Computer applications for prediction of protein–protein interactions and rational drug design

Solène Grosdidier¹
Max Totrov²
Juan Fernández-Recio¹
¹Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain; ²Molsoft LLC, La Jolla, CA, USA

Abstract: In recent years, protein–protein interactions are becoming the object of increasing attention in many different fields, such as structural biology, molecular biology, systems biology, and drug discovery. From a structural biology perspective, it would be desirable to integrate current efforts into the structural proteomics programs. Given that experimental determination of many protein–protein complex structures is highly challenging, and in the context of current high-performance computational capabilities, different computer tools are being developed to help in this task. Among them, computational docking aims to predict the structure of a protein–protein complex starting from the atomic coordinates of its individual components, and in recent years, a growing number of docking approaches are being reported with increased predictive capabilities. The improvement of speed and accuracy of these docking methods, together with the modeling of the interaction networks that regulate the most critical processes in a living organism, will be essential for computational proteomics. The ultimate goal is the rational design of drugs capable of specifically inhibiting or modifying protein–protein interactions of therapeutic significance. While rational design of protein–protein interaction inhibitors is at its very early stage, the first results are promising.

Keywords: protein–protein interactions, drug design, protein docking, structural prediction, virtual ligand screening, hot-spots

Introduction

Protein–protein interactions (PPI) are involved in most of the essential processes that occur in living organisms, such as cellular communication, immunological response, and gene expression control. A detailed energetic and structural knowledge of these interactions is necessary to understand the complex regulatory and metabolic interaction networks that occur in living organisms, with the ultimate goal of designing drugs for blocking or modifying interactions of therapeutic interest. Thus, targeting PPI of therapeutic interest with small-molecule compounds is becoming the Holy Grail of drug discovery. A number of experimental and computational methods have been reported to contribute to all the stages of the drug discovery process targeting PPI. High-throughput experimental methods, such as coexpression analysis¹ and the yeast two-hybrid test,² may be used to establish an interaction, and random mutagenesis³ to locate the interaction surfaces. Finally, X-ray crystallography and/or nuclear magnetic resonance (NMR) provide the most detailed structural information of the atomic interactions in a protein–protein complex. However, although the number of three-dimensional (3-D) protein structures deposited in
the Protein Data Bank (PDB) is rapidly growing, only a small fraction of numerous protein–protein complexes, frequently transient, has been experimentally characterized so far. The increasing availability of high-performance computing has favored the development of computer tools that can help in this task.

Computational prediction of a protein–protein complex geometry from the 3-D coordinates of the individual proteins involved has a relatively short history. Early docking methods used purely geometrical criteria to evaluate the resulting solutions, and considered the conformation of the molecule side-chains as fixed (rigid-body assumption). While this approach demonstrated the ability to rebuild a protein–protein complex from its already mutually adjusted subunits, it was not accurate enough to model the induced fit of the interacting surfaces upon binding, and therefore the prediction results were clearly poorer when using the uncomplexed subunits. In order to perform more realistic simulations, recently developed docking methods include interface explicit or implicit flexibility, and a more accurate energy function. While a 100% reliable automated prediction of the association of two proteins is beyond the reach of current methods, advances in energy calculations and in global minimization algorithms, together with the increasing availability of computing power, may lead to useful predictions at a proteomic scale in the next few years.

The development of computational methods to model protein–protein docking, identify promising binding pockets, and predict protein-ligand association will facilitate the discovery of small molecules capable of inhibiting or modifying PPI, a major new challenge in drug design. We can envisage a general strategy for a multidisciplinary drug discovery process that targets PPI of therapeutic interest, involving four major stages (Figure 1). In the following sections each of these stages will be discussed in detail, with the emphasis on the computerized techniques (Table 1).

Identification of PPI of biomedical importance

A plethora of biochemical and genetic experiments, such as cross-linking, co-immunoprecipitation and co-fractionation by chromatography, among others, have been traditionally used to establish specific interactions between proteins. Biophysical assays have been also developed to experimentally measure kinetic and thermodynamic binding constants between two given proteins. From all the experimental methods, the yeast two-hybrid assay and correlation of mRNA expression profiles have propelled large-scale detection of PPI.

In recent years, with all the available information derived from the genome sequencing projects, several computational tools have been applied to find and recognize PPI from genome sequences at a proteomic level. Analysis of co-evolution of proteins and gene fusion events can be used to detect putative PPI. A well-known study described the use of combination of techniques (correlated evolution, correlated mRNA expression profiles, and domain fusion patterns) to find 93,750 pairwise links between 4,701 (76%) functionally related yeast proteins, from which 4,130 links (between 1,223 proteins) were of the ‘highest confidence’ (validated by direct experimental techniques or by two of the three prediction techniques). Other methods are based on interacting domains, interacting motifs, and a variety of criteria such as similarity of phylogenetic trees, protein interaction network topology, signature products or genome-wide coevolutionary networks. There are some good recent reviews that give a complete view of the currently available methods to identify PPI.

Other computational approaches focus on mining the literature, the whole ‘googleome,’ for PPI. A system for automatic detection of PPI extracted from scientific abstracts was able to rebuild key interactions of the Drosophila cell cycle control for 33 of the 91 protein names used in the bibliography screening. A similar system, based on a general-purpose information extraction engine, identified interactions between two proteins from Medline abstracts with an accuracy of 77% and a coverage of 58% of the total interactions. Another method used discriminating words to identify Medline abstracts that described protein interactions, with an accuracy of >77% and a coverage of ~50% (or 100% of accuracy with a coverage of ~30%). A new text-mining method (PIE: Protein Interaction information Extraction system) is available on the web to extract PPI from literature (http://pie.snu.ac.kr/). This tool, consisting on an article filter followed by a sentence filter, has been trained on the BioCreAtIve II workshop dataset, enriched by other selected known-interactions. Using a 10-fold cross validation and 0.5 probability cutoff, the method showed a precision of 87.4% for the article filter, and 92.1% for the sentence filter.

All the experimental and computational data on existing PPI were soon organized in various public databases: YPD and WormPD – Yeast and Caenorhabditis elegans Proteome Databases; MIPS – Munich Information Center for Protein Sequences; DIP – Database of Interacting Proteins;
BIND – Biomolecular Interaction Network Database;\textsuperscript{41} and private ones, such as PathCode\textsuperscript{TM} from GPC-Biotech (http://www.gpc-biotech.com/). This facilitated large-scale studies that aimed to map the network of PPI of complete living organisms. The first described maps were those of the hepatitis C virus;\textsuperscript{42} vaccinia virus;\textsuperscript{19} Saccharomyces cerevisiae;\textsuperscript{43–45} Caenorhabditis elegans;\textsuperscript{46} or Helicobacter pylori.\textsuperscript{47} Other organisms followed, at different levels of completeness.\textsuperscript{48–51} The field is rapidly growing, and there are currently many web tools and data collections that are publicly available online (http://www.imb-jena.de/jcb/ppi/jcb_ppi_databases.html).

**Target characterization:** Structural information about the protein–protein complex and drugability of the targeted PPI

**Location of the interface**

Mutational studies

Alanine-scanning mutagenesis\textsuperscript{3} combined with kinetic and thermodynamic measurements can be used to experimentally locate and characterize residues involved in PPI. A comprehensive database of energetic data for different

---

**Figure 1** Flow-chart of the drug discovery process targeting protein–protein interactions.
protein–protein complexes, determined by alanine mutations (ASEdb), has been compiled and made publicly available (http://www.asedb.org). This database is actively updated and it is commonly used both by experimentalists and by computational benchmark studies.

**Computational prediction of protein–protein interfaces**

Can protein–protein interfaces be predicted from the structures of their components, or, in other words, are there specific chemical and physical characteristics on a protein–protein interface that we could use to predict protein-binding sites on a protein surface? The question is far from being solved. Pioneer studies found that PPI sites have specific structural characteristics that differentiate them from other areas of the protein surface. However, when oligomeric proteins were excluded from the analysis, the results showed that chemical composition of protein–protein interfaces does not seem to differ greatly from the rest of the solvent-accessible surface.

Although chemical and physical complementarity between the interacting surfaces is essential for the recognition, it is difficult to find simple chemical or structural patterns on the surface of proteins that unequivocally define a protein recognition site.

Alternative strategies for prediction of protein–protein interaction sites have been recently developed (Zhou and Qin recently reviewed all the different approaches). Amongst them, methods based solely on sequence information have been reported. Receptor-binding domains were predicted by analyzing hydrophobicity distribution on protein sequences. Predictions were 59%–80% correct, depending on the database of protein interactions used. A neural network method that uses sequence profiles and solvent exposure of neighboring residues has been reported. The method was trained on 615 pairs of nonhomologous protein–protein complexes (homodimers and heterodimers), and was tested on different sets of bound and unbound proteins. In the case of unbound proteins, 70% of the predicted residues were correctly located at the protein–protein interfaces. More recently, Ofran and Rost developed a machine learning-based method called ISI (Interaction Sites Identified from Sequence) to identify interacting residues from protein sequences only. They combined predicted structural features with evolutionary information with no reference to the 3-D structure of the protein, and the strongest interface residue predictions reached 90% of accuracy in a cross experiment. Another method, based on a 3-D cluster analysis that evaluates residue conservation on a set of 35 protein families, can identify interfaces and functional residues.

In addition to conservation, a combination of physical and empirical methods can give promising results for interface prediction, as in the Promate server (http://bioinfo.weizmann.ac.il/promate/). Considering energy-based approaches, the optimal docking area (ODA), a method based on the hypothesis that desolvation must play a central role during protein–protein binding, identifies continuous surface patches with optimal docking desolvation. This approach has been validated on 66 unbound non homologue protein structures involved in nonobligate protein–protein heterocomplexes and the ODA predicted regions were correct in 80% of the cases. The strategy has been applied to numerous cases of biological and therapeutic interest, with excellent predictive results.

**Structure determination of protein–protein complexes**

Once a target protein–protein interaction has been established, it is desirable to obtain the most detailed structural information at atomic resolution of the protein–protein complex by X-ray crystallography and/or NMR experimental techniques. During the last decades, a number of protein–protein complex structures of therapeutic interest have been solved and deposited in the PDB. However, solving the 3-D structure

---

**Table 1 Computer approach to rational design of inhibitors/enhancers of protein–protein interactions**

<table>
<thead>
<tr>
<th>Target identification</th>
<th>Correlated mRNA expression profiles; correlated elution; domain fusion patterns; automated literature mining</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target characterization</strong></td>
<td>Surface analysis; hydrophobicity profiles; 3-D cluster analysis; residue conservation</td>
</tr>
<tr>
<td>1. Establish protein–protein interaction</td>
<td>Rigid-body docking; energy minimization; side-chain refinement; flexible docking</td>
</tr>
<tr>
<td>2. Locate interface</td>
<td>Analysis of “hot-spots”; surface cavities</td>
</tr>
<tr>
<td>3. Modeling protein–protein interaction</td>
<td>Energy minimization; graphic modelling</td>
</tr>
<tr>
<td>4. Finding putative small-molecule pockets</td>
<td>Flexible ligand docking; grid or explicit receptor representations; MC minimization</td>
</tr>
<tr>
<td><strong>Lead discovery and optimization</strong></td>
<td></td>
</tr>
<tr>
<td>5. Mimicking interface</td>
<td></td>
</tr>
<tr>
<td>6. Ligand docking</td>
<td></td>
</tr>
</tbody>
</table>
of a protein–protein complex is still a long and difficult task, and the number of available coordinates of protein heterodimers is relatively small compared to the number of deposited individual protein structures. Therefore, there is a need for reliable computational tools that can predict protein–protein complex formation and help to theoretically analyze the phenomenon of association between proteins.

Protein–protein docking prediction

The so-called “protein docking” problem, that is the prediction of a protein–protein complex using only the coordinates of its separate subunits, is one of the major challenges in structural biology. Apart from the intrinsic academic interest in characterizing the determinants of molecular recognition, the scientific community increasingly requires computational tools to model the physiological interactions in which a protein is involved, once its 3-D structure has been solved. Given that the number of available 3-D structures of individual structures is significantly increasing with the upcoming structural genomics projects, and considering that solving the structure of a protein–protein complex is often qualitatively more difficult than solving its individual subunits, one can easily deduce the importance of computational prediction of protein–protein complexes for the proteomics era. For that reason, during the last 20 years, a variety of computational algorithms for automatic protein–protein docking have been developed.

Geometry-based protein–protein docking methods

The analysis of protein–protein complex structures at atomic resolution gave the first glimpse of the determinants of protein docking. From the analysis of several protein–protein complex structures, the most obvious observation was that protein surfaces of interacting proteins at binding sites were often highly complementary (Figure 2). For that reason, early protein–protein docking algorithms were based on purely geometric criteria, aiming to maximize the shape complementarity between the two interacting molecules. Conformational search of the best fit was performed on the rigid-body (ie, fixed backbone and side-chain conformation) representation of molecules, by geometric methods such as ‘sphere-matching’ in the original DOCK algorithm.

Many of the recently developed docking methods are still mainly based on the shape complementarity criterion and the rigid-body assumption. In these methods, efforts have been directed towards improvement of spatial conformational search by introduction of new minimization techniques. Simulated annealing by using Monte Carlo simulations facilitated the use of constraint-driven docking. One of the most important advances was the use of Fourier transformation techniques to rapidly evaluate all possible translations between the molecules in a given orientation in order to find the best geometry matching. This method actually constitutes the basis of some of the most popular rigid-body docking approaches nowadays (eg, FTDOCK or ZDOCK). Other successful geometric-based docking methods are Hex and MolFit.

In general, geometry-based rigid-body docking methods were able to find and score properly the correct solution when using the 3-D coordinates of the complexed subunits during simulations. However, when these methods were tested on real cases, using the 3-D coordinates of the uncomplexed subunits, the correct solution was often not properly discriminated from the false positives, or even was not found at all. Clearly, the geometry criterion was valid to rebuild a complex after separation of its bound subunits, given the additional induced shape complementarity of their

Figure 2 Shape complementarity at the interface of a protease-inhibitor complex (chymotrypsin/APPi; PDB code 1ca0).

Abbreviation: PDB, Protein Data Bank.
surfaces, but it was not able to correctly dock the unbound subunits, because their interacting surfaces are not always complementary enough. Thus, in order to model the induced fit that occur upon protein–protein association, it was necessary to overcome the rigid-body assumption, and to include in the scoring function other binding determinants than pure geometrical complementarity.

**Protein docking as a global energy minimization problem**

During formation of a protein–protein complex, the interacting interfaces of the approaching subunits fit each other to reach the native bound conformation. Since protein complexes are, in general, thermodynamically stable systems, the native bound conformation should represent the global minimum of the free energy, and therefore the docking problem can be reduced to finding this global minimum. From this point of view, geometry-based docking methods considered that the interaction energy was proportional to the contact area. Whereas this geometry-based approach can account reasonably well for the van der Waals interactions, it is clearly insufficient to describe other contributions to the interaction energy. Thus, different docking methods were developed to include other binding determinants, such as hydrogen bonding, electrostatic energy, solvation or hydrophobicity.

At the same time, finding the global minimum of the free energy for a protein–protein association presented a conformational search challenge. As the rigid-body approach was insufficient to simulate the induced conformational fit upon binding, docking methods started to include strategies to mimic molecular flexibility during the optimization (Bonvin reviewed all the diverse strategies developed to deal with molecular flexibility upon binding during the docking process). The most practical strategy was the softening of the scoring function by imposing some limiting values to the steric energy terms, thus allowing some overlap of the interacting surfaces. This strategy overcame, to some extent, the difficulties stemming from the use of the unbound conformations of the interacting molecules. Explicit treatment of flexibility could lead to a more accurate description of the protein–protein complex formation phenomenon, but a full conformational search is currently impractical. However, since molecular association involves only small conformational changes in most of the known protein–protein complexes, computational requirements can be dramatically lowered by limiting conformational flexibility to interface side chains.

The first docking method that considered continuous flexibility of interface side-chains during the global minimization process was based on internal coordinate mechanics (ICM). The ICM flexible docking procedure, successfully applied to the prediction of an antibody-lysozyme complex, was tested in a blind prediction contest. Although the ICM pseudo-Brownian method with subsequent global optimization of the interface side chain rotations lead to promising results, it was computationally too expensive to be tested on large databases of complexes. Therefore, an alternative two-step docking procedure (rigid body docking followed by ICM side-chain optimization) was proposed. The docking method used a fast soft interaction energy function pre-calculated on a grid, similar to the fast ligand docking procedures, which drastically increased the speed of the procedure. The scoring function used to evaluate the rigid-body docking poses was further optimized, for a better selection of docking solutions before the refinement step. The scoring function, composed by Coulombic electrostatics and ASA-based desolvation energy with atomic solvation parameters optimized for protein–protein docking, was later implemented in a docking protocol called pyDock, which was able to score docking sets generated by different docking methods. Other docking and/or scoring schemes that are based on energy description are Haddock, ClusPro/SmoothDock, RosettaDock and ATTRACT.

Baker and colleagues improved side-chain modeling during docking significantly using a rapid and efficient method for sampling off-rotamer side-chains conformations by torsion space minimization. Their approach to include flexibility yielded better energetic discrimination between correct and incorrect docking models and a significant improvement in the quality of their predictions. Other approaches aim to include backbone flexibility by using conformational ensembles of the unbound subunits previously computed by Molecular Dynamics or by precalculating soft collective degrees of freedom by normal mode analysis (NMA) that are later used to include flexibility during docking. However, fully unrestricted molecular dynamics are too costly for routine application during docking. Nevertheless, there are important advances, as the use of steered molecular dynamics to give insights into the energy determinants and mechanism of TCR-pMHC association.

**Benchmarking protein–protein docking methods**

To be used in practical applications, the protein–protein docking methods first have to be validated on a sufficiently large and diverse set of experimentally solved complex structures, ideally with individual subunit structures also experimentally determined. The problem is that there are
not so many cases suitable for benchmarking. In one of the earliest benchmarking efforts, Norel and colleagues tested a rigid-body docking method on a set of complexes, starting from bound and unbound components. When the unbound subunits for both partners (in four different complexes) were used, the near-native solution had the lowest energy (eg, was identified as best docked) in only one case. A FTDOCK docking method with refinement of binding side-chains was also benchmarked in a set of complexes (five cases when using unbound subunits). The near-native solution was ranked below 20 in all five cases, but it was predicted as the lowest energy solution in only one. Bigger rigid-body method was later applied to 11 protein–protein complexes using the unbound subunits. The near-native solution was found among the docked conformations in eight out of the 11 cases, but was not ranked first in any of them. The ICM protein–protein docking was applied to a set of twenty-four protein–protein complexes (starting from the 3-D coordinates of bound and unbound subunits). When the unbound subunits were used, the near-native conformation was ranked below 20 in 85% of the complexes with no major backbone rearrangement upon binding (it was ranked 1 in 64% of the protease-inhibitor complexes). Recently, the laboratory of Weng has made an effort in providing suitable sets of cases for benchmarking of protein–protein docking methods. Nowadays, it is almost standard to provide success rates on these benchmarks.

In 2001, an international protein–protein docking experiment called Critical Assessment of PRedicted Interactions (CAPRI; http://www.ebi.ac.uk/msd-srv/capri/capri.html) was launched, based on the CASP experiment model for single protein structural prediction. CAPRI is a blind test to evaluate the capacity of protein–protein docking algorithms to predict the binding-mode of two interacting proteins. This experiment allows direct comparison of different docking algorithms and permits also to follow the evolution of the performance of the most popular docking methods along time. Table 2 shows the results of all the CAPRI rounds that have been assessed so far, where the performance of the most active groups can be compared. In Figure 3 we can see three different examples of CAPRI results for the ICM and pyDock methods for targets 6, 14, and 25.

Evaluate suitability for small-molecule binding (drugability)

Computational methods to analyze the small-molecule druggability of a target protein–protein interface focus on the existence of ‘hot-spots’ and/or small pockets. Although the overall chemical composition of protein interfaces does not significantly differ from the rest of the solvent-accessible surface, structural analysis and experimental studies on protein–protein complexes underline the existence of ‘hot-spots’, eg, a few residues that confer most of the binding energy. These ‘hot-spots’ can be potential targets for small molecule drug discovery. Indeed, a specific interaction may be disrupted by targeting one or several of its hot spots. Consequently, low molecular weight compounds satisfying the requirements for orally deliverable drugs can be used to interfere with recognition sites in protein–protein interfaces that are usually above 800 Å². Hot-spots can also be particularly helpful in difficult cases in which no small cavities are identified in flat protein–protein interfaces.

Experimental measurement of residue contributions to binding energy can be done by Alanine Scanning Mutagenesis combined with biophysical methods but this is a quite costly way to identify hot-spots. Therefore efforts have focused on computational prediction of these residues, and a variety of approaches have been reported based on residue conservation, machine-learning algorithm from protein sequence alone, hydrogen bonding, complete binding energy evaluation or propensity calculation from rigid-body docking. In Figure 4 we can see the high correlation between the hot-spot predictions from docking and the known experimental data for the IL4–IL4 receptor α chain.

The protein–protein interfaces most easily targeted with small molecule drugs typically contain a sufficiently deep surface pocket suitable for small molecule binding. Experimental and computational prediction of binding pockets on the surface of proteins has been successfully applied to rational drug design, and thus they can be one of the first-choice computational tools to characterize a protein–protein interface in search of potential pockets.

Discovery of inhibitors/promoters of PPI

Given the role of PPI in regulating the majority of biological functions, PPI inhibition has long been one of the major goals in drug design. Empirical discovery of small compounds that can disrupt PPI has been traditionally difficult, and structure-based design of PPI inhibitors is currently limited to only a few successful cases. However, recent development of computational methods for protein–protein and protein-ligand docking is expected to facilitate the rational discovery of small compounds that can modify PPI. Several reviews of PPI modulation by small molecules have been published. Although the current review focuses on...
Table 2 Prediction results of different protein–protein docking methods in the first three CAPRI editions (from target 1 to 28)\(^{12-14}\)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>1st CAPRI edition targets</th>
<th>2nd CAPRI edition targets</th>
<th>3rd CAPRI edition targets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
<td>T3</td>
</tr>
<tr>
<td>Abagyan</td>
<td>0</td>
<td>0</td>
<td>**</td>
</tr>
<tr>
<td>Fernandez-Recio</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Baker</td>
<td>0</td>
<td>0</td>
<td>**</td>
</tr>
<tr>
<td>Bates</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Bonvin</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Camacho</td>
<td>*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CLUSPRO</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Eisenstein</td>
<td>*</td>
<td>*</td>
<td>0</td>
</tr>
<tr>
<td>Gray</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Palma</td>
<td>–</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Krippahl</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ritchie</td>
<td>0</td>
<td>0</td>
<td>**</td>
</tr>
<tr>
<td>Sernberg</td>
<td>0</td>
<td>*</td>
<td>0</td>
</tr>
<tr>
<td>Smith</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Takeda-Shikata</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ten Eyck</td>
<td>*</td>
<td>*</td>
<td>0</td>
</tr>
<tr>
<td>Vakser</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Weng</td>
<td>0</td>
<td>**</td>
<td>0</td>
</tr>
<tr>
<td>Wolfson</td>
<td>*</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Zacharias</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**Notes:** Only the groups and servers that participated in more than half of the targets are listed in this table. The symbol "–" means that no predictions were submitted. The quality of the best submitted model for each group is described as follows: 0, no acceptable models were submitted; *Acceptable quality; **Medium quality; ***High quality model. The type of targets are indicated as: U, unbound; B, bound; M, model by homology.
computational approaches, we will overview here several examples of experimental discovery of compounds that can modify PPI, and we will give later more extended information on rational design of new PPI inhibitors based on structural data and computer simulations.

Discovery of PPI agonists/antagonists: experimental approaches
Phage-display selection of peptides
Phage display has been used to probe hot-spots as well as to identify novel peptide agonist/antagonists of PPI.\(^{158}\) For example, it has been used to select small peptides that can induce oligomerization in different cytokine receptors.\(^{157,159,160}\) Especially interesting was the case of the 20 residue cyclic peptide EMP-1,\(^{161}\) which induced dimerization of the EPO receptor. A small change in this sequence transformed its agonist character into antagonist.\(^{162}\) Cwirla and colleagues also identified small peptides that can induce receptor dimerization in thrombopoietin receptor (TPOR).\(^{163}\) An interesting drug discovery platform that used phage peptide libraries and HTS of small molecules based on the selected peptides was reported.\(^{164}\) This strategy was applied to the discovery of agonist/antagonist peptides and small molecules in the insulin-like growth factor-1 (IGF-1)/IGF-1 receptor system, and has been also used to identify a binding epitope and potential protein–protein interaction partners of a given protein, by searching in the sequence databases. In the case of the insulin receptor, both agonists and antagonists have been discovered using phage display. This technology even allowed a better understanding of the receptor molecular architecture with identification of critical regions required for its biological activity.\(^{165}\) Potent antagonists called “zeta” peptides of the high-affinity immunoglobulin E (IgE) receptor have also been identified and prevent histamine release from cultured cells. Moreover, these peptides that acts as competitive IgE inhibitors can be used for further design of IgE inhibitors.\(^{166}\)

Discovery and optimization of natural ligands
Natural products didemnaketals A and B were used to synthesize simplified analogs that inhibited HIV protease homodimerization.\(^{167}\) Chalcone derivatives, with known anticancer properties, were recently described to inhibit interactions between the human oncoprotein MDM2 and p53 tumor suppressor protein.\(^{168,169}\) Cyclodextrin dimers (CD) that disrupt PPI by targeting hydrophobic patches have been also reported.\(^{170}\) Interestingly, small molecules and peptides can also induce an unwanted stabilization of a protein–protein complex. This is the case of Brefeldin A, a small hydrophobic compound produced by toxic fungi that has dramatic effects on mammalian cells. It has been proposed that brefeldin A works as an uncompetitive inhibitor stabilizing a transient “dead-end” complex between Arf exchange factor and Sec7 domain of Gea1, Gea2 and Sec7 proteins.\(^{171}\) On the other
side, a large number of natural compounds have been known to target the tubulin-tubulin interaction to stop cancer cells division, and some new molecules are currently used in clinical trial for cancer therapy.

**High-throughput screening of small molecules**

High-throughput screening (HTS) methods have been used to discover small compounds capable of inhibiting PPI, especially when no structural information is available about the target proteins. In general, HTS is less successful in identifying PPI inhibitors than in identifying any other type of inhibitor: PPI are extended over a big interface (average binding energy per unit area: 9 cal/mol-Å²) and are difficult to target with a small site-specific drug (average binding energy per unit area: 31 cal/mol-Å²).

Tian and colleagues used a high-throughput, cell-based screen to detect small compounds that activated the murine granulocyte-colony-stimulating factor (G-CSF) receptor. They found a small molecule (SB 247464; Figure 5a) that replaced the natural protein ligand G-CSF in its role of inducing oligomerization of G-CSF receptor chains, thus triggering the corresponding signal transduction pathways. Later, it was shown that SB 247464 could dimerize the G-CSF receptor in a different manner than G-CSF, through a Zinc mediated interaction. It also appeared that SB 247464 and G-CSF bound to different sites on the receptor, given that the small compound was unable to compete with G-CSF receptor natural ligand to initiate the dimerization. This constituted one of the first examples of a synthetic small molecule capable of dimerizing cell surface receptors.

Similarly, Kimura and colleagues screened many compounds capable of inhibiting the binding of thrombopoietin (TPO) to the cell surface receptor c-Mpl, necessary for triggering megakaryopoiesis and platelet production cascades after receptor oligomerization. They found two small inhibitor compounds, TM4 and TM41 (Figure 5b), which were able to replace the natural TPO in its role of inducing c-Mpl oligomerization.

HTS techniques have also been used to find a molecule that inhibited interaction between EPO and EPO receptor (Figure 5c). A multimeric form of this molecule was also synthesized and shown to induce dimerization in the EPO receptor, thus mimicking the physiological role of EPO. A cell-based screening assay was also used to select a molecule (L-783,281) that activated insulin receptor (Figure 5d). The mechanism proposed was that L-783,281 bound the tyrosine kinase domain of the insulin receptor, triggering the corresponding signal transduction pathways.

Combinatorial piperazinone libraries have been used to find compounds that disrupt the interaction between the transcription factor LEF-1 and the protein β-catenin, which accumulates in a majority of colorectal tumors. The complex formed between the Tcf4 transcription factor and the β-catenin, also involved in colorectal tumors, has been investigated by screening several thousand of natural compounds, among which six inhibitors in the low micro-molar range were found. A combinatorial chemical library based on a pyrimidineimidazole core has been designed to find inhibitors of the inducible nitric oxide synthase (iNOS). This enzyme that generates NO from L-arginine is involved

---

**Figure 4** Normalized interface propensity (NiP) calculation from rigid-body docking for the prediction of hot-spots on the complex between the IL-4 and its IL-4 receptor α chain (complex PDB code 1iar). A) The IL-4 receptor α chain is represented as grey ribbon, and the dots represent the center of coordinates of each of the 12000 rigid-body docking solutions poses colored according to the pyDock scoring function on a scale from red to blue: lowest energy values are shown in red. B) The IL-4 receptor α chain surface is colored according to the NiP value of each residue obtained from the 100 lowest-energy docking solutions. The predicted hot-spots corresponding to the highest NiP values (cutoff = 0.4) are shown in red. C) The IL-4 receptor α chain surface is colored according to the available experimental data. The residues shown in red are the ones experimentally known as hot-spot, and the residues in white have no available experimental data. **Abbreviations:** IL, interleukin; PDB, Protein Data Bank.
in tissue damage during inflammation and is fully active as a dimer. By screening, Devlin and colleagues found a class of potent, selective and cell permeable iNOS inhibitors capable of preventing its dimerization.182,183

The oncoprotein c-Myc, over expressed in many human tumors (lung, colon, Burkitt’s lymphoma), requires binding to its activation partner Max in order to interact with DNA and achieve its transcription factor function. Because of its potential therapeutic applications, this interaction was studied by HTS and allowed the discovery of two potent and selective dimerization inhibitors: Mycro1 and Micro2, both in the low micromolar range.184

The complex formed between TNF-alpha and its receptor TNFRc1 were known to be inhibited by antibodies and soluble receptors, but no potent small molecule was reported until Muckelbauer and colleagues performed screening on this system and discovered two inhibiting small-molecule compounds acting through covalent modification of the receptor via a photochemical reaction.185 HTS has also been used to discover two classes of competitive antagonists for the interaction B7.1/CD28, involved into the T-cell response augmentation, with potential therapeutic applications in immunotherapy after transplantation or autoimmune diseases.186 More recently, a Rac activation-specific inhibitor of the Rac1-GEF interaction that could be useful for therapeutic targeting at Rac deregulation has also been found in this way.187

**Figure 5**

A) Molecule SB 247464; B) Molecules TM4 and TM41; C) Inhibitor of EPO-EPOR interaction; D) Molecule L-783, 281.

**Fragment-based lead discovery**

Fragment assembly is a recent approach developed to help finding or optimizing leads during the drug discovery process. A set of small fragments are screened against the protein of interest and the binders are then combined to form small-molecule compounds, which significantly increases the search process efficiency. Indeed, the chances of finding a hit are higher than in conventional HTS.188 Hajduk and Greer analyzed the impact of fragment-based methods in drug design over the last decade,
showing a list of all the targets studied through fragment-based approach that lead to potent inhibitors discovery. For example, a potent inhibitor of the anti-apoptotic Bcl-2 family proteins (Bcl-2, Bcl-XL and Bcl-w) was discovered using NMR-based screening of small fragments combined with structure-based drug design. The molecule, called ABT-737 showed a strong capacity to reduce regression of solid tumors in mice. Tethering is a fragment-based method relying on the reversible formation of a disulfide bridge between the target protein and the fragment. In this way, the search region is controlled by the introduction, by site-directed mutagenesis, of a cysteine residue near the site of interest of the target protein, which in addition facilitates the computational analysis of potential binding modes. With this approach, a known inhibitor (in the millimolar affinity range) of the interaction between IL-2 and its receptor got its affinity significantly increased. The X-ray structure of the previously known IL-2/inhibitor complex revealed an adaptive IL-2 interface, in which a small molecule binding site was created. The application of the tethering approach resulted in a clear improvement of the original molecule to the nanomolar affinity range. Rational design of PPI inhibitors: computational approaches

In principle, the problem of PPI inhibition seems to be just a particular case of the broader drug design field, but a deeper analysis shows intrinsic characteristics that make it a distinct field. While drug design, in general, is focused on the discovery of small compounds that can bind into natural ligand-binding pockets or active centers of proteins of therapeutic interest, inhibition of PPI requires identification of small compounds capable of disrupting a large and highly complementary interaction surface between two proteins (Figure 6). The absence of well defined, deep pockets in protein–protein interfaces, and the large number of intermolecular contacts arising from their high geometrical and chemical complementarity makes the problem especially difficult. Nevertheless, several methods for rational design of PPI inhibitors have been recently applied to particular cases with some success. The constant increase in computational power and the development of new efficient and accurate computer tools for drug design are starting to yield promising results in this very challenging area.

Mimicking protein–protein interfaces:
from peptides to small molecules

Antibodies capable of blocking or enhancing PPI have been reported, for example a monoclonal antibody that can induce homodimerization of erythropoietin receptor and triggers cell proliferation cascades, or a monoclonal antibody that may block critical PPI of HIV-1 integrase. The use of antibodies in cancer therapy is highly promising.

Figure 6 An example of a protein-small molecule interface A: complex trypsin/RPR128515 inhibitor (PDB 1f0u) compared to a protein–protein interface B: complex EPO receptor/EPO (PDB 1eer). The protein-small molecule interface A), with fewer inter-atomic contacts, is noticeably smaller and deeper than the shallow, large protein–protein interface B). The molecules are shown in the same scale.

Abbreviation: PDB, Protein Data Bank.
However, clinical use of antibodies presents in practice numerous problems (high cost for large-scale production, drug delivery, immunoreactivity, etc.). Fortunately, there are some reported examples of design of small molecule inhibitors that mimic an antibody binding function. Based on this strategy, Chrunyk and colleagues proposed to design small compounds to mimic the binding of antibodies that can act as blocking agents of PPI. They selected a monoclonal antibody to block interaction between proteins IL-1β and IL-1R and they found that a simpler single chain antibody retained the same blocking capacity, leaving the door open for future design of PPI inhibitors. Similarly, a calyx[4]arene scaffold with pendant cyclic peptide units was designed as a mimetic of antibody Fab fragments, and was shown to bind cytochrome c in the same region of the protein as its natural protein partners (cytochrome oxidase, cytochrome c peroxidase).

Structure-based design of peptides that mimic structural elements of a protein–protein interface has been widely applied for the inhibition of PPI. Some examples of the so-called “interface peptide” strategy include: peptide inhibitors of different adhesive proteins such as α–actinin and vinculin, short peptides that inhibit homodimerization of HIV-1 protease; a stabilized helicoidal peptide that inhibits a domain-domain interaction between the N-terminal and C-terminal domains of the HIV-1 envelope glycoprotein gp41, disabling thus membrane fusion between the virus and target cells; synthetic peptides that inhibit homo-dimerization of thymidylate synthase (TS); a β-sheet peptide that inhibits dimerization of the small E47 transcription factor; a synthetic cyclic heptapeptide that inhibits interaction between CD4 and major histocompatibility complex (MHC) class II proteins; synthetic peptides that block interaction between CD8 and MHC class I proteins; synthetic peptides that inhibit dimerization of the HIV reverse transcriptase; inhibition of HIV-1 protease homodimerization by a small tethered peptide; inhibition of the herpes simplex virus ribonucleotide reductase dimerization by a small hexapeptide resulting in a stronger effect on replication than the Acyclovir; and peptides targeting SH3-mediated PPI. Ferrer and colleagues used a combinatorial chemical library to find elements that, when covalently attached to a peptide derived from the outer layer α-helix, could inhibit gp41-mediated HIV-1 cell entry. Based on the X-ray structure of the inhibitor in complex with the HIV-1 gp41 trimeric core, they showed that blocking a small cavity was sufficient to inhibit the interaction between the core coiled-coil and the outer-layer α-helix of gp41. However, the small molecule alone had no inhibitory activity, although it increased the potency of the 30-mer mimetic peptide.

The discovery of peptide, peptidomimetic and small molecule inhibitors of the association between integrin α4β1 (VLA-4) and the endothelial surface protein vascular cell adhesion molecule (VCAM) was reported. A review of structure-based design of phosphopeptides and small molecule inhibitors of Grb2-SH2 mediated PPI has been published. A peptide sequestering the anti-apoptotic protein Bcl-2 has been optimized to increase its “mimicking” capability with respect to the BH3 domain of BID (a pro-apoptotic BH3-only protein) by hydrocarbon stapling. The resulting BH3 domain alpha-helix is more rigid, protease-resistant, cell permeable and binds with increased affinity to Bcl-2. This inhibitor suppresses the growth of human leukemia cells in vitro, and it prolongs the survival of leukemic mice in vivo. Furet and colleagues applied a structure-based approach to improve 1700-fold the binding affinity towards hdm2 of their initial peptide derived from the N-terminal domain of p53. They discovered potent antagonists of the p53-hdm2 interaction, which constitutes an attractive approach for cancer therapy. Several “two turns” structural mimics of the myosin light chain kinase present functional homology in its high affinity binding to calmodulin, and are able to inhibit the calmodulin activation of PDE enzyme in the nanomolar range.

Although design of peptide molecules that mimic protein–protein interfaces or antibody binding is an interesting approach, the ultimate goal is the design of small nonpeptidic PPI inhibitors (generally with MW < 500), more desirable for therapeutic use than peptides or peptidomimetics. Tilley and colleagues designed a series of acylphenylalanine derivatives intended to mimic the proposed binding region of interleukin-2 (IL-2) to the α subunit (IL-2Rα), based on a combination of structural information of IL-2 (by X-ray and NMR data) and site-directed mutagenesis. Structure-activity studies led to a small compound with an IC_{50} of 0.23 μM. Similarly, Sarabu and colleagues designed a series of small molecule antagonists for the interaction between interleukin-1 alpha protein (IL-1α) and the Type I receptor, with potential interest to treat inflammation related diseases. The design was based on the 3-D structure of the proposed binding epitope for IL-1β (derived from the X-ray structures of IL-1 ligands and site-directed mutagenesis data).

Nonpeptidic inhibitors of the interaction between fibrinogen and GPIIB-IIIa integrin, association that is essential for platelet aggregation, have been designed based on the tripeptide sequence Arg-Gly-Asp (RGD). Several
of these molecules (xemilofiban, orbofiban, sibrafiban, and lotrafiban; Figure 7) have progressed until phase III clinical trials but unfortunately they did not reach the market due to both a lack of efficacy and safety concerns. Other nonpep-tidic RGD mimics have been designed based on spirocyclic structures.

Based on the structure of the complex between the B domain (Fb) of Staphylococcus aureus protein A (SpA) and the Fc fragment of IgG (Figure 8a), Li and colleagues used computer-aided molecular modeling to design a molecule mimetic for protein A (Figure 8b) that is an effective competitive inhibitor for its interaction with IgG (Figure 8c).

Another interesting strategy for PPI inhibition is the use of transition metal complexes to target distinctive patterns of histidine residues on the surface of a protein. A review of rational design of PPI inhibitors involving the TNF family cytokines has been published. A different area of therapeutic interest involving PPI is the formation of amyloid fibrils. Klabunde and colleagues discovered small compounds that can inhibit transthyretin (TTR) fibril formation by stabilizing the native tetrameric conformation of TTR. They used a structure-based drug design approach based on the crystal structures of TTR complexed with known amyloid fibril inhibitors. Their work represents a good example of modulating PPI by enhancing stability of the complexed conformations avoiding unbound conformations that lead to disease.

Protein interfaces can be artificially re-engineered. A particularly difficult task is to break strong PPI in which two monomers are interlocked through extensive interactions and side-chain mutations are insufficient. Borchert and colleagues re-engineered the backbone of loop3 at the interface between two triose-phosphate isomerase monomers, which led to predicted monomeric structures. Engineered protein–protein interfaces, artificially disrupted after the introduction of cavities by using alanine-scanning mutagenesis, can be restored with small molecules bound to the cavity, thus generating artificial small molecule switches for PPI. Although rational design of the protein–protein interfaces themselves has limited therapeutic interest, it could be useful to understand the physicochemical basis of PPI modulation, and also to generate manipulated organisms in biotechnology that functionally respond to specific molecules.

Computer-aided design: virtual screening and docking simulations

Computational simulations is increasingly facilitating rational design of small molecules that can inhibit or stimulate the biological activity of specific proteins, mostly by targeting a clearly defined binding pocket. However, so far very few inhibitors of PPI have been designed using computer simulations (see recent reviews focused on virtual screening for the identification of inhibitors of PPI). Computational approaches have been successfully applied to optimize peptidic ligands in several systems. Zeng and colleagues used a combinatorial algorithm based on the MCSS approach for the optimization of

![Figure 7](https://www.dovepress.com/)

Figure 7 Some small molecule antagonists of GP Ib-IIIa: A) xemilofiban; B) orbofiban; C) sibrafiban; D) lotrafiban.
peptides that inhibit the association between Ras and Raf, proteins involved in signal transduction pathways and in many oncogenic events. Furet and colleagues optimized the inhibition properties of the phospho-tripeptide pTyr-Ile-Asn by molecular modeling and found a derivative capable of blocking the interaction between the activated tyrosine kinase growth factor receptors (TKGFR) and the SH2 domain of Grb2 (see a review of SH2 domain and drug discovery). Proline-rich peptides targeting SH3 domains were computationally optimized using the programs GRID and LUDI, obtaining an increment of 100-fold in affinity and 1000-fold in selectivity.
Fewer computational methods have been developed for rational design of small nonpeptidic compounds to inhibit PPI. Li and colleagues applied computer screening to select small nonpeptidic organic molecules that can inhibit interaction between CD4 and MHC class II proteins. Based on the X-ray structure of the human CD4 D1 domain, and using a combination of theoretical prediction and synthetic peptide experiments, the authors identified a surface pocket potentially involved in functional binding to MHC class II (Figure 9a). The identification of such a surface pocket was critical for the success of the strategy. The authors used the computer program DOCK3.5 to screen the available chemicals directory (ACD) (Molecular Design Limited, San Leandro, CA, USA), that included around 150,000 commercially available small organic compounds, in search for possible ligands to that particular pocket. They finally selected four compounds with significant inhibitory activity (45%–75% at 100 µM) for the CD4-MHC class II interaction (Figure 9b).

A novel class of low molecular weight hydantoins, which inhibits the interaction between the lymphocyte function-associated antigen-1 (LFA-1) and the intercellular adhesion molecule (ICAM-1) by allosteric regulation, represented an alternative example of PPI regulation. Based on an integrated immunochemical, chemical, and molecular modeling approach, the following allosteric inhibition mechanism was proposed: the hydantoins bind to LFA-1 and drive the equilibrium between active and inactive states of LFA-1 towards the conformation that is unable to interact with ICAM-1. Bushweller and colleagues found four new inhibitors that effectively blocked the interaction between Runx1 and CBFβ with low micromolar affinity, amongst 35 potential candidates selected by virtual screening. An NMR spectroscopy screening study showed later that none of these compounds were directly bound to the protein–protein interface, which suggested the existence of allosteric effects in the inhibition.

Virtual screening has been used to identify 13 nonpeptide drug-like inhibitors targeting the p56Lck SH2-domain from an initial screening of 25,000 compounds. Amongst the 13 inhibitors, two were identified as potential lead compounds for further development. In another example, virtual screening of 640,000 compounds was performed with DOCK4.0.1 in order to target the interaction between S100B and p53, which lead to the discovery of seven inhibitors in the micromolar range. Five of these compounds inhibited growth of primary malignant melanoma cells and are currently being optimized to find higher affinity inhibitors for potential applications in cancer therapy. The extracellular kinases ERK1/2, which play an important role in a signaling pathway involved in proliferation, are believed to be interesting targets to arrest cell proliferation in cancer. Only two proteins are known to turn on ERK1/2 kinases,
which then are able to phosphorylate dozens of proteins in vitro. Shapiro and colleagues applied a virtual screening approach to specifically target the ERK phosphorylation of two substrates: RSK-1 and ELK-1. The discovered compounds were able to inhibit the proliferation of several cancer cell lines in vitro. A recent study combined virtual fragment analysis and selection by molecular docking (using five different scoring functions) with an NMR-screening experiment called fluorine chemical shift anisotropy and exchange for screening (FAXS). The approach permitted the identification of a molecule displaying a strongly favorable binding enthalpy as tested by isothermal titration calorimetry (ITC), which suggested an enhanced selectivity for the v-src tyrosine kinase SH2 domain. Finally, computational modelling of the interaction nicely helped to explain the high binding enthalpy of this compound.

**Drug design and regulation of PPI: clinical applications**

Very few of the designed molecule inhibitors of PPI have been clinically tested. One example is a synthetic cyclic heptapeptide that inhibits interaction between CD4 and MHC class II proteins and that has been approved by the United States Food and Drug Administration for a phase I clinical trial in graft-versus-host disease (GVHD) prophylaxis in bone marrow transplant patients. Another example is a new thrombopoietic growth factor, eltrombopag (or SB-497115), which is actually in phase III clinical trials as an oral and nonpeptide thrombopoetin receptor agonist for the treatment of idiopathic thrombocytopenic purpura. Genetech (San Francisco, CA, USA) is also developing pertussis toxin (IAP) antagonists, a novel class of cancer therapeutics, and one of the molecules is now in phase I clinical trials.

Numerous factors can affect the output of an interaction network in a living organism. Some studies suggest that small changes in effector concentration can be more significant than absence or presence of a particular component, and the response can depend highly upon the biology of the system. The complexity of the response of the interaction networks in living organisms upon small changes in the environment makes the possibility of controlling signaling pathways with small compounds extremely challenging, although in the near future it will undoubtedly become one of the hottest areas in medicinal chemistry.

**Conclusions**

Targeting protein–protein interfaces with a small molecule is much more difficult than targeting a natural ligand pocket with another compound, due to the large and distributed set of interactions, the frequent lack of deep pockets, and the induced fit of the protein interfaces. A careful analysis of a protein–protein interface in search of putative small-molecule binding pockets, together with extensive computational protein-ligand docking simulations, will help to improve the rational design of PPI inhibitors. Computational prediction of “hot-spots” (for protein and ligand binding) at the surface of proteins can help to focus virtual screening or protein-ligand docking studies onto specific areas of a protein surface, and thus prioritize a large number of putative protein–protein interaction targets according to their potential to lead to a small molecule modulator. Finally, new improved protein–protein docking methods will be essential to predict the protein interfaces, and evaluate the PPI inhibition or oligomerization modulation capability of the selected compounds.

A combination of experimental and computational techniques, together with a deep knowledge of the determinants of protein–protein and protein-ligand interactions is necessary for the successful design of small compounds that can specifically modify PPI of therapeutic interest. The field is at its very early stage, but it constitutes a highly promising area of therapeutic proteomics.

**Disclosures**

The authors report no conflicts of interest in this work.

**References**


ADVANCES AND APPLICATIONS IN BIOINFORMATICS AND CHEMISTRY


Dovepress

**Advances and Applications in Bioinformatics and Chemistry**

**Publish your work in this journal**

Advances and Applications in Bioinformatics and Chemistry is an international, peer-reviewed open-access journal that publishes articles in the following fields: Computational biomodelling; Bioinformatics; Computational genomics; Molecular modelling; Protein structure modelling and structural genomics; Systems Biology; Computational Biochemistry; Computational Biophysics; Chemoinformatics and Drug Design; In silico ADME/Tox prediction. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: http://www.dovepress.com/advances-and-applications-in-bioinformatics-and-chemistry-journal