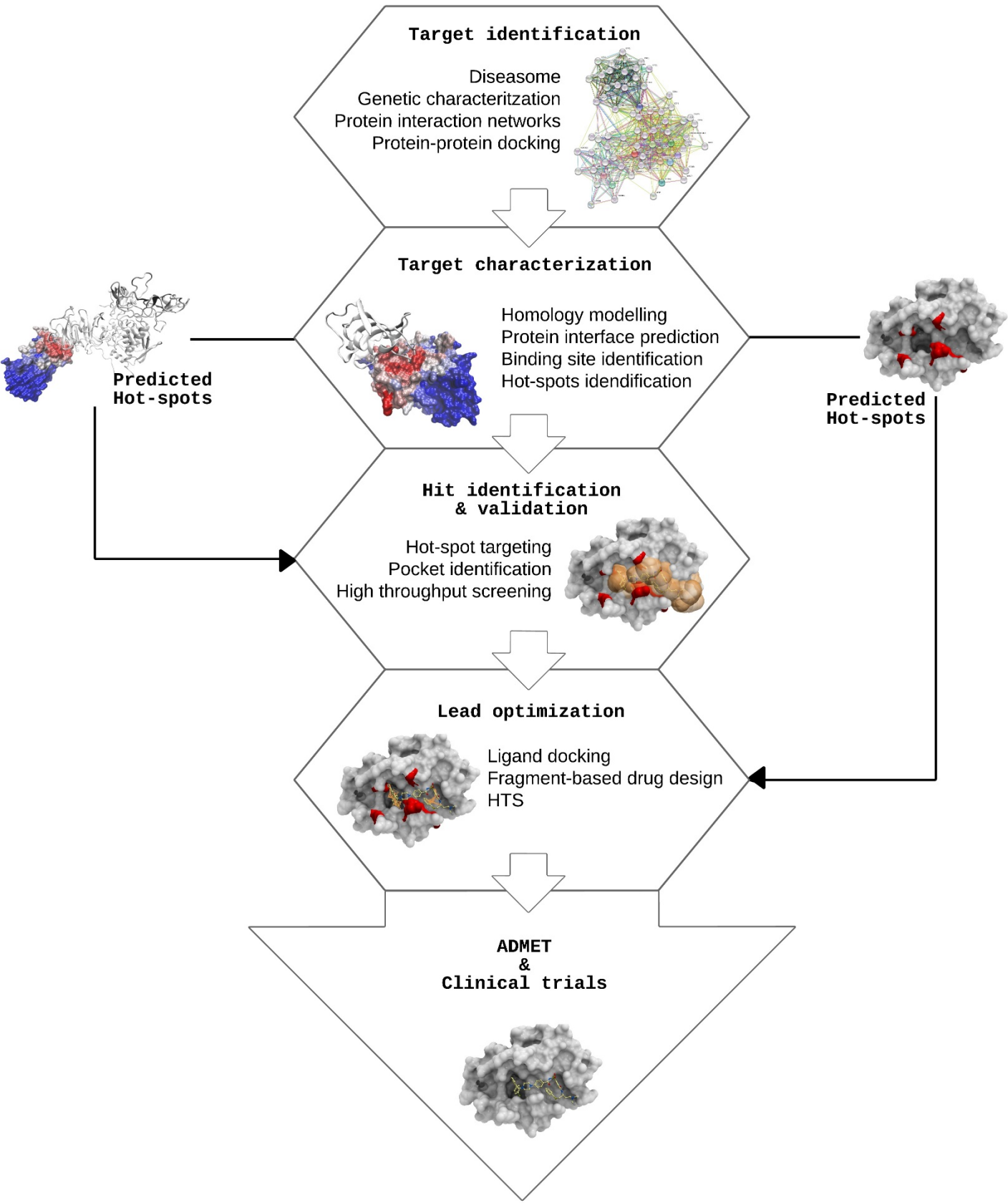


FIGURE 5

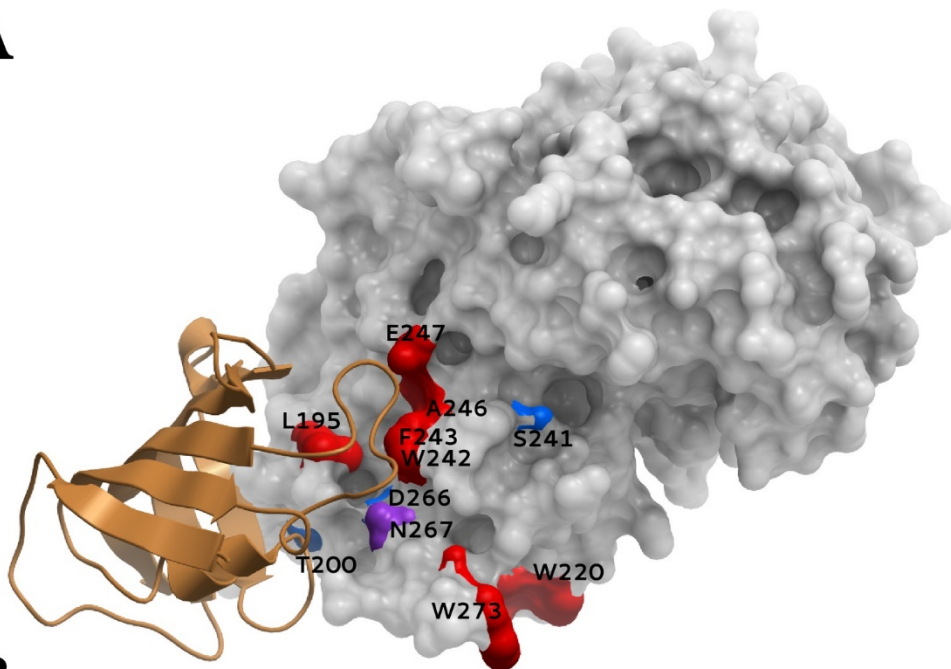
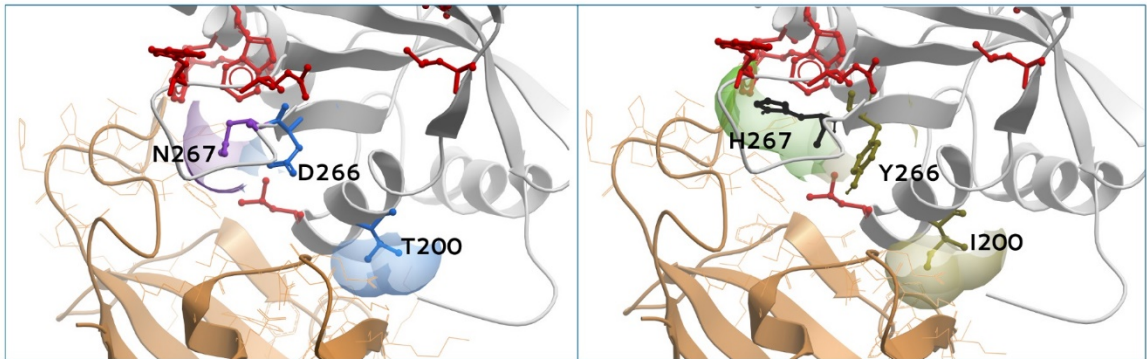
A**B**

FIGURE 6