World Research Journal of Peptide and Protein
ISSN: 2278-4586 & E-ISSN: 2278-4608, Volume 1, Issue 1, 2012, pp.-09-20.
Available online at http://www.bioinfo.in/contents.php?id=131

FINDING INHIBITORS OF PROTEIN-PROTEIN INTERACTIONS (i-PPIs): A SUPPORT FROM BIOINFORMATICS

ANNA MARABOTTI* AND LUCIANO MILANESI

ITB-CNR, Segrate (MI), Italy.
*Corresponding Author: Email- anna.marabotti@itb.cnr.it

Received: January 17, 2012; Accepted: February 14, 2012

Abstract- Finding drugs with traditional approaches is becoming more and more difficult especially for the need of finding targets different from the traditional ones, to cope unmet clinical needs. However, the increasing amount of knowledge about cross-talks between proteins in the cell has prompted researchers to extend the range of "druggable" targets to include protein-protein interactions. Disruptors of protein complexes appear as a new paradigm in the field of drug discovery and development. This approach has long been considered as extremely challenging, and to date, only few successes were achieved with some molecules modulating protein-protein interactions that are currently on the market. Nevertheless, the number of inhibitors of protein-protein interactions in pre-clinical and clinical trials is increasing and this is encouraging for the future.
Bioinformatics represents a valid support for scientists, and several tools and software are available practically for every field of research. This paper reviews those bioinformatics tools to support people in developing drug-like molecules to target protein-protein interactions, at all levels of this process.
Keywords- Interactomics, protein-protein interactions, bioinformatics, drug discovery, fragment-based drug design, docking.
Abbreviations- i-PPIs: inhibitors of protein-protein interactions; 3D: three-dimensional; SVM: Support Vector Machine; MD: molecular dynamics; FBDD: fragment-based drug discovery.

Citation: Anna Marabotti and Luciano Milanesi (2012) Finding Inhibitors of Protein-Protein Interactions (i-PPIs): A Support from Bioinformatics. World Research Journal of Peptide and Protein, ISSN: 2278-4586 & E-ISSN: 2278-4608, Volume 1, Issue 1, pp.-09-20.

Introduction

At the end of the 2nd millennium, a new paradigm for science has been developed to look at complex systems in a holistic way. Starting from genomics, this new approach has been extended soon to transcriptomics and proteomics, and soon after, proteins became to be considered not only as individual entities, but as elements interacting with each others and with other macromolecules within a cell. The term "interactome" appeared in literature in 1999 [1], but a substantial increase in the interest for interactomics has been registered only in the last 5 years, with the publication of a high number of papers devoted to the study of interactomes in different conditions and organisms [2], including humans [3].

Simultaneously with the first examples of the basic research on large-scale protein-protein interactions, more applicative works appeared in the biomedical literature, focused on the use of this information to find molecules acting as inhibitors of protein-protein interactions (i-PPIs). The increase of interest for this class of compounds arose from the fact that the discovery of new drug targets is a constant challenge for pharmaceutical companies. In fact, despite remarkable efforts made to increase the chances of success [4], there remains a significant number of unmet clinical need for many diseases. Thus, there is the need to expand the range of potential targets using new molecules with mechanisms of action different from the "traditional" ones (that are essentially addressed to the active or binding sites of single enzymes or receptors, re-

spectively). i-PPIs appear then as a promising new class of potential therapeutic compounds. A very first example of work focused on the selection of new molecules able to disrupt crucial protein-protein interactions was published in 1997 [5]. During the last years, studies targeted on the discovery of promising drug candidates able to interfere in protein-protein interaction processes were intensified and some successful results were achieved (for reviews, see e.g. [6-12]). In particular, active fields of research were focused on: **i.** integrins, cell surface receptors that mediate a variety of functions involving cell-cell interaction and communication. Two molecules (epifibatide - a peptidic molecule; tirofiban - a non-peptidic molecule) that target the interaction between integrin αIIbβ3 and proteins such as vitronectin and fibronectin are the first i-PPIs that have been approved for clinical use, and are currently used in the prevention of platelet aggregation (reviewed in [7]); **ii.** disruptors of the interaction between p53 and its inhibitor MDM2, which binds p53 and negatively regulates its transcriptional activity and stability. These molecules represent an attractive new approach to cancer therapy, since many tumors overproduce MDM2 to impair p53 function (reviewed in [13]); **iii.** inhibitors of the interaction between IL-2 and its receptor IL-2R, that have been investigated as potential drugs for a range of immune-cell disorders (reviewed in [14]); **iv.** inhibitors of the interactions of antiapoptotic factors Bcl-2 and Bcl-XL with other members of the family of critical mediators of apoptosis. These inhibitors promote cell death in specific pathological conditions such as cancer (reviewed in [15]); **v.** antiinfective drugs. An i-PPI, maraviroc, targets the interaction between the viral protein gp120 of HIV and the cell receptor CD4, thus impairing the membrane fusion and the entry of viral particles into the cell. This i-PPI was approved for human use by FDA (reviewed in [16]). Another example in this field is the development of i-PPIs against human papilloma virus, targeting the interaction between viral proteins E1 and E2, crucial for viral genome replication (reviewed in [17]); **vi.** modulators of the interactions between the transcription factor c-Myc (involved in fundamental cellular processes such as cell cycle progression, growth, and oncogenic transformation, as well as apoptosis) and its heterodimerizing partner Max, whose interaction is crucial for the vast majority of c-Myc functions (reviewed in [18]).

Despite these and other efforts, the percentage of success in finding i-PPIs that can actually be marketed as drugs in the near future is still extremely low. The main difficulties arise from several factors. Most of the protein-protein surfaces are large and flat, therefore the interaction with a small molecule requires a high number of weak interactions widely spaced. On the contrary, generally the interaction between a "traditional" druggable binding site and a small molecule is driven by a limited number of strong interactions. Protein-protein interaction surfaces are more likely to be topologically complex, whereas the binding sites in proteins traditionally targeted by drugs are simple-shaped pockets. Post-translational events such as phosphorylations regulate most of the protein-protein interactions, but a protein site that has evolved to interact with a phosphorylated partner is not likely to be an attractive target, since any small molecule mimicking the charged species would be probably too polar to show good bioavailability. Finally, another issue concerns the specificity of the interaction, that is more difficult to achieve for i-PPIs [7,9]. During the time, most of these problems have been overcome. However, it is undeniable that standard approaches such as high-throughput screenings are uneffective to reach a high percentage of success if focused on the search for i-PPIs [9]. Therefore, the development of alternative strategies has soon become a main goal to tackle this problem. The knowledge of the three-dimensional (3D) structure of the protein-protein interaction site is obviously a great advantage, providing valuable insight into protein-ligand interactions, thus accelerating the iterative process of inhibitor design. Most of the successful projects have exploited this information (see e.g. [19-23]). Over the years, there has been a progressive rationalization of the research to enable the discovery of i-PPIs in a quick, efficient and money-saving way, and a relevant role has started to be played by computer-aided drug design approaches [11].

Bioinformatics has played a main role since the outset of interactomics, providing scientists with essential tools for interpretation, storage and data analysis, in parallel with the massive application of experimental techniques [24]. For example, the first databases to collect data on known molecular interactions were developed very early, even before the term "interactome" was coined [25], and to date, almost 100 resources are listed in the "Protein-protein interactions" subcategory of the 2012 NAR Database Issue: http://www.oxfordjournals.org/nar/database/subcat/6/26.

Other tools were developed for visualizing molecular interaction networks and biological pathways, and to integrate these networks with annotations, gene expression profiles and other data, allowing users to manage this plethora of information in an easier and immediate way. The most popular of these tools is probably Cytoscape (http://www.cytoscape.org), an open source platform that is continuously developing thanks to a community-based effort [26].

The next goal for bioinformatics applied to the investigation of protein-protein interactions was the development of computational tools for the prediction of protein-protein interactions. The first computational methods were based on sequence and genomic information, such as the presence or absence of genes in related species, the conservation of gene neighborhood, analysis of gene fusion events, the evaluation of the similarity of phylogenetic trees and *in silico* two-hybrid methods [27]. Then, machine-learning approaches such as Support Vector Machines (SVM) or Random Forest methods were introduced [28]. Since the knowledge of the 3D structure of the proteins adds further information, several structure-based predictive methods were then developed [29-31]. Several challenges are still to be faced, despite notable progress during these last years [32]. However, there is a continuously increasing interest in this field, as inferred by the fact that in the last 3D-SIG (Structural Bioinformatics and Computational Biophysics Meeting, a satellite ISMB/ECCB Meeting, held in Vienna on July 2011), two out of eight sessions were devoted to computational approaches to study protein-protein interactions.

During last years, computational biology has rapidly provided many tools to help researchers to develop i-PPIs, and it has become essential for an efficient and successful project. The purpose of this review is to give the reader an overview of the various computational strategies that have evolved over time to address and, hopefully, solve this problem. Since the focus of this review is based on drug-like molecules acting as i-PPIs, computational tools to design peptides and peptidomimetic able to modulate protein-protein interactions will not be included here. However, the reader is referred to some excellent and recent reviews covering these

topics [33,34]. A schematic illustration of the different computational approaches to find i-PPIs is shown in Fig. (**1**).

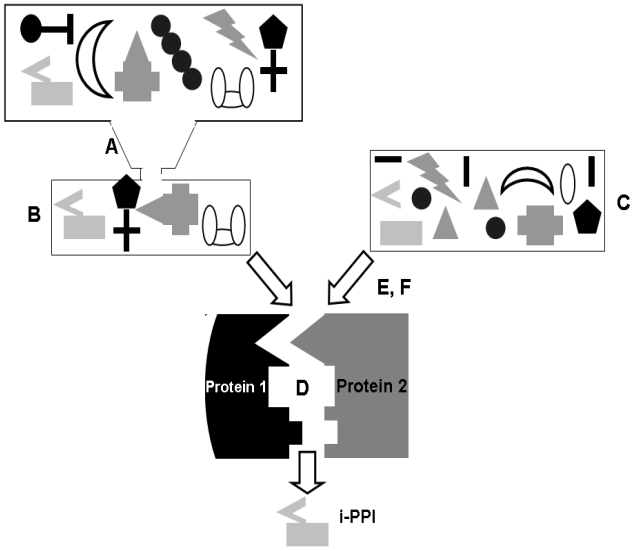

Fig.1- Scheme of the different computational approaches to find i-PPIs described in the text and of their role in the discovery steps. Starting from a library of compounds, computational descriptors of chemical space (A) can help in filtering the more promising putative i-PPIs using detailed analysis of the chemical and structural features of these molecules. This allows to create a library enriched in putative i-PPIs (B), thus increasing the chances of success for high-throughput screenings. Fragment-based drug design (C) is another approach by which promising fragments are bound together to form putative i-PPIs. Tools for the analysis of protein-protein interfaces (D) allow to access to structural information about protein-protein interfaces, that can be exploited for the development of new i-PPIs. Finally, docking (E) and, whenever possible, molecular dynamics simulations (F) allow to identify molecules that would bind the protein interface, to select the most promising i-PPI to disrupt that interaction.

Computational strategies to find i-PPIs

Computational descriptors of the chemical space of i-PPIs and libraries enriched in i-PPIs

When the data about the first i-PPIs started to be available in literature, it became evident that the chemical and structural features of these molecules are quite different from those of the traditional drug-like compounds. Due to the fact that they bind to regions very different from classic pockets or active sites, i-PPIs are generally larger in size than enzyme inhibitors, ion channel modulators, or receptor ligands [35]. Moreover, given that several studies have underlined the abundance of aromatic and/or hydrophobic residues at the protein-protein interfaces, most of the i-PPIs contain several aromatic and/or hydrophobic moieties to establish favourable interactions with their macromolecular targets [36]. A number of following studies tried to identify less empirical descriptors of the physico-chemical features of known i-PPIs, in order to compare them with the chemical spaces covered by the existing libraries of chemicals, with the aid of computational tools (listed in Table 1, part **a**). Probably, one of the earliest investigations in this sense was the one performed by Pagliaro and coworkers. In that work, a principal component analysis was performed on the basis of physico-chemical properties of the molecules calculated with MOE, a package for computational chemistry, cheminformatics and bioinformatics (Chemical Computing Group Inc. Montreal, Canada, http://www.chemcomp.com). The analysis showed that only 50% of the 19 i-PPIs available to date were covered by the diversity chemical space of three popular commercial databases [37]. Therefore, one of the first strategies suggested to improve the rate of success of screening procedures focused on the search for promising i-PPIs was the creation of libraries enriched in i-PPIs [37]. Computational chemistry techniques, allowing a detailed analysis of the chemical and structural features of these molecules, were used to help researchers in selecting the most promising compounds.

In 2007, Neugebauer and colleagues used chemoinformatics and artificial intelligence methods to retrieve information from a collection of known i-PPIs and to develop an algorithm to discriminate possible i-PPIs among a group of compounds [38]. Starting from 25 small molecules extracted from the literature and used as training set (excluding peptides and small proteins acting as i-PPIs), they identified descriptors for these compounds using a software for molecular descriptor calculations (Dragon, http://www.talete.mi.it) [39]. Then, they built a decision tree using the three most relevant descriptors (related to molecular shape, number of ester functions in molecule and 3D structure of the chemical compound), to discriminate, among different molecules, those that are more likely to be i-PPIs. The final algorithm showed a good classification power towards the training set with respect to the background (1137 non-i-PPIs compounds extracted from FDA approved drug database), and a high predictive power. Using this algorithm onto another very popular dataset, ZINC [40], 185 compounds were predicted to be potential i-PPIs.

Another attempt to find rules to better define the chemical space of i-PPIs to improve screening approaches was made very recently [36], substantially with the same approach as the one just described. 145 experimentally validated i-PPIs and 4857 existing drugs taken from the small subset of DrugBank database, filtered to obtain a final set composed of 66 diverse drug-like i-PPIs and 557 traditional drugs with an improved chemical diversity, were evaluated again with Dragon. 1666 different chemical descriptors were evaluated to obtain as much information as possible to characterize the two subsets and to allow their discrimination. Using those descriptors that separate the two subsets significantly (P-values <0.05) and that are uncorrelated to each others (one related to molecular shape, as in [38], whereas the other one related to the presence of multiple or aromatic bonds), two decision trees were built up and used in succession. The application of the two models on the MayBridge Screening collection (57,200 compounds) and on the diversity set of the ChemBridge database (50,000 compounds) allowed to select, respectively, 13,799 and 9,622 i-PPIs-like molecules. In a further work of the same group, the best model has been transposed into a computer program, PPI-HitProfiler (freely downloadable from http://www.CDithem.com), that is able to build, from any drug-like compound collection, a chemical library enriched in putative i-PPIs [41].

The first examples of libraries dedicated to i-PPIs have been recently created (Table 1, part **a**). The first database fully dedicated

to i-PPIs, TIMBAL (http://www-cryst.bioc.cam.ac.uk/timbal), was published in 2009 by Higueruelo and colleagues [42]. It is a fully hand-curated relational database holding a collection of i-PPIs retrieved from literature. At the date of publication, the database contained information about 104 molecules disrupting 17 protein-protein complexes, retrieved from 40 papers. The comparison of the properties of these collected i-PPIs with those of ligands retrieved from 3D structures of complexes collected in CREDO database [43], or with those of commercialized drugs extracted from MDDR database, or with those of compounds randomly selected from the catalogues of three different suppliers (Enamine, Asinex and Maybridge) confirmed that in general i-PPIs are larger and more hydrophobic, with more rings and less rotatable bonds than drugs and ligands from the PDB. Moreover, they have on average a smaller number of hydrogen bond donors and acceptors than the drugs set. The analysis of their Ligand Efficiency (LE), a parameter related to the free energy of binding per heavy atom [44] confirmed a previous observation [9] : i-PPIs show a slightly lower LE than that of typical medicinal chemistry leads with the same number of atoms.

One year after, the database 2P2I was published. This is a hand-curated structural database collecting information about the protein-protein interfaces for which both the protein-protein and protein-inhibitor complexes have been structurally characterized [45]. This allows users to extract the best descriptors of protein-protein interactions with a known inhibitor. The dataset was built through data mining from the literature and by exhaustive search of the Protein Data Bank [46], and it was finally compiled into a relational database that was used to further analyze the general properties of protein-protein interfaces with a known inhibitor, using several computational tools. A web interface at the address http://2p2idb.cnrs-mrs.fr was then developed to facilitate the access to the data calculated for the different i-PPIs. In 2010, 17 protein-protein complexes corresponding to 14 families and 56 small molecule inhibitors bound to the corresponding target were collected in the database. The limited number of targets was due to the structural prerequisites that were applied. The same group published very recently a review in which they propose, in contrast with the well-known Lipinski's "Rule-of-Five" [47], a "Rule-of-Four" to define the generic profile of a i-PPI (MW >400 Da, ALogP >4, number of rings >4 and number of hydrogen bond acceptors >4) [48]. Obiously, this means that these compounds will have difficulties in the following phases of drug optimization, and that there is an urgent need for pharmaceutical companies interested in the development of i-PPIs using parallel technologies for optimal drug delivery, such as nanoparticles delivery systems.

In conclusion, these computational tools have allowed to create databases more focused on i-PPIs and more precise rules to describe them in terms of physico-chemical features, thus increasing the chances of success for traditional approaches such as high-throughput screening.

Tools for the analysis of protein-protein interfaces

The increasing availability of structures of protein complexes solved by experimental methods and the increasing accuracy of methods to predict the structure of protein-protein complexes [49] has allowed an extended access to structural information about protein-protein interfaces, that can be exploited for the development of new i-PPIs. This information is extremely important and, when possible, a detailed analysis of the interaction surfaces should be made in order to increase the percentage of success in developing i-PPIs. It has been already stressed in the Introduction that protein-protein interfaces are quite different from the classical pockets targeted by traditional drug-like molecules, and that not all of them are suitable to bind molecules acting as disruptors of macromolecular interactions. During the time, computational biology and bioinformatics have provided tools to estimate the "druggability" of the protein interfaces, characterizing their conformational features [50]. They will be discussed in the following part. A summary of these tools is listed in Table 1, part **b**.

The first condition to identify a suitable binding site for i-PPIs is the presence of a pocket on the protein's surface. In fact, it is more difficult to obtain potent inhibitors for flat interfaces than for interfaces which contain well-defined cavities [8]. Unfortunately, the protein-protein complexes most attractive for drug discovery (i.e. those formed transiently between different proteins) have rather flat interfaces. Therefore, the presence of cavities or pockets at the contact region should be searched very carefully during the evaluation of a protein-protein interaction. Several computational tools are currently available to find pockets on protein surface, based on different criteria [51]. Two main types of tools were developed: those based on evolutionary algorithms and those that use structure-based algorithms. This last category can be subdivided in geometry- and energy-based algorithms [52]. Some popular tools, most of which also have a Web interface for the analysis are: SURFNET [53], LIGSITE [54], CASTp [55], AVP [56], Q-SiteFinder [57], PocketPicker [58], AutoLigand [59], fpocket [60] and SiteMap [61]. Some further evolutions of these methods, such as LIGSITEcsc [62] and SURFNET-ConSurf [63], take into account also the degree of conservation of the residues in the pocket. A very recent tool, MDpocket, has been developed to detect small molecule binding sites and gas migration pathways on conformational ensembles such as molecular dynamics (MD) trajectories [64]. Also meta-methods that combine results from several methods have been implemented recently [65,66].

The presence of a pocket on the protein-protein interface is a necessary but not sufficient condition for the development of drug-like small molecules disrupting this interaction [67]. For example, it is more difficult to generate potent competitive i-PPIs if the two interacting chains are closely packed and make an extensive number of direct interactions [8]. In addition, the role of water cannot be ignored. Protein interfaces rich in cavities contain a large number of water molecules that are involved in bridging H-bonds between the two macromolecular chains. This should be taken into account, since the displacement of these bound water molecules by the inhibitor could provoke a favorable entropic effect that might enhance its affinity [8]. Several tools have been developed that provide a more complete analysis of protein-protein interfaces. For example, the very popular web server PDBePISA [68], provided by the EBI (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) is an interactive tool for the exploration of macromolecular interfaces that allows the investigation of the structural and chemical properties of macromolecular surfaces and/or interfaces, the probable quaternary structure of the complex and the probable dissociation pattern, either for pre-calculated or interactively calculated structures. PIC (http://

pic.mbu.iisc.ernet.in/) is another Web server which recognizes various kinds of interactions within a protein or between proteins in a complex [69]. PROTORP (http://bioinformatics.sussex.ac.uk/protorp/) is a server that calculates interaction properties from 3D structures of individual proteins of interest and for entire datasets in real time [70]. This server provides an efficient way of characterizing protein-protein associations of new or existing proteins, and a mean of putting these values in the context of previously observed associations.

To help in considering the suitability of a protein-protein interface for drug development on the basis of its structural and physico-chemical properties, an empirical decision tree was proposed that takes into account parameters such as the presence of cavities, their hydrophobicity and size, the shape complementarity between the two interacting subunits within the cavity [8]. A more formal mathematical model to evaluate the "druggability" of a binding site based only on its physico-chemical properties was then developed [67]. With this method, the druggability of a set of 27 target binding sites was calculated and, despite the simplicity of the model, surprisingly good discrimination was achieved between non druggable and easily druggable protein-protein interfaces. A possible use of this model could be the prioritization of the choice of targets suitable for drug discovery efforts [67].

In addition to the generic physico-chemical properties, the study of protein-protein interactions have highlighted the presence of some regions of the interfaces that are more important than others for protein recognition and binding. These regions are called "hot spots" [71] and a large part of the interaction energy is concentrated there. In addition, hot spots are smaller than the full interface, and often there are well-defined pockets near them. For all these reasons, it might be easier to identify low-molecular-weight compounds binding to the hot spots that inhibit protein-protein interaction and therefore the identification of hot spots is a good starting point for drug development [72]. Useful experimental data to identify hot spots often comes from mutational studies, in particular alanine scanning mutagenesis studies [73]. Mutations may cause significant changes in the affinity of protein-protein interactions and thus help in identifying important residues, although it should be stressed that only if the mutation influences the conformational ensemble of the complex, measured binding free energy differences between mutant and wild type protein can be related to specific contact differences [50]. The increasing importance of these data has allowed the development of databases to collect many results of alanine scanning experiments and to make them available through the Web. An example is the ASEdb database, a searchable database of binding energy changes reulting from mutations of protein side chains to alanine, which is freely available at the address: http://nic.ucsf.edu/asedb/ [74]. In addition, given the growing interest in hot spot determination and prediction, an increasing number of computational approaches have been developed to predict the presence of hot spots in protein-protein interfaces. The first one developed was an *in silico* alanine scanning approach, which used free energy functions (including van der Waals potentials, electrostatic interactions, hydrogen bonds, and desolvation energy) to calculate the change of binding free energy when alanine replaces (virtually) the residues of the hot spot region. This approach was performed using MD simulations associated to free energy perturbation and thermodynamic cycles, combining explicit molecular mechanical energies and continuum solvation models for calculating interaction free energies [75]. Methods based on evaluations of binding free energy were further developed, also with the aim of reducing the high computational costs of these procedures [76-81]. Other methods were based on component analysis: the contributions of molecular mechanical energies and solvation free energies are assigned to those atoms that participate in the respective interaction, and these contributions are then summed to produce the total binding free energy of the residue [82]. These methods, however, have to deal with the fact that the total binding free energy is a state function, but their free energy components in general are not, and their determination is affected by the decomposition scheme chosen. Another alternative topological approach was developed representing protein-protein complexes as small-world networks [83].

It was commonly thought that there was no general pattern of hydrophobicity, shape or charge that can be used as a basis for predicting which atoms of the protein will participate in hot spots [73]. Nevertheless, in more recent years, feature-based prediction methods have been developed thanks to the increasing amount of experimental information on hot spot regions. Most of them are based on machine-learning algorithms such as SVM, decision trees and so on [84-91]. Other methods implement also evolutionary information based on sequence conservation [92-94]. The increasing amount of information available has also prompted to develop databases of predicted and experimentally defined hot spots [95-96].

When the structure of the protein-protein complex is not available, detecting the clefts in unbound protein interfaces by computational tools would provide starting points for the further rational design of i-PPIs. The most direct computational approach to deal with the conformational plasticity of proteins is MD simulations [97]. Several studies showed that even short MD simulations of proteins in their unbound state were able to sample the conformational fluctuations of key residues at the interfaces in the bound state [98-100]. Moreover, MD simulations should be able to discover the presence of "cryptic" binding sites that are not present in the unbound form of the interactors [101,102]. For example, MD simulations were able to reveal the opening of the cryptic sites at the complex interface between IL-2 and IL-2R, opening the way to the finding of i-PPIs inhibitors for this interaction [103], as described below in more details. Although this suggests that predicting conformations of protein-protein interactions suitable for the binding of i-PPIs using MD simulations might be feasible, the routinely application of this computational technique is still far from reality, essentially because of the high computational demands. In addition, the force fields used require further refinement to take into account phenomena in which quantum effects are important [97]. An alternative to MD simulations is normal mode analysis (NMA) [104], in which an analytical solution to the equations of motions yields collective variables (normal modes) that describe the dynamics of the system. This method is particularly interesting in view of predicting bound protein states from unbound ones, and it can be applied to identify potential conformational changes of proteins upon binding.

Docking

Once the binding site(s) of a protein-protein complex are known,

the main computational biology approach to identify molecules that would bind those binding site(s) is essentially the docking. This methodology predicts the preferred reciprocal orientation of one molecule to a second one when they are bound together to form a stable complex, and may be used to predict the strength of association or the binding affinity between two molecules, using for example scoring functions somehow related to the binding energy. In general, the screening of large *in silico* databases is divided into two steps: a first one in which a coarse-grained quick selection of compounds is made at the expense of accuracy, and a second step in which selected compounds are docked with a higher level of accuracy. In both steps the ligand is generally treated as flexible, whereas typically the protein structure is kept rigid in the first step to reduce computational costs, and some degree of flexibility is possibly introduced during the second step. Docking is a widely used procedure in drug discovery, and the procedures applied to dock i-PPIs are in principle not dissimilar from those used to dock small molecules into "classical" binding sites, such as enzyme active sites. Therefore, an extensive description of docking programs is not the focus of the present review (the reader is referred to several complete and recent reviews that already exist on this subject, see for example [105-107]). Rather, the specific problems that are encountered when docking is applied to the discovery of i-PPIs are summarized here. The first problem is related to the fact that often the contribution of solvent in docking calculation is ignored [105]. It should be pointed out that the particularities of the interaction surfaces, stressed previously, do not allow to introduce much simplifications in the evaluation of the binding interactions between a small molecule and a protein-protein interface. In particular, solvent contribution plays a much greater role in these cases than in the interaction between a ligand and a pocket deeply buried in the protein core. Therefore, to treat correctly these simulations, it would be necessary to deal with explicit solvent [50]. However, this would result in a computationally expensive simulation. Alternative approaches to treat the solvent implicitly using continuum or macroscopic models in which solvent properties are described in terms of average values [108] may also be applied. Continuum methods that involve solution of the Poisson-Boltzmann equation [109] are more accurate, but with higher computational costs. Implicit solvent models such as the Generalized Born/surface area method allow to obtain a good accuracy in binding energy estimate saving computational costs [110].

Another great challenge to face when docking is applied to the discovery of ligands at protein-protein interfaces is the protein flexibility. In fact, side chains are on average more flexible at the protein's surface or at protein-protein interfaces than in traditional binding pockets within the core of the proteins [111]. A comprehensive review on methods to deal with flexibility in docking has been published recently [112], and these methods could be applied for the docking of i-PPIs. Many methods address this problem either using multiple receptor representations for docking, or performing a conformational search within the space of discrete side chain conformers [113]. This last approach is typically the most expensive one in terms of computational resources, because of the change of the coordinates of wide parts on the receptor. Moreover, in general such methods allow for limited adjustments of the side chains and backbone conformation in the proximity of the binding pocket, but they do not consider large scale conformational changes that can occur upon ligand binding [105].

Despite these problems, it is expected that docking procedures that take into account these considerations, with an optimal balance between computational requirements and accuracy in the description of the protein-protein interaction surface would be able to greatly increase the percentage of success in finding promising i-PPIs.

i-PPIs developed using fragment-based drug discovery approach.

Fragment-based drug discovery (FBDD) is an approach to identify lead compounds in drug discovery that allows to probe a large chemical space and to generate molecules with high ligand efficiency [114-117]. Fragments are organic molecules smaller than traditional leads identified by high-throughput screening procedures, and their physico-chemical features can be summarized by an empirical "rule of three" (MW <300, CLogP <3 and number of hydrogen bond donors and acceptors <3) [118]. Fragments can be considered as "building blocks" for more complex molecules: they bind to their target with lower affinity compared to drug-like molecules, but they can identify subpockets to which they bind with higher efficiency. When they are optimized for their potency (usually taking advantage of structural information about their binding mode and/or connecting more fragments to each others), new molecules are generated that retain high ligand efficiency, thanks to a number of high-quality interactions with the key pockets of the protein. Moreover, the creation of a lead compound by FBDD potentially enables the assembly only of small molecules that explore and capture available surface features for high-affinity binding, removing poor binders early in the discovery process. Such optimized lead compounds are more likely to have improved pharmacokinetic properties than those obtained by traditional approaches [114]. In addition, with FBDD strategies smaller libraries of fragments can be screened while maintaining the same probability of sampling a large chemical space [119,120]. One of the first examples of application of such strategy was published in a pioneering work in 1996, in which researchers used a technique called "SAR by NMR" consisting in the use of ^{1}H-^{15}N heteronuclear single quantum correlation NMR as a screening tool for identifying fragments that bind to a site of interest on a protein. A combination of structure-based design and structure-activity relationship (SAR) analysis was then used to link the fragments together [121]. Other experimental methods to provide additional information about binding site and/or binding stoichiometry of fragments are X-ray crystallography [122,123], surface plasmon resonance [124] and mass spectrometry-based methods [125]. In particular, a method called "Tethering" developed at Sunesis Pharmaceuticals [126] relies on the rapid identification by mass spectrometry of fragments with the greatest affinity for a binding site of a protein through the detection of the formation of a stable disulfide bond between the fragment and a cysteine residue in the target binding site.

In the last years, also computational methods have gained importance in this field, as shown afterwards. Extensive reviews of the recent advancements in both computational and experimental approaches are available in literature [127-129], as well as sever-

al examples of studies that apply FBDD for the discovery of new leads [115,130,131]. At present, FBDD has gained popularity in both academic and industrial worlds, and now it is often conducted in parallel with high-throughput screenings [129].

The fragment-based approach seems to be especially appropriate to find i-PPIs because it is particularly suited for binding sites with multiple and nearby subsites, such as protein-protein interfaces [131]. A successful example of a fragment-based approach for the development of an i-PPI that reached clinical trials is that of small organic molecules developed at Abbott Laboratories that bind to the hydrophobic helical domain of Bcl-XL, Bcl-2 and another anti-apoptotic molecule, Bcl-W. Using a SAR-by-NMR approach, two weak binding fragments were identified and subsequently improved by fragment linking, parallel synthesis and structure-based design, to obtain ABT-737 [132], which was found to bind to Bcl-2 with nanomolar affinity [133,134]. However, this molecule is not orally available and showed low aqueous solubility. Further optimizations of multiple individual sites of the molecule gave additive effects, resulting in ABT-263 [135,136], that is orally available and has entered several clinical trials to test its efficacy in chronic lymphocytic leukemia, lymphoma, small-cell lung cancer, and solid tumor, alone or in combination with other anticancer agents [15].

Another noteworthy example of the application of FBDD approach for the development of i-PPIs was the discovery of inhibitors of the interaction between IL-2 and its receptor IL-2R. After the design of compound Ro26-4550 [19], and the structural characterization of its interaction with the complex IL-2/IL-2R [137,138], a fragment-based approach was used to develop new compounds [139-142], the best of which showed an improved IC50 of 60 nM, compared to the original IC50 of 3 μM [141]. Computational simulations of IL-2 have provided important clues toward understanding the ability of Ro26-4550 and fragments to bind synergistically [103]. In particular, MD simulations revealed highly correlated movements of side chains and backbone that form tightly coupled networks, and that would induce the opening of the cryptic sites at the complex interface [14].

Given the increasing importance of the fragment-based approach in drug discovery in general, and in particular in finding i-PPIs, several computational tools and methods have been developed to support this strategy (Table 1, part **c**). Some computational tools can support the experimental phases of this approach. An example in this case is a platform developed by Astex Technology, called “Pyramid”, in which libraries of fragments enriched by virtual screening procedures for fragments likely to bind the target are screened using X-ray crystallography. Then, electron densities are analysed with automated procedure using a software called AutoSolve [123], and fragments identified to bind the target are subsequently optimized [133].

In the last years, there has been an exponential increase in purely computational methods to support FBDD. They are applied in all stages of the discovery process, going from the creation of fragment libraries, to the selection of fragments to be tested experimentally, and to the evolution of fragments into potent compounds [144-147].

As illustrated previously for libraries enriched in i-PPIs, a good fragment library can be obtained by filtering commercial libraries using the above-mentioned “Rule-of-three” or other filters specific for fragments. Alternatively, libraries are created based on diversity-driven fragment selection, to maximise the fraction of chemical space covered by the components, or starting from the deconstruction of molecules into fragments. Diversity enhancements algorithms were developed independently from the needs of FBDD, but were successfully applied also to this field. They are typically based on clustering techniques, which group chemical structures using several criteria, and pick one or more representative molecule from each cluster. The two main approaches for clustering are the similarity-based and the grid-based strategies [148]. In the first case, descriptors reporting the presence or absence of specific atoms or chemical groups are used to characterise the structures, and fragments considered to be in the same chemical space are grouped together [149]. Also descriptors that take into account the 3D structure of a molecule can be employed, but it seems that they do not offer significant advantages over the other descriptors [150,151]. Statistical functions such as the Tanimoto coefficient are used to perform the clustering.

A grid-based strategy divides the chemical space into a grid of distinct cells with the positions of the molecules in those cells determined by their physico-chemical features (e.g., charge, polarizability, H-bond capability, etc) [152]. Grid-based approaches are much more intuitive and visually informative than the similarity-based ones, and they are also better suited for an easy identification of chemical regions not covered in a particular library [128]. Several tools have been developed to implement the above-mentioned rules in order to automatically design fragment libraries. For examples, a number of computational tools were developed in house at Astra Zeneca from 1993 to 1999 [153].

The alternative approach for building fragment-based libraries is to break bigger molecules into fragments and to collect them. This approach is especially useful to create a focused fragment library for a particular target when known inhibitors are already available. An example of this approach is the SHAPES strategy applied at Vertex Pharmaceuticals [154]. Some popular computational methods, such as RECAP [155], DAIM which has been described more recently [156], and a workflow integrated in the Pipeline Pilot package, have been developed to build collections of privileged fragments [157], breaking drug-like chemical compounds and applying then several filters to the resulting fragments. However, fragment deconstruction has not proven to be the most fruitful pathway for a FBDD approach [158]. Several examples of designed fragment libraries are available in the literature [129].

Once the fragment library has been obtained, in addition to experimental strategies, also computational approaches can be used to test their affinity within the macromolecule of target. Docking and virtual screening procedures (see above) were applied also to FBDD. The low complexity of fragments and other pitfalls linked to the low reliability of the scoring functions used in fragment docking (caused by the fact that these functions have historically been optimised for drug-like compounds and not for fragments), has however highlighted the limitations of state-of-the-art docking programs in FBDD. Only very recently, approaches tailored for fragment docking have been developed [159], and programs such as Glide have implemented protocols for fragment docking [159,160]. In addition, the group of A. Caflisch has created several computational tools dedicated to high-throughput screening by fragment-based docking of large collections of small molecules

[161] that are freely available at http://www.biochem-caflisch.uzh.ch. Another very recent tool that performs the search of fragments into a database, and their docking has been developed in the group of McCammon [162].

Once the fragments that bind to the macromolecule(s) have been identified, there are several strategies to evolve them in a drug-like molecule. A possible strategy is to create chemical analogues with new groups that could potentially improve the affinity towards the targets. Otherwise, the "fragment-linking" approach consists in binding together fragments that show individual high affinity for different cavities or pockets of the target macromolecule. An interesting computational strategy, called 3D-RISM [163] has been developed to identify the most probable positions and orientations of fragments on a protein surface, using 3D spatial distribution functions of the atomic sites of the ligand, calculated using the molecular theory of solvation. Other programs specialized in fragment-linking strategies are GANDI [164] and CONFIRM [165]. The first one uses an island-based genetic algorithm to sample linkers in order to join pre-docked fragments; the second one screens a library of "bridges" to find suitable linkers between pre-docked fragments that have been selected by Glide. Linked candidates are selected on the basis of their strain energy and of the deviation of the original fragments between their initial position and their position in the linked compound.

The "fragment growing" strategy is an alternative strategy to evolve fragment into drug-like compounds by adding chemical complexity to the structure. Several programs have been proposed to aid in this process, such as LUDI [166], SPROUT [167] SMoG [168], LigBuilder [169], and AutoGrow [170]. These tools can be applied also in the *de novo* design strategy [171-173]. A *de novo* design workflow developed within a collaboration between medicinal and computational chemists, and software developers was created recently combining a suite of already published programs [174]. Another strategy is the "fragment shuffling" for the identification of novel lead compounds combining central elements from fragment-based lead identification and structure-based *de novo* design. An automated workflow was recently developed to exploit calculated ligand fragment data sets derived from aligned and scored protein-ligand complex structures by recombining these fragments to novel molecules [175].

When examples of ligands for selected targets are known, a general strategy to develop new ligands is the pharmacophore approach, that can be applied also in the context of FBDD [176], essentially to enrich fragment libraries with chemotypes of interest or to analyze their complexity/diversity. For example, Gozalbes and colleagues developed an approach in which several pharmacophore hypotheses were developed based on the alignment of the most active inhibitors of heparanase known to date. The selected pharmacophore model, once validated, was applied virtually to a database of around 700 chemical fragments [177].

In conclusion, FBDD are new reliable methods that can be applied in drug discovery, and that are promising in the field of discovery of i-PPIs. Computational methods can be applied at different levels in this field, although the peculiarity of fragments highlights the need for more specialized tools for these small molecules.

Conclusions

Small-molecules i-PPIs are an emerging family in the scenario of future drugs, and their development is seen as a new frontier that, thanks to the increasing knowledge about protein-protein interactions and their meaning in the most important processes of cell life and metabolism, can be surely of outmost importance for the generation of new therapeutic approaches. The application of bioinformatics and computational biology to this field has allowed to adopt a more rational approach for the finding of new molecules of this class, reducing the costs, saving the time and increasing the probabilities of success in the drug development pipeline. This is a rather new approach and therefore these progress are not immediately visible to the scientific community; however, it is likely that the advantages will become more and more evident in the near future, and we will benefit of new and more effective drugs to treat diseases.

Acknowledgements

We would like to thank Dr. Angelo Facchiano, of the Institute of Food Science, CNR, Avellino, for useful suggestions and comments about the review. The authors are supported by: FIRB ITALBIONET project (RBPR05ZK2Z and RBIN064YAT_003) and the Flagship Project InterOmics.

References

[1] Sanchez C., Lachaize C., Janody F., Bellon B., Röder L., Euzenat J., Rechenmann F., Jacq B. (1999) *Nucleic Acids Res*, 27(1), 89-94.
[2] Devos D., Russell R.B. (2007) *Curr. Opin. Struct. Biol.*, 17(3), 370-377.
[3] Figeys D. (2008) *Cell Res.*, 18(7), 716-724.
[4] Overington J.P., Al-Lazikani B., Hopkins A.L. (2006) *Nat Rev Drug Discov.*, 5(12), 993-996.
[5] Huang J., Schreiber S.L. (1997) *Natl. Acad. Sci.*, 94(25), 13396-13401.
[6] Arkin M.R., Wells J.A. (2004) *Nat. Rev. Drug. Discov.*, 3(4), 301-317.
[7] Fry D.C. (2006) *Biopolymers*, 84(6), 535-552.
[8] Chène P. (2006) *Chem. Med. Chem.*, 1(4), 400 - 411.
[9] Wells J.A., McClendon C.L. (2007) *Nature*, 450(7172), 1001-1009.
[10] Dömling A. (2008) Curr Opin Chem Biol, 12(3), 281-291.
[11] Betzi S., Guerlesquin F., Morelli X. (2009) *Comb. Chem. High Throughput Screen.*, 12(10), 968-983.
[12] Meireles L.M., Mustata G. (2011) Curr. Top. Med. Chem., 11 (3), 248-257.
[13] Vu B.T., Vassilev L. (2011) *Curr. Top. Microbiol. Immunol.*, 348, 151-172.
[14] Wilson C.G., Arkin M.R. (2011) *Curr. Top. Microbiol. Immunol.*, 348, 25-59.
[15] Zeitlin B.D., Nör J.E. (2011) *Curr. Top. Microbiol. Immunol.*, 348, 115-137.
[16] Fletcher S., Hamilton A.D. (2006) *J. R. Soc. Interface*, 3(7), 215-233.
[17] White P.W., Faucher A.M., Goudreau N. (2011) *Curr. Top. Microbiol. Immunol.*, 348, 61-88.
[18] Berg T. (2011) *Curr. Top. Microbiol. Immunol.*, 348, 139-149.
[19] Tilley J.W., Chen L., Fry D.C., Emerson S.D., Powers G.D., Biondi D., Varnell T., Trilles R., Guthrie R., Mennona F., Kaplan G., LeMahieu R.A., Carson M., Han R.J., Liu C.M.,

Palermo R., Ju G. (1997) *J. Am. Chem. Soc.*, 119(32), 7589-7590.
[20]Rush T.S., Grant J.A., Mosyak L., Nicholls A. (2005) *J. Med Chem*, 48(5), 1489-1495.
[21]He M.M., Smith A.S., Oslob J.D., Flanagan W.M., Braisted A.C., Whitty A., Cancilla M.T., Wang J., Lugovskoy A.A., Yoburn J.C., Fung A.D., Farrington G., Eldredge J.K., Day E.S., Cruz L.A., Cachero T.G., Miller S.K., Friedman J.E., Choong I.C., Cunningham B.C. (2005) *Science*, 310(5750), 1022-1025.
[22]Fry D.C., Graves B., Vassilev L.T. (2005) *Methods Enzymol.*, 399, 622-633.
[23]Mastrangelo E., Cossu F., Milani M., Sorrentino G., Lecis D., Delia D., Manzoni L., Drago C., Seneci P., Scolastico C., Rizzo V., Bolognesi M. (2008) *J. Mol. Biol.*, 384(3), 673-689.
[24]Shoemaker B.A., Panchenko A.R. (2007) *PLoS Comput. Biol.*, 3(4), e42.
[25]Mohr E., Horn F., Janody F., Sanchez C., Pillet V., Bellon B., Röder L., Jacq B. (1998) *Nucleic Acids Res.*, 26(1), 89-93.
[26]Cline M.S., Smoot M., Cerami E., Kuchinsky A., Landys N., Workman C., Christmas R., Avila-Campilo I., Creech M., Gross B., Hanspers K., Isserlin R., Kelley R., Killcoyne S., Lotia S., Maere S., Morris J., Ono K., Pavlovic V., Pico A.R., Vailaya A., Wang P.L., Adler A., Conklin B.R., Hood L., Kuiper M., Sander C., Schmulevich I., Schwikowski B., Warner G.J., Ideker T., Bader G.D. (2007) *Nat. Protoc.*, 2(10), 2366-2382.
[27]Valencia A., Pazos F. (2002) *Curr. Opin. Struct. Biol.*, 12(3), 368-373.
[28]Shoemaker B.A., Panchenko A.R. (2007) *PLoS Comput. Biol.*, 3(4), e43.
[29]Szilágyi A., Grimm V., Arakaki A.K., Skolnick J. (2005) *Phys. Biol.*, 2(2), S1-16.
[30]Skrabanek L., Saini H.K., Bader G.D., Enright A.J. (2008) *Mol. Biotechnol.*, 38(1), 1-17.
[31]Tuncbag N., Gursoy A., Keskin O. (2011) *Phys. Biol.*, 8(3), 035006.
[32]Wass M.N., David A., Sternberg M.J. (2011) *Curr. Opin. Struct. Biol.*, 21(3), 382-390.
[33]Rubinstein M., Niv M.Y. (2009) *Biopolymers*, 91(7), 505-513.
[34]Audie J., Boyd C. (2010) *Curr. Pharm. Des.*, 16(5), 567-582.
[35]Buchwald P. (2010) *IUBMB Life*, 62(10), 724-731.
[36]Sperandio O., Reynès C.H., Camproux A.C., Villoutreix B.O. (2010) *Drug Discov. Today*, 15(5-6), 220-229.
[37]Pagliaro L., Felding J., Audouze K., Nielsen S.J., Terry R.B., Krog-Jensen C., Butcher S. (2004) *Curr. Opin. Chem. Biol.*, 8 (4), 442-449.
[38]Neugebauer A., Hartmann R.W., Klein C.D. (2007) *J. Med. Chem.*, 50(19), 4665-4668.
[39]Todeschini R., Consonni V. (2009) *Molecular descriptors for chemoinformatics* (2). 1257
[40]Irwin J.J., Shoichet B.K. (2005) *J. Chem. Inf. Model*, 45(1), 177-182.
[41]Reynès C., Host H., Camproux A.C., Laconde G., Leroux F., Mazars A., Deprez B., Fahraeus R., Villoutreix B.O., Sperandio O. (2010) *PLoS Comput. Biol.*, 6(3), e1000695.
[42]Higueruelo A.P., Schreyer A., Bickerton G.R.J., Pitt W.R., Groom C.R., Blundell T.L. (2009) *Chem. Biol. Drug Des.*, 74 (5), 457-467.
[43]Schreyer A., Blundell T. (2009) *Chem. Biol. Drug Des.*, 73(2), 157-167.
[44]Hopkins A., Groom C., Alex A. (2004) *Drug Discov.,* Today, 9 (10), 430-431.
[45]Bourgeas R., Basse M.J., Morelli X., Roche P. (2010) *PLoS ONE*,, 5(3), e9598.
[46]Berman H., Henrick K., Nakamura H., Markley J.L. (2007) *Nucleic Acids Res.,* 35(Database issue), D301-303.
[47]Lipinski, C.A. (2004) *Drug Discov. Today: Technol.,* 1(4), 337-341.
[48]Morelli X., Bourgeas R., Roche P. (2011) *Curr. Opin. Chem. Biol.,* 15(4), 475-481.
[49]Lensink M.F., Wodak S.J. (2010) *Proteins*, 78(15), 3073-3084.
[50]Gonzàlez-Ruiz D., Gohlke H. (2006) *Curr. Med. Chem*. 13(22), 2607-2625.
[51]Laurie, A.T., Jackson, R.M. (2006) *Curr. Protein Pept.Sci*. 7 (5), 395-406.
[52]Schmidtke P., Souaille C., Estienne F., Baurin N., Kroemer R.T. (2010) *J. Chem. Inf. Model*. 50(12), 2191-2200.
[53]Laskowski, R.A. (1995) *J. Mol. Graph*, 13(5), 323-330.
[54]Hendlich M., Rippmann F., Barnickel G. (1997) *J. Mol. Graph Model*, 15(6), 359-363.
[55]Liang J., Edelsbrunner H., Woodward C. (1998) *Protein Sci,* 7 (9), 1884-1897.
[56]Cuff A.L., Martin A.C. (2004) *J. Mol. Biol.* 344(5), 1199-1209.
[57]Laurie, A.T.R., Jackson, R.M. (2005) *Bioinformatics.* 21(9), 1908-1916.
[58]Weisel M., Proschak E., Schneider G. (2007) *Chem. Cent. J.*, 1, 7.
[59]Harris R., Olson A.J., Goodsell D.S. (2007) *Proteins*, 70(4), 1506-1517.
[60]Le Guilloux V., Schmidtke P., Tufféry P. (2009) *BMC Bioinformatics,*1, 168.
[61]Halgren T.A. (2009) *J. Chem. Inf. Model,* 49(2), 377-389.
[62]Huang B., Schroeder M. (2006) BMC Struct Biol, 6, 19.
[63]Glaser F., Morris R.J., Najmanovich R.J., Laskowski R.A., Thornton J.M. (2006) *Proteins,* 62(2), 479-488.
[64]Schmidtke P., Bidon-Chanal A., Luque F.J., Barril X. (2011) *Bioinformatics*, 27(23), 3276-3285.
[65]Huang B. (2009) *Omics*, 13(4), 325-330.
[66]Zhang Z., Li Y., Lin B., Schroeder M., Huang B. (2011) *Bioinformatics*, 27(15), 2083-2088.
[67]Cheng A.C., Coleman R.G., Smyth K.T., Cao Q., Soulard P., Caffrey D.R., Salzberg A.C., Huang E.S. (2007) *Nat. Biotechnol*., 25(1), 71-75.
[68]Krissinel E., Henrick K. (2007) *J. Mol. Biol*., 372(3), 774-797.
[69]Tina K.G., Bhadra R., Srinivasan N. (2007) *Nucleic Acids Res*., 35(Web Server issue), W473-W476.
[70]Reynolds C., Damerell D., Jones S. (2009) *Bioinformatics*, 25 (3), 413-414.
[71]Clackson T., Wells J.A. (1995) *Science*, 267(5196), 383-386.
[72]Li X., Keskin O., Ma B., Nussinov R., Liang J. (2004) *J. Mol. Biol.*, 344(3), 781-795.
[73]DeLano W.L. (2002) *Curr. Opin. Struct. Biol.,* 12(1), 14-20.
[74]Thorn K.S., Bogan A.A. (2001) *Bioinformatics*, 17(3), 284-285.
[75]Massova I., Kollman P.A. (1999) *J. Am. Chem. Soc,,* 121(36), 8133-8143.
[76]Pitera J.W., Kollman P.A. (2000) *Proteins*, 41(3), 385-397.
[77]Kortemme T., Baker D. (2002) *Natl. Acad. Sci. USA*, 99(22),

14116-14121.
[78]Verkhivker G.M. (2004) *J. Mol. Graph Model*, 22(5), 335-348.
[79]Almlöf M., Åqvist J., Smalås A.O., Brandsdal B.O. (2006) *Biophys J.*, 90(2), 433-442.
[80]Lafont V., Schaefer M., Stote R.H., Altschuh D., Dejaegere A. (2007) *Proteins*, 67(2), 418-434.
[81]Benedix A., Becker C.M., de Groot B.L., Caflisch A., Bockmann R.A. (2009) *Nat. Methods*, 6(1), 3-4.
[82]Gohlke H., Kiel C., Case D.A. (2003) *J. Mol.Biol*., 330(4), 891-913.
[83]del Sol A., O'Meara P. (2005) *Proteins*, 58(3), 672-682.
[84]Darnell S.J., Page D., Mitchell J.C. (2007) *Proteins*, 68(4), 813 -823.
[85]Pavelka A., Chovancova E., Damborsky J. (2009) *Nucleic Acids Res*., 37(Web Server issue), W376-W383.
[86]Tuncbag N., Gursoy A., Keskin O. (2009) *Bioinformatics*, 25 (12), 1513-1520.
[87]Tuncbag N., Keskin O., Gursoy A. (2010) *Nucleic Acids Res*., 38(Web Server issue), W402-W406.
[88]Lise S., Archambeau C., Pontil M., Jones D.T. (2009) *BMC Bioinformatics*, 10, 365.
[89]Lise S., Buchan D., Pontil M., Jones D.T. (2011) *PLoS One*, 6 (2), e16774.
[90]Xia J.F., Zhao X.M., Song J., Huang D.S. (2010) *BMC Bioinformatics*, 11, 174.
[91]Li Z., Wong L., Li J. (2011) *BMC Syst. Biol.*, 5(1), S5.
[92] Chen R., Chen W., Yang S., Wu D., Wang Y., Tian Y., Shi Y. (2011) *BMC Bioinformatics*, 12, 311.
[93] Deng L., Guan J.H., Dong Q.W., Zhou S.G. (2011) *Protein Pept. Lett.*, 18(9), 896-905.
[94] Ma D.C., Diao Y.B., Guo Y.Z., Li Y.Z., Zhang Y.Q., Wu J., Li M.L. (2011) *Protein Pept. Lett*., 18(9), 906-911.
[95] Segura J., Fernandez-Fuentes N. (2011) *Nucleic Acids Res.*, 39, D755-D760.
[96] Cukuroglu E., Gursoy A., Keskin O. (2012) *Nucleic Acids Res.*, 40, D829-D833.
[97] Durrant J.D., McCammon J.A. (2011) *BMC Biology*, 9, 71.
[98] Rajamani D., Thiel S., Vajda S., Camacho C.J. (2004) *Natl Acad Sci USA*, 101(31), 11287-11292.
[99] Sotriffer C.A., Krämer O., Klebe G. (2004) *Proteins*, 56(1), 52-66.
[100] Smith G.R., Sternberg M.J., Bates P.A. (2005) *J. Mol. Biol.*, 347(5), 1077-1101.
[101] Eyrisch S., Helms V. (2009) *J. Comput. Aided Mol. Des.*, 23 (2), 73-86.
[102] Boehr D.D., Nussinov R., Wright P.E. (2009) *Nat. Chem. Biol.*, 5(11), 789-796.
[103] McClendon C.L., Friedland G., Mobley D.L., Amirkhani H., Jacobson M.P. (2009) *J. Chem. Theory Comput.*, 5(9), 2486-2502.
[104] Case D.A. (1994) *Curr. Opin. Struct. Biol.*, 4(2), 285-290.
[105] Zhong S., Macias A.T., MacKerell Jr. A.D. (2007) *Curr. Top. Med. Chem.*, 7(1), 63-82.
[106] Bottegoni G. (2011) *Front Biosci*, 17, 2289-2306.
[107] Meng X.Y., Zhang H.X., Mezei M., Cui M. (2011) *Curr. Comput. Aided Drug Des.*, 7(2), 146-157.
[108] Honig, B., Nicholls A. (1995) *Science*, 268(5214), 1144-1149.
[109] Grant J.A., Pickup B.T., Nicholls A. (2001) *J. Comp. Chem.*, 22(6), 608-640.
[110] Bashford D., Case D.A. (2000) *Annu. Rev. Phys. Chem.*, 51, 129-152.
[111] Marabotti A. (2008) *Curr. Chem. Biol.*, 2, 200-214.
[112] Cozzini P., Kellogg G.E., Spyrakis F., Abraham D.J., Costantino G., Emerson A., Fanelli F., Gohlke H., Kuhn L.A., Morris G.M., Orozco M., Pertinhez T.A., Rizzi M., Sotriffer C.A. (2008) *J. Med. Chem*., 51(20), 6237-6255.
[113] Antes I. (2010) *Proteins*, 78(5), 1084-1104.
[114] Erlanson D.A., McDowell R.S., O'Brien T. (2004) *J. Med. Chem.*, 47(14), 3463-3482.
[115] Erlanson D.A. (2006) *Curr. Opin. Biotechnol.*, 17(6), 643-652.
[116] Carr R.A., Congreve M., Murray C.W., Rees D.C. (2005) *Drug Discov. Today*, 10(14), 987-992.
[117] Schulz M.N., Hubbard R.E. (2009) *Curr. Opin. Pharmacol.*, 9 (5), 615-621.
[118] Congreve M., Carr R., Murray C., Jhoti H. (2003) *Drug Discov Today*, 8(19), 876-877.
[119] Jacoby E., Davies J., Blommers M.J. (2003) *Curr Top Med Chem*, 3(1), 11-23.
[120] Fink T., Bruggesser H., Reymond J.L. (2005) *Angew Chem Int Ed Engl*, 44(10), 1504-1508.
[121] Shuker S.B., Hajduk P.J., Meadows R.P., Fesik S.W. (1996) *Science*, 274(5292), 1531-1534.
[122] Nienaber V.L., Richardson P.L., Klighofer V., Bouska J.J., Giranda V.L., Greer J. (2000) *Nat Biotechnol*, 18(10), 1105-1108.
[123] Carr R., Jhoti H. (2002) *Drug Discov Today*, 7(9), 522-527.
[124] Hamalainen M.D., Zhukov A., Ivarsson M., Fex T., Gottfries J., Karlsson R., Bjorsne M. (2008) *J. Biomol Screen*, 13(3), 202-209.
[125] Breuker K. (2003) *Angew. Chem. Int. Ed*, 43(1), 22-25.
[126] Erlanson D.A., Braisted A.C., Raphael D.R., Randal M., Stroud R.M., Gordon E.M., Wells J.A. (2000) *Natl. Acad. Sci.*, 97(17), 9367-9372.
[127] Fattori D., Squarcia A., Bartoli S. (2008) *Drugs RD*, 9(4), 217 -227.
[128] Gozalbes R., Carbajo R.J., Pineda-Lucena A. (2010) *Curr. Med. Chem*, 17(17), 1769-1794.
[129] Hoffer L., Renaud J.P., Horvath D. (2011) *Comb. Chem. High Throughput Screen*, 14(6), 500-520.
[130] Congreve M., Chessari G., Tisi D., Woodhead A.J. (2008) J Med Chem, 51(13), 3661-3680.
[131] Fuller J.C., Burgoyne N.J., Jackson R.M. (2009) *Drug Discov. Today*, 14(3-4), 155-161.
[132] Petros A.M., Dinges J., Augeri D.J., Baumeister S.A., Betebenner D.A., Bures M.G., Elmore S.W., Hajduk P.J., Joseph M.K., Landis S.K., Nettesheim D.G., Rosenberg S.H., Shen W., Thomas S., Wang X., Zanze I., Zhang H., Fesik S.W. (2006) *J. Med. Chem*., 49(2), 656-663.
[133] Oltersdorf T., Elmore S.W., Shoemaker A.R., Armstrong R.C., Augeri D.J., Belli B.A., Bruncko M., Deckwerth T.L., Dinges J., Hajduk P.J., Joseph M.K., Kitada S., Korsmeyer S.J., Kunzer A.R., Letai A., Li C., Mitten M.J., Nettesheim D.G., Ng S., Nimmer P.M., O'Connor J.M., Oleksijew A., Petros A.M., Reed J.C., Shen W., Tahir S.K., Thompson C.B.,

Tomaselli K.J., Wang B., Wendt M.D., Zhang H., Fesik S.W., Rosenberg S.H. (2005) *Nature*, 435(7042), 677-681.
[134] Wendt M.D., Shen W., Kunzer A., McClellan W.J., Bruncko M., Oost T.K., Ding H., Joseph M.K., Zhang H., Nimmer P.M., Ng S.C., Shoemaker A.R., Petros A.M., Oleksijew A., Marsh K., Bauch J., Oltersdorf T., Belli B.A., Martineau D., Fesik S.W., Rosenberg S.H., Elmore S.W. (2006) *J. Med. Chem.*, 49 (3), 1165-1181.
[135] Park C.M., Bruncko M., Adickes J., Bauch J., Ding H., Kunzer A., Marsh K.C., Nimmer P., Shoemaker A.R., Song X., Tahir S.K., Tse C., Wang X., Wendt M.D., Yang X., Zhang H., Fesik S.W., Rosenberg S.H., Elmore S.W. (2008) J. Med. Chem., 51(21), 6902-6915.
[136] Tse C., Shoemaker A.R., Adickes J., Anderson M.G., Chen J., Jin S., Johnson E.F., Marsh K.C., Mitten M.J., Nimmer P., Roberts L., Tahir S.K., Xiao Y., Yang X., Zhang H., Fesik S., Rosenberg S.H., Elmore S.W. (2008) *Cancer Res.*, 68(9), 3421-3428.
[137] Emerson S.D., Palermo R., Liu C.M., Tilley J.W., Chen L., Danho W., Madison V.S., Greeley D.N., Ju G., Fry D.C. (2003) *Protein Sci.*, 12(4), 811-822.
[138] Arkin M.R., Mike R., Delano W.L., Hyde J., Luong T.N., Oslob J.D., Raphael D.R., Taylor L., Wang J., McDowell R.S., Wells J.A., Braisted A.C. (2003) *Natl Acad. Sci.*, 100(4), 1603-1608.
[139] Braisted A.C., Oslob J.D., Delano W.L., Hyde J., McDowell R.S., Waal N., Yu C., Arkin M.R., Raimundo B.C. (2003) *J. Am. Chem. Soc.*, 125(13), 3714-3715.
[140] Hyde J., Braisted A.C., Randal M., Arkin M.R. (2003) *Biochemistry*, 42(21), 6475-6483.
[141] Thanos C.D., Randal M., Wells J.A. (2003) *J. Am. Chem. Soc.*, 12550 15280-15281.
[142] Raimundo B.C., Oslob J.D., Braisted A.C., Hyde J., McDowell R.S., Randal M., Waal N.D., Wilkinson J., Yu C.H., Arkin M.R. (2004) *J. Med. Chem.*, 47(12), 2111-3130.
[143]Blundell T.L., Sibanda B.L., Montalvao R.W., Brewerton S., Chelliah V., Worth C.L., Harmer N.J., Davies O., Burke D. (2006) *Phil Trans R. Soc. B.*, 361(1467), 413-423.
[144]Hubbard R.E., Chen I., Davis B. (2007) *Curr. Opin. Drug. Discov. Devel.*, 10(3), 289-297.
[145]Villar H.O., Hansen M.R. (2007) *Curr. Top. Med. Chem.*, 7 (15), 1509-1513.
[146]Law R., Barker O., Barker J.J., Hesterkamp T., Godemann R., Andersen O., Fryatt T., Courtney S., Hallett D., Whittaker M. (2009) *J. Comput. Aided Mol. Des.*, 23(8), 459-473.
[147]Joseph-McCarthy D. (2009) *J. Comput. Aided Mol. Des.*, 23 (8), 449-451.
[148] Van Drie J., Lajiness M. (1998) *Drug Discov Today*, 3(6), 274-283.
[149]Lewis R.A., Mason J.S., McLay I.M. (1997) *J. Chem. Inf. Comput. Sci.*, 37(3), 599-614.
[150]Patterson D.E., Cramer R.D., Ferguson A.M., Clark R.D., Weinberger L.E. (1996) *J. Med. Chem.*, 39(16), 3049-3059.
[151]Horvath D., Jeandenans C. (2003) *J. Chem. Inf. Comput. Sci.*, 43(2), 691-698.
[152] Bergmann R., Linusson A., Zamora I. (2007) *J. Med. Chem.*, 50(11), 2708-2717.
[153]Blomberg N., Cosgrove D.A., Kenny P.W., Kolmodin K. (2009) *J. Comput. Aided Mol. Des.*, 23(8), 513-525.
[154] Fejzo J., Lepre C.A., Peng J.W., Bemis G.W., Ajay, Murcko M.A., Moore J.M. (1999) *Chem. Biol.*, 6(10), 755-769.
[155] Lewell X.Q., Judd D.B., Watson S.P., Hann M.M. (1998) *J. Chem. Inf. Comput. Sci*., 38(3), 511-522.
[156] Kolb P., Caflisch A. (2006) *J. Med. Chem*., 49(25), 7384-7392.
[157] Gianti E., Sartori L. (2008) *J. Chem. Inf. Model.*, 48(11), 2129-2139.
[158] Babaoglu K., Shoichet B.K. (2006) *Nat. Chem. Biol.*, 2(12), 720-723.
[159] Loving K., Salam N.K., Sherman W. (2009) *J. Comput. Aided Mol. Des.*, 23(8), 541-554.
[160] Kawatkar S., Wang H., Czerminski R., Joseph-McCarthy D. (2009) *J. Comput. Aided Mol. Des.*, 23(8), 527-539.
[161] Huang D., Caflisch A. (2010) *J. Mol. Recognit.*, 23(2), 183-193.
[162] Durrant J.D., Friedman A.J., McCammon J.A. (2011) *J. Chem. Inf. Model.*, 51, 2573-2580.
[163] Imai T., Oda K., Kovalenko A., Hirata F., Kidera A. (2009) *J. Am. Chem. Soc.*, 131(34), 12430-12440.
[164] Dey F., Caflisch A. (2008) J Chem Inf Model, 48(3), 679-690.
[165] Thompson D., Denny R., Nilakantan R., Humblet C., Joseph-McCarthy D., Feyfant E. (2008) *J. Comput. Aided Mol. Des.*, 22 (10), 761-772.
[166] Bohm H.J. (1992) *J. Comput. Aided Mol. Des.*, 6(6), 61-78.
[167] Gillet V.J., Johnson A.P., Mata P., Sike S., Williams P. (1993) *J. Comput. Aided Mol. Des.*, 7(2), 127- 153.
[168]DeWitte R.S., Shakhnovich E.I. (1996) *J. Am. Chem. Soc*, 118(47), 11733-11744.
[169] Wang R., Gao Y., Lai L. (2000) *J. Mol. Model*, 6(7-8), 498-516.
[170] Durrant J.D., Amaro R.E., McCammon J.A. (2009) *Chem. Biol. Drug. Des.*, 73(2), 168-178.
[171]Schneider G., Lee M., Stahl M., Schneider P. (2000) *J. Comput. Aided Mol. Des.*, 14(5), 487-494.
[172] Schneider G., Fechner U. (2005) *Nature Rev. Drug Discov.*, 4(8), 649-663.
[173] Proschak E., Sander K., Zettl H., Tanrikulu Y., Rau O., Schneider P., Schubert-Zsilavecz M., Stark H., Schneider G. (2009) *Chem. Med. Chem*, 4(1), 45-48.
[174] Zaliani A., Boda K., Seidel T., Herwig A., Schwab C.H., Gasteiger J., Claussen H., Lemmen C., Degen J., Parn J., Rarey M. (2009) *J. Comput. Aided Mol. Des.*, 23(8), 593-602.
[175] Nisius B., Rester U. (2009) *J. Chem. Inf. Model*, 49(5), 1211-1222.
[176] Ji H., Stanton B.Z., Igarashi J., Li H., Martasek P., Roman L.J., Poulos T.L., Silverman R.B. (2008) *J. Am. Chem. Soc.*, 130(12), 3900-3914.
[177] Gozalbes R., Mosulen S., Carbajo R.J., Pineda-Lucena A. (2009) *J. Comput. Aided Mol. Des*., 23(8), 555-569.

Table 1- List of the computational tools cited in the text.

a) Libraries enriched in i-PPIs				
Library	**Type**	**Availability**	**Reference**	**URL**
MOE	Software	Commercial	-	http://www.chemcomp.com
Dragon	Software	Commercial	39	http://www.talete.mi.it
PPI-HitProfiler	Software	Free for academic users	41	http://www.cdithem.fr/ppiHitProfiler.php?lg=en
TIMBAL	Database	Free	42	http://www-cryst.bioc.cam.ac.uk/timbal
2P2I	Database	Free	45	http://2p2idb.cnrs-mrs.fr
b) Tools for the analysis of protein-protein interfaces				
SURFNET	Software	Free for academic users	53	http://www.biochem.ucl.ac.uk/~roman/surfnet/surfnet.html
LIGSITE^{CSC}	Web server	Free	62	http://projects.biotec.tu-dresden.de/pocket/
CASTp	Web server	Free	55	http://sts.bioengr.uic.edu/castp/
AVP	Software	Free	56	http://www.bioinf.org.uk/software/avp/
Q-SiteFinder	Web server	Free	57	http://www.modelling.leeds.ac.uk/qsitefinder/
PocketPicker	Software	Free	58	http://gecco.org.chemie.uni-frankfurt.de/pocketpicker/index.html
AutoLigand	Software	Free	59	http://autodock.scripps.edu/resources/autoligand
fpocket	Software	Free	60	http://fpocket.sourceforge.net/
SiteMap	Software	Commercial	61	http://www.schrodinger.com/products/14/20/
ConSurf	Web server	Free for academic users	63	http://consurf.tau.ac.il/
MDpocket	Software	Free	64	http://fpocket.sourceforge.net/
MetaPocket	Web server	Free	65	http://metapocket.eml.org/
MetaPocket 2.0	Web server	Free	66	http://sysbio.zju.edu.cn/metapocket/
PDBePISA	Web server	Free	68	http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html
PIC	Web server	Free	69	http://pic.mbu.iisc.ernet.in/
PROTORP	Web server	Free	70	http://bioinformatics.sussex.ac.uk/protorp/
ASEdb	Database	Free	74	http://nic.ucsf.edu/asedb/
KFC Server	Web server	Free	84	http://kfc.mitchell-lab.org/
HotSpot Wizard	Web server	Free	85	http://loschmidt.chemi.muni.cz/hotspotwizard/
HotPoint	Web server	Free	86,87	http://prism.ccbb.ku.edu.tr/hotpoint/
HSPred	Web server	Free	88,89	http://bioinf.cs.ucl.ac.uk/hspred
PCRPi-DB	Database	Free	95	http://www.bioinsilico.org/PCRPIDB
HotRegion	Database	Free	96	http://prism.ccbb.ku.edu.tr/hotregion
c) Tools for fragment-based drug design				
Pyramid	Software	Commercial	123,133	http://astx.com/technology/pyramid-platform/
RECAP	Software	Commercial	155	http://www.chemaxon.com/jchem/doc/user/fragment_recap.html
DAIM	Software	Free	156,161	http://www.biochem-caflisch.uzh.ch/download/
Pipeline Pilot	Software	Commercial	157	http://accelrys.com/products/pipeline-pilot/
Glide	Software	Commercial	159,160	http://www.schrodinger.com/
CrystalDock	Software	Free	162	http://www.nbcr.net/crystaldock/
3D-RISM	Software	Commercial	163	http://www.scm.com/Products/Overview/ADFinfo.html
GANDI	Software	Free	164	http://www.biochem-caflisch.uzh.ch/download/
CONFIRM	Software	Commercial	165	http://accelrys.com/products/pipeline-pilot/
LUDI	Software	Commercial	166	http://accelrys.com/services/training/life-science/FragmentBasedApproachesDescription.html
SPROUT	Software	Commercial	167	http://www.simbiosys.ca/sprout/index.html
SMoG	Software	Free	168	http://www-shakh.harvard.edu/~smog/
LigBuilder V1.2	Software	Free	169	http://mdl.ipc.pku.edu.cn/cgi-bin/down.cgi?kind=e
AutoGrow	Software	Free	170	http://autogrow.ucsd.edu/