



**HAL**  
open science

## **In silico-in vitro screening of protein-protein interactions: towards the next generation of therapeutics.**

Bruno Villoutreix, Karine Bastard, Olivier Sperandio, Robin Fahraeus, Jean-Luc Poyet, Fabien Calvo, Benoit Déprez, Maria A. Miteva

### ► To cite this version:

Bruno Villoutreix, Karine Bastard, Olivier Sperandio, Robin Fahraeus, Jean-Luc Poyet, et al.. In silico-in vitro screening of protein-protein interactions: towards the next generation of therapeutics.. *Current Pharmaceutical Biotechnology*, 2008, 9 (2), pp.103-122. 10.2174/138920108783955218 . inserm-00273685

**HAL Id: inserm-00273685**

**<https://inserm.hal.science/inserm-00273685>**

Submitted on 13 Nov 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## ***In silico-in vitro* screening of protein-protein interactions: towards the next generation of therapeutics**

Bruno O. Villoutreix<sup>#&\$</sup>, Karine Bastard<sup>#</sup>, Olivier Sperandio<sup>#&</sup>, Robin Fahraeus<sup>\*&</sup>, Jean-Luc Poyet<sup>\*&</sup>, Fabien Calvo<sup>\*&</sup>, Benoit Déprez<sup>&%</sup>, Maria A. Miteva<sup>#&</sup>

<sup>&</sup>c-Dithem, Inserm Consortium for Discovery and Innovation in Therapy and Medicine

\*Centre d'Investigations Cliniques CIC 9504  
and INSERM U 716  
Hôpital Saint-Louis, Paris, France

<sup>#</sup>Inserm U648  
University of Paris 5  
75006 Paris, France

<sup>%</sup>Inserm U761  
Institut Pasteur Lille  
1 rue du Professeur Calmette  
59019 Lille, France

<sup>\$</sup>Correspondence should be addressed to:  
Bruno O. Villoutreix, PhD  
Email: [bruno.villoutreix@univ-paris5.fr](mailto:bruno.villoutreix@univ-paris5.fr)

**Key words:** virtual ligand screening, structure-based drug design, docking, scoring, drug discovery, protein-protein interaction.

## **Summary**

Protein-protein interactions (PPIs) have a pivotal role in many biological processes suggesting that targeting macromolecular complexes will open new avenues for the design of the next generation of therapeutics. A wide range of “in silico methods” can be used to facilitate the design of protein-protein modulators. Among these methods, virtual ligand screening, protein-protein docking, structural predictions and druggable pocket predictions have become established techniques for hit discovery and optimization. In this review, we first summarize some key data about protein-protein interfaces and introduce some recently reported computer methods pertaining to the field. URLs for several recent free packages or servers are also provided. Then, we discuss four studies aiming at developing PPI modulators through the combination of *in silico* and *in vitro* screening experiments.

## Introduction

Protein-protein interactions (PPIs) control most biological processes and the prospect of manipulating macromolecular complexes, either by blocking or stabilizing molecular interactions, holds promise for generating novel therapeutic agents as well as tools to improve our understanding of biochemical pathways [1-6]. However, finding drug-like compound modulators of PPIs has long been regarded as intractable. Indeed, the concept was that macromolecular protein-protein complexes could only be modulated (impeded or stabilized) by relatively large molecules, like, for examples, heparin [7-11], peptides (e.g., aptamers) or (miniature) proteins [12-17] or large antibodies (as for example seen *in vivo* in pathological conditions such as Haemophilia A where antibodies usually impede macromolecular interactions [18]) or through the design (or lucky finding) of monoclonal antibodies that can enhance protein-protein interaction (e.g., the FIXa-FVIIIa interaction) [19]. Also, one of the possible reasons for slow progress with the design of drug-like modulators of PPIs (or other targets) in academic settings is linked to the fact that screening compound collections to find “hits” was considered almost up to the 90’, as anything but science. Then, from around 1985 to 1995, the concept went from deep skepticism to marginal enthusiasm while today many groups that were against this strategy claim that screening, whether using experimental high-throughput screening (HTS) methods or *in silico* approaches, represents an exciting paradigm in academic biological research !

Projects involving the development of small-molecule modulators of PPIs can clearly be considered at first as very risky: the interface region is usually very large, often fantasized as featureless and undergoing conformational changes while the biological assays to validate the effects of a small molecule on PPIs are usually truly difficult to design and to miniaturize for HTS studies. However, over the years, many research/clinical studies were pointing to the direction of targeting PPIs. For instance, some 10 years ago, alanine-scanning mutagenesis was used to identify what is commonly called today “hot-spots” (residues that contribute over 2 kcal/mol to the binding free energy) [20], suggesting that it may not be necessary for a small molecule to span a large part of the interface. Along the same line of reasoning, but after analysis of mutation data in patients (thus *in vivo*) and confirmation of the possible impact of the substitution *in vitro* via site directed mutagenesis, it has also been observed that naturally occurring mutations could impede protein-protein interaction (see for example the factor VIII-von Willebrand factor interaction [21] or the discoidin domain receptor 1-collagen interaction [22]), suggesting that a small perturbation in one specific region (i.e., most likely a hot-spot region but allosteric sites can also impede association) can significantly alter a macromolecular interaction, opening the door for drug-like molecule interventions and rational design. While PPI antagonists would be very valuable in many disease indications, it is important to keep in mind that PPI stabilizers could also be beneficial. Indeed, some molecules are well-known to perform this task [23], for examples: antibodies, peptides and heparin (e.g., bridging effects) but drug-like molecule stabilizers or promoting protein-protein associations are less documented [23, 24]. In addition, when some stabilizers are identified, the mechanisms of action(s) are usually not fully understood (allosteric effects, bridging effects, stabilization of one partner, etc), making the rational design of drug-like stabilizers very difficult.

In order to identify drug-like molecules able to interfere with a protein function, several methodologies can be employed. While biophysical methods (NMR, crystallography, fragment-based approaches, etc)[25-28] and HTS experiments have all been shown efficient for hit-finding (although the hit rates can be low with HTS methods [29]), a major drawback with these approaches is the cost, with screening campaigns easily reaching hundred thousands to millions of US dollars (screening 100,000 compounds can cost from 100,000 to 1 million US \$ while solving the X-ray structure of some proteins can cost well over 1 million US \$) [30-32]. *In silico* simulations can reduce the time and cost of the overall process in many cases (e.g., a 3D structure of the target is known or can be generated through

threading/comparative model building or some active ligands are known) [33-35]. It is also important to mention that *in silico* approaches can be relatively expensive as well, depending on the protocols and selected algorithms (e.g., stochastic methods are usually very CPU demanding) [36, 37]. While *in silico* strategies have been successfully applied to catalytic sites, targeting PPIs is still considered as extremely challenging. In fact, it may not be more challenging to target PPIs than some catalytic sites: it all depends on the topology, physico-chemical nature and malleability of the interface area. In fact, several molecules disrupting protein-protein interactions have already been reported, often through a combination of *in silico-in vitro* screening strategies [1, 38-40]. These studies definitively open new avenues for therapeutic interventions.

Computational techniques yield useful insights into an ever-wider range of biomolecular systems but also suffer from limitations that will have to be addressed in the coming years. Protein three-dimensional structures and molecular functions can be predicted in some circumstances, non-covalent binding of biomolecules can be understood by considering structural, thermodynamic and kinetic issues, and theoretical simulations of such events can be attempted. Chemoinformatics approaches combined with structural bioinformatics and appropriate experimental approaches are now well established and help to prioritize, design, analyze and rationalize experiments and ultimately can facilitate the design of new therapeutic compounds (Fig. 1). Drug discovery requires an integrated, multi-disciplinary approach considering the complexity of the task at hand. With the above data in mind, the purpose of the present review is to present some recent studies in which scientists were able to modulate protein-protein interactions through combined *in silico-in vitro* screening experiments. Yet, prior to discuss these examples, we will introduce several concepts pertaining to the field of protein-protein interactions (i.e., nature of the interface, prediction of residues potentially involved in protein-protein interaction) and to the field of chemoinformatics (i.e., druggable pockets, compound collections and virtual screening methods).

## Protein-protein interface: a brief overview

Proteins are usually involved in interactions with an estimated average of 5-10 protein partners [41], the binding site(s) with these partners may or may not overlap, illustrating the complexity ahead with regard to identifying, understanding and predicting protein interaction networks. In addition, different types of protein complexes have been observed, like homo- and hetero-complexes (i.e., the interaction occur between identical or non-identical chains), obligate and non-obligate complexes (a related view refers to interactions as transient or permanent) [42-45]. Well-known examples can be, for a permanent complex, the interleukin 8 homodimer (PDB code 1il8), and for a non-obligate complex, the cytochrome c peroxidase associated with cytochrome c (PDB code 2pcb). Depending on the types of complexes (permanent, transient...), the nature of the interface usually differs, from somewhat hydrophobic/aromatic to a mixture of hydrophobic/aromatic and polar properties [45]. Different *in silico* methods have been developed to analyze the physical properties of proteins and learn more about how these properties are associated with binding. Interactions at the interface are usually measured *in silico* in terms of accessible surface area, shape complementarity, molecular surface and the associated mathematical objects, Voronoi diagram, Delaunay tessellation, Laguerre polyhedral decomposition and  $\alpha$ -shape [46-49]. Topographical information can also be determined using an algorithm based on the measurement of atomic density near the molecular surface [50]. Results (in term of statistics) obtained with these methods have to be taken with cautions since computations are usually performed on static structures. Nevertheless, such investigations have obviously been shedding lights on PPIs. Interfaces have been found to be either relatively flat or to present convex and concave shapes, indicating in these latter cases that shape-shape complementarity in addition to chemical complementarity can be of major importance for PPIs (just like for protein-small ligands). The importance of the “shape descriptor” essentially applies to homodimers, enzyme-inhibitor complexes and permanent heterocomplexes [51]. In some other cases, electrostatic steering plays an essential role and can drive the formation of an encounter complex and of the final functional complex [52]. In addition, analysis of packing at the interface was also performed and suggests that for homodimers, enzyme-inhibitor complexes and permanent heterocomplexes, the interfaces are usually well-packed [53].

Following the discovery of hot spots [20], computer analysis of crystallized complexes has provided new insights about protein interfaces [54-61]. However, this topic is still an area of intense investigations [44, 62-68]. Yet, it has been shown that contiguous regions on a protein could form the signature for protein-binding, these are commonly called ‘patches’ [69]. A surface area, less than 800 Å<sup>2</sup>, would form a key interaction point, composed of specific amino acids, surrounded by a zone, which is supposed to isolate this contact zone from the solvent. This finding joins the theory of ‘O-ring’ described by Bogan et al. [54]. Regions surrounding the hot-spots residues are highly dense, leading to solvent exclusion, thereby decreasing the local dielectric constant, and enhancing electrostatic interactions [70]. Identification of conserved residues alone is generally not sufficient for complete and accurate prediction of protein-protein interfaces. The interface is rarely significantly more conserved than other surface patches and when an interface is among the most conserved surface patches, it tends to be part of an enzyme active site [71]. Thus, this structural/physico-chemical diversity renders theoretical predictions of hot-spots rather difficult. Interestingly, molecular dynamics simulations on sets of complexes have highlighted that few key interface residues generally adopt the same conformation in the unbound and bound-form of the partners whereas peripheral interface residues adopt a wider range of possible rotamers [72-74]. Help to investigate interfaces will in part come from further analysis of protein-protein complexes and unbound partners. To this end, databases of protein interactions are instrumental [75-89].

The energy of interaction between two proteins can be computed to help predict or analyze complexes and interfaces. Several types of scoring functions have been reported, ranging from physics-

based methods to statistical potentials (see for example [57, 90-93]). Several tools have been proposed to predict binding sites, ranging from computation of the relative free energy of interaction, identification of hydrophobic clusters, search for planarity, protrusion, sequence similarity, phylogenetic profiling to solvation potentials (many free tools are listed at [www.vls3d.com](http://www.vls3d.com) [94] and in Table I) [57, 93, 95-107]. It is still very challenging to obtain the experimental structure of protein-protein complexes but many recent and promising docking have been reported and should assist the process [108-124]) (see some protein-protein docking packages in Table I and in [94]). Ranking of the proposed binding modes after protein-docking experiments is very difficult, and therefore the scoring functions are usually supplemented with experimental data (i.e., concept of protein docking via hybrid approaches), like site-directed mutagenesis, epitope mapping or site-directed glycosylation [125, 126], yeast-two-hybrid [127], FRET/BRET [128], surface plasmon resonance [129] and calorimetry [130] coupled with for instance mutagenesis, SAXS [131], electron microscopy [132], affinity purification-mass spectrometry [133], protein microarrays [134] or NMR [57].

In some situations, it will be possible to obtain a docked model that matches with some experimental data and presents some “hot-spot” residues at the interface together with a so-called druggable pocket (see below). When such situation occurs, it should then be possible to perform *in silico* screening experiments [57, 135]. Such a project (interface unknown or not well defined) would seem very challenging and most likely would require the testing of a relatively large number of compounds (e.g., from 2000 to 50,000) *in vitro* to increase the chance of finding hits. Yet, even these very difficult cases, the *in silico* approach may still be valuable and time/money-saving as compared to the screening of hundreds of thousands of compounds. Along the same line, it could be valuable to act next to the interface regions, and, in this case one may search for allosteric sites. However, this exercise can be extremely challenging to available *in silico* methods as the predictions and handling of allosteric sites are only possible in some rare cases. New methods should allow such projects to be undertaken effectively in the coming years [57, 136, 137].

The notion of druggable pocket is very important for *in silico* and experimental screening alike (Fig. 2) [138] but before commenting about binding pockets, it is important to define the terms “hot-spots” and druggability. Hot-spots, as mentioned above, are mainly used for protein-protein interactions but some authors also use “hot-spots” for residues within the binding pocket that contribute the most to ligand binding (i.e., affinity). The term druggability can be confusing since it does not only stand for the ability of a protein’s binding pocket to bind a drug-like compound with high affinity [138] but it can also refer to as predicting if a protein is druggable in terms of tissue distribution, number of pathways in which the target is involved, the number of homologues outside its family, etc [3]. In the following paragraph, we discuss druggability in term of pocket and not directly in term of validation/selection of a target (i.e., the two can be linked as if they are no pockets, the target may not be amenable to *in silico-in vitro* methods). Also, we will not use “hot-spots” for druggable pocket *per se*.

Druggable pockets may not be always visible in a static 3D structure but if they do, they usually have to display some key structural features and physico-chemical characteristics to be considered as druggable and as such they tend to be different from any pocket and small cavity that are often present at a protein surface (i.e., a druggable pocket tends to have some specific shape, volume, content in hydrophobic residues, depth...)[139-143]. Here again, just like in the case of the prediction of protein-protein interfaces, the rules to define a pocket as druggable are not strict, although, in general, a certain volume is required (e.g., from about 300 to ~1000 Å<sup>3</sup>) [139, 144, 145]. Many different algorithms have been developed to investigate pockets, including neural network and support vector machine approaches, that can help in some situations but not all due to the dependence of these tools on the training set and the subjective definition of what is and what is not a binding pocket [144, 146-151]. Binding pockets, interfaces and local

geometries [152] can be analyzed and/or compared, like for instance via SitesBase [153], Cavbase [154], LigASite [155], SimDock [156] and sc-PDB [157] (please see a detailed list in [94]). If a druggable pocket overlaps with a hot-spot, then it is likely that *in silico* methods would succeed in identifying drug-like modulators for this particular PPI. The nature of the binding pocket can also help in deciding which docking tools is more appropriate and which scoring functions are more likely to find hits [37, 140, 158]. Many free and commercial tools are available to help identify and analyze these cavities [94] (Table II), they usually rely on geometric algorithms and/or probe matching algorithms (in average, energy-based approaches, i.e., probe matching, tend to perform better than purely geometry-based methods) [159]. As mentioned above, complex formation is accompanied in some cases by various structural changes: these include changes in the position of side chains, motions of the main chain or entire domains. These structural changes can influence the detection of druggable pockets and require special considerations (see below).

The rationale behind interfering with PPIs is still under debate. For instance, some authors suggest that inhibiting protein association is not really possible from a thermodynamic standpoint assuming that protein-protein interfaces should be mostly flat [160]. This is understandable but among the many protein-protein interactions taking place in living organisms, certainly a sizable number will present a druggable pocket at the interface. Moreover, we have already succeeded in blocking a rather flat interface with small drug-like molecules (manuscript in preparation) suggesting that additional thinking is needed before deciding if this mechanism is druggable or not. In order to investigate PPIs, we have analyzed about 50 complexes obtained from two databases of experimentally determined protein-protein complexes, PROTCOM [85] and DOCKGROUND [86]. We, for instance, investigated the representative unbound collection of protein complexes recently reported by Gao *et al.* [86], removed antibody-protein complexes as well as enzyme-inhibitor complexes and focused our attention on a small set of 15 non-redundant, representative and diverse protein complexes. Among them, about 60-70% present with a well defined druggable pocket at the interface (Fig. 3). These pockets are clearly visible on the crystal structure of the isolated partners and are very similar in the complexes, indicating that no major structural changes are taking place upon complex formation and that with some experimental knowledge of the interface area and druggable pocket prediction algorithms, it would be possible to use *in silico* screening tools to target these interactions. Furthermore, we investigated several complexes from the bound collection [86] and noticed that next to the interface, druggable pockets are often present, suggesting that allosteric sites are likely to be present at many interfaces, definitively opening possibilities for the design of tomorrow's new drugs (i.e., antagonists and stabilizers). Our identification of potential allosteric sites is in agreement with the investigation of interfaces recently reported by Block *et al.* [160] (see below).

## Virtual ligand screening: strengths and limitations

Among the different *in silico* screening methods, one can usually distinguish two main virtual screening strategies: ligand-based screening and structure-based screening (Fig. 4). Most of these high-throughput tools have been recently reviewed and interested readers can find detailed information in [28, 161-171]. Also, several of these methods are freely available to academic groups, they have been recently listed in [94]. The ligand-based and structure-based approaches can be combined and fully integrated with the other experimental methods as shown in some recent reports [172-174] while in the case of PPIs, since in most cases there will not be any known ligand, the use of structure-based (SB-VLS) methods appear more appropriate.

In the ligand-based methods (similarity and substructure search, clustering, QSAR, pharmacophore matching or three-dimensional shape matching), the concept is to use information provided



by compounds that are known to bind to the desired target and to use these data to identify molecules in the databases with (more or less) similar properties (notion of scaffold hopping [175-178]). For structure-based methods (SB-VLS), it is assumed that the three-dimensional (3D) structure of the target is known either by X-ray crystallography or NMR experiments or predicted by homology (see comments about homology modeling and recent methods allowing structural prediction using this approach in [67, 179-183]). The principle with SB-VLS methods is to dock all the ligands (or most of the ligands if the protocol involves a hierarchical procedure) present in a database into the binding pocket of the selected target and evaluate the fit between the molecules. To perform *in silico* screening computations, appropriate compound collections are needed, often the compounds have to be transformed in 3D and some ADME/Tox (for absorption, distribution, metabolism excretion and toxicity) computations may have to be performed and often tuned according to the project (i.e., in case of PPIs, most likely the commonly used ADME/Tox parameters will have to be modified) (Fig. 5). As such, both the ligands and the proteins have to be prepared prior to running *in silico* computations.

Databases of small molecules are needed for basically all virtual screening projects. These compound collections (purchasable molecules, about 10 millions in total in 2007) can be obtained in 2D and in general in SDF format (in some cases in SMILES) from Chemical vendors. Some of these libraries contain only marketed drugs and/or natural products while others are for cherry picking. Many of these collections have been analyzed in depth these last few years [184-190] (Table III). Virtual compound collections can also be generated with the risk that some molecules will not be possible to synthesize easily. In addition, pharmaceutical companies have usually proprietary compound collections that are not available to academic groups. As introduced above, many purchasable compounds or virtual compounds may not have satisfactory ADME/Tox properties to be considered as promising drug candidates. It is then possible to remove from the libraries some of these molecules by computing properties such as: absorption prediction via computation of logP, logD, molecular weight, polar surface area (e.g., if greater than 140 Å<sup>2</sup>, poor intestinal absorption), number of H-bond donors-acceptors, number of rotatable bonds (see Lipinski rule of 5 reviewed in [191]), frequent hitters, reactive chemical groups, etc) using *in silico* methods ([191-199]) (Table III). It is commonly accepted that absorption, distribution and excretion are dependent on similar descriptors while metabolism and toxicity are of a quite different nature and depend of numerous factors. Clearly all *in silico* ADME/Tox prediction methods (this observation also applies to *in vitro* and to animal models) should be considered with care because the human organism is an immensely complicated system and it is not possible at present to simulate accurately all the different events in a computer. Yet, these tools, whether *in silico* or *in vitro* can provide valuable information in many situations.

For a screening project, small molecules usually go through several *in silico* ADME/tox filtering steps in an attempt to generate a database of molecules that have physical properties and chemical functionality consistent with known drugs/leads/hits (i.e., we aim at keeping in the library molecules having characteristics, topological descriptors, physicochemical descriptors (etc), similar to those of known marketed drugs that are, by definition, classified as drug-like). Several suggestions have been made to facilitate the design of a database suitable for *in silico* screening experiments, interested users can find recommendations in the following articles [200-203]. An important concept is that compound collections do not need to contain millions of compounds, smaller, more focused or more diverse collections of higher quality can be much more appropriate for a screening campaign. For some projects, it may be necessary to design a database of small fragments and here also, several “recipes” have been reported [204, 205]. For many *in silico* screening experiments, the small molecules have to be in 3D, several tools are available to generate 3D structures of small molecules [94] while some servers already deliver the collections in 3D (single conformer or multi-conformer and usually ADME/Tox filtered) [206-210] (Table III). Yet, it is important to note that predicting 3D structures, computing charges, multiple protonation states, stereo-

chemistries (e.g., racemic mixtures) and regio-isomeric forms (E/Z isomerism), tautomeric, and conformational states (multiple conformers) for large databases are challenging and obviously not always possible while not always needed, it usually depends on the sensitivity of the methods and algorithms used for the docking-scoring experiments [211-213].

Recently, an *in silico* attempt to extract potential PPI inhibitors from compound collections has been reported. In that study, the authors extracted known PPI data from the literature (25 representative molecules were selected) and computed various descriptors (e.g., molecular shape) for each compound and used machine-learning methods to try to discriminate PPIs inhibitors from non-PPI inhibitors. The results, although preliminary, show that it is possible to enrich a collection 10-fold in PPI inhibitors with the consideration of only a few descriptors [214], but it remains to be seen if a database prepared along this line will effectively facilitate the *in vitro* discovery of PPIs modulators as compared to a non-PPI-filtered collection.

The target 3D structures of proteins usually need to be prepared for SB-VLS experiments. This step often involves the addition of hydrogen atoms and the prediction of the correct protonation state for the titratable residues, adding or removing water molecules, counterions, metal ions, cofactors, sugar molecules, removing subunits not involved in ligand binding or far from the binding site, introducing corrections to the tautomeric states of histidine residues and re-orientations of hydroxyl groups. Pocket predictions, as mentioned above, have to be performed if the zone that needs to be screened is not known experimentally. In the present context of structural genomics and the small number of known drug-like compounds interfering with protein-protein association, it is likely that SB-VLS approaches are the methods of choice to modulate PPIs (Fig. 5). In fact, SB-VLS methods have already been successfully applied to basically all types of targets (experimental structures or homology models) and biological mechanisms: catalytic sites [215, 216], protein-protein interaction [216, 217], protein-membrane interaction [218], protein-DNA interaction [219], protein-RNA interaction [220], protein-heparin interaction [221]. In addition, these computer methods start to be used on extremely complex mechanisms such as protein-misfolding diseases [222]. Yet, in the case of PPIs, it is likely that difficulties will arise on some targets due to possible flexibility of the binding regions and special considerations will have to be taken for this target class (see below). Assuming that the appropriate receptor 3D structure is available, a primary challenge in hit discovery via SB-VLS is to predict both ligand orientation and binding affinity; the former is often referred to as ‘molecular docking’ while the latter is referred to as ‘scoring’ (Table IV). Docking protocols can be described as a combination of two components: a search strategy and a simple scoring/fitting function to assess the poses prior to the real scoring step. The search algorithm should generate an optimum number of conformations for the ligands and of poses that should include the experimentally determined binding mode. The complexity of molecular docking implies several approximations, from rigid body docking (where both partners are treated as rigid but “flexibility” can be generated prior to docking), to (pseudo)-flexible ligand docking (where the receptor is held rigid and the ligand is partially flexible) to flexible docking (where both receptor and ligand flexibility are considered). Algorithms dealing with flexibility from the ligand side can be divided in three types, namely (a) (pseudo-) systematic (e.g., incremental construction algorithms; database methods with libraries containing pre-generated conformations for each molecules), (b) stochastic or random algorithms (e.g., Monte Carlo methods, Tabu search/evolutionary algorithms, they make random changes on some variables and usually require multiple independent runs), and (c) deterministic searches or simulation methods (e.g., energy minimization and molecular dynamics) [223]. Many free tools performing such tasks are listed in a recent review [94] (see also Table IV).

Generating a broad range of binding modes is ineffective without a model to rank each conformation that is both accurate and efficient/fast. The scoring functions commonly used are said to be:

force-field based, empirical and knowledge-based [224-227] while consensus scoring can provide interesting insights/results [228] and that in some situations, quantum mechanics can improve the results [229]. It is however important to note that the more accurate the computations the more sensitive they become to small errors in some parameters (e.g., partial charges) and in the position of the ligands in the binding pocket. In general, some scoring functions perform well on some target classes but not on others. There are no clearly defined rules to predict *a priori* which scoring functions will be successful on a given target although the nature of the binding pocket can give some insights [37] since it governs the overall shape, size and physico-chemical nature of the ligands. A strong emphasis is now being put on post-processing strategies rather than on designing more accurate generic scoring functions. Post-processing can involve consensus docking and/or scoring [161, 228] and tools that facilitate the interactive analysis of the docked molecules (e.g., Post-Dock [230]). Hierarchical protocols that, for instance, involve rigid-body docking followed by flexible ligand docking and possible rescoring using relative free energy computations have been proposed [37, 135, 231-233]. Another way of conceiving post-processing (or pre-processing) is through the notion of fingerprint [234, 235]. In this situation, a 1D binary fingerprint representation of the intermolecular interactions in a 3D protein-ligand complex is computed. For example, amino acids in contact with the ligand are represented as a bitstring with each bit switched to on or off depending whether it is involved in a contact, the contact is with the main chain or the side chains of the protein, and if the interaction is polar, apolar or involve hydrogen-bond donors/acceptors. This approach is implemented in the SIFt package that uses 7 bits for each binding site residue representing 7 different types of interactions [234] but other methods have also been developed like FLAP [235]. The whole interaction fingerprint is then constructed by concatenating the bit strings for each binding site residue. Fingerprints can be used for comparing ligands and learn about target specificity, for ranking and for the construction of focused libraries. These methods are promising in the context of screening large compound collections with relatively modest computer power but additional investigations are needed to validate the superiority of these programs as compared to for instance packages build on decades of physics-based developments.

We and many other research groups have performed comparative studies and the overall picture indicates that no one docking method currently outperforms others across diverse protein targets (see below). In other words, some docking/scoring approaches are very efficient on some targets but not on others [167]. What is generally observed is that docking methods are able to reproduce the experimental poses as seen after comparisons with crystallographic studies but that the scoring does not always allow for the selection of the right binding mode and does have problems with ranking, leading to many high-scoring false positives, thereby lowering enrichment rates. These observations indirectly point to the concept that important information are present in the binding pocket but that they are not appropriately considered with the current docking/scoring methods. In our opinion, structural and physico-chemical properties have to be taken into account during the docking and the scoring steps to improve the process and most likely existing methods are already at hand to perform these tasks. We advocate that now, studies should be undertaken to correlate pocket descriptors with the docking/scoring algorithms. This represents an area of intense investigation in our group (manuscript in preparation) that remains extremely challenging.

Several metrics are commonly used to analyze hit lists including calculation of enrichment factor, hit rate and area under the receiver operating characteristic curve (ROC). The enrichment factor or rate is the increase in the proportion of hits found in any given sample of compounds, compared with the proportion expected from a random sample [236]. This value is often used to benchmark *in silico* methods. If we imagine having 100 actives in a collection of 100,000 compounds, the chance of picking one active at

any toss should be 100/100,000 thus 0.001. If we generate a list of 1000 molecules at random, we should have in theory,  $0.001 \times 1000 = 1$  active. Now, if a list of 1000 molecules is proposed after *in silico* screening and that 10 binders are found experimentally while we know that we have a total of 100 binders in our collection, then the enrichment factor at that point (1% of the database, i.e., our list of 1000 molecules) equal 10, which is the number of binders found divided by the number of known binders expected from random selection (here equal to 1). EF is equal to:  $\{\text{found binders}/\text{number of molecules in the short list selected } in\ silico\}/\{\text{total number of binders}/\text{total number of compound in the collection}\}$ . In practice, achieving a good enrichment rate of active compounds in the chosen subset is more important than predicting the exact binding affinity of individual compounds, this can be attempted at a later stage. For hit rate, if we imagine that we test experimentally 50 molecules and, that among them, 7 are bioactive compounds, then the hit rate = 14% (i.e.,  $7/50 \times 100$ ). ROC curve and area under the ROC curve can also be used to assess screening studies in addition to or instead of EF [237]. Yet, Truchon and Bayly suggest that none of the methods used so far are really appropriate to evaluate *in silico* screening and have recently derived new mathematical formulas [238]. It will be interesting to see in the future, if one can really improve the understanding of *in silico* screening procedures by using these new equations. Along the same line but concerning the evaluation of docked poses, the accuracy is most of the time assessed by computing the RMS deviation between the predicted pose and the crystal structure. This approach is sometimes misleading and several authors have proposed to evaluate the presence of key interactions between the docked ligands and the protein versus the ones present in the crystal structure. One method of this type has been coined IBAC, for Interactions-Based Accuracy Classification, it highlights that in a number of cases, there are important discrepancies between IBAC and RMSD assessments. The suggestion is that IBAC-like evaluations are more meaningful to measure docking accuracy than RMS deviation [239].

Many benchmarking studies have been performed these last few years in an attempt to evaluate the performance of the diverse *in silico* screening methods/protocols [240]. Yet, validation of a method is extremely complex and subject to errors. A compound collection may not be proper to assess a method (e.g., too small, the actives can be very different from the others and be easily identified...), the docking/scoring parameters are optimized in some studies while they are not in some others making comparisons relatively unfair. The main message from these studies is that some packages are successful on some targets while fail on others but the reasons for success and failure are unclear because of the complex interplay among the many parameters acting on the docking and scoring steps (see for instance [167, 168]). Another point that we noticed with regard to *in silico* screening studies is the race for “hit rate”, something that could be called “over hit rate”. More and more studies report very high hit rates that do not reflect the actual strengths and limitations of *in silico* screening methods. For instance, SB-VLS methods carried out on a well-known pocket for which many co-crystal structures are known (protein with drug-like ligands co-crystallized, not peptides) can yield a hit rate of 50% or more. This gives the impression that studies with hit rate of 1% or less have failed. However, in these latter cases, it could be that the binding pocket is not well-defined, that it is filled with water molecules, that several amino acid side chains could be flexible and that co-crystals of the target with a drug-like ligand are not available. Also, the final selection is usually performed through visual inspections and authors may want to take the risk of selecting compounds that have little chance of binding as seen on the computer screen but that are originals and presenting with interesting scaffolds. Thus, screening more molecules with an extra list of risky compounds will most of the time damage hit rates but could be very valuable from a biological standpoint. In all cases, a hit rate of 0.5% or less as obtained after *in silico* computations could still be interesting as compared to published HTS hit rates that can easily be around 0.01% or less. Thus, hit rates have to be examined with care.

As mentioned above, a major problem with some targets is due to flexibility (or plasticity or malleability) in the area of the interface [241]. Indeed, with regard to protein-protein complexes, only about one third can be docked without considerations of induced conformational changes [242], further highlighting the need of addressing this issue. The inadequacies or limitations of some experimental structures can be overcome in some cases by using simulation methods (e.g, molecular dynamics, normal mode analysis, Gaussian network models, Monte Carlo simulations) [243-245] (Fig. 6). Flexible receptor-flexible ligand methods suited for *in silico* screening experiments and protocols are emerging (see for example [180, 246-251]), but clear successes are still missing because flexibility, if not properly treated, can increase the “noise” instead of facilitating identification of relevant hits. The first approach to try to address the flexibility problem has been the use of soft potentials, where some atom overlaps are tolerated [216]. Another way is to dock on multiple structures (experimental or simulated) of the receptor. The ultimate approach would be to explicitly consider protein and ligand flexibility during docking. Yet, this latter approach is very computationally demanding and, for the time being, basically no packages are able to address the problem efficiently. The difficulties of performing docking against multiple receptor conformations lie with how to deal with the ranked lists since for each structure of the receptor, the docking/scoring procedure can propose over 1000 likely molecules. If computations are performed on 3 structures, one may have a list of 3000 molecules to test experimentally with possibly very few consensus compounds. The merge-and-shrink procedure has been recently proposed to address this problem [248]. With this approach, for each compound, the best rank -not the best score- is kept. This was shown efficient on some targets but, in general, to use this approach, one needs known actives to test the protocol and calibrate the computations. The merge-and-shrink protocol or the consensus approach [37] can be efficient in some cases or dilute the correct results for some targets. As such, much work is needed to address this difficult question of flexibility from the receptor side. Interestingly, a recent molecular dynamics study investigating transient pockets on protein surfaces directly located at the interfaces has been reported [252]. The authors investigated Bcl-xL, IL-2 and MDM2 and propose a new pocket detection protocol. They observed that the native inhibitor-binding pocket was partly present in the unbound proteins. They performed 10 ns molecular dynamics simulations and scanned, on many structural snapshots collected along the trajectory, the surface of the proteins, in search of transient or more permanent cavities. They noticed that some cavities were opening up within 2.5 ps and that most of them appeared multiple times. In the druggable region, they noticed that pockets of similar size as compared with the experimental structures in complex with the ligand could be generated. Docking was successfully performed into these transient pockets, suggesting that this protocol is viable to study protein interfaces with the goal of designing antagonists and possibly stabilizers. An interesting point noted by Luque and Freire about catalytic sites is that these regions appear to have a dual nature characterized by regions of high stability and regions of high flexibility [253]. Analysis of pockets at the interfaces or near the interfaces is required to confirm this observation for PPIs.

## Case studies

PPIs are challenging but attractive targets for small chemical drugs. By focusing on PPIs, the number of latent and novel drug targets can be expected to increase dramatically. This is because the size of the human interactome is much larger than that of the human proteome. With a human genome coding for about 25,000-40,000 proteins, the size of the interactome is estimated to be around 250,000 based on extrapolation from the yeast interactome. However, for the time being, the number of PPIs registered in the public interactome database is limited to about 40,000, suggesting that many interactions have yet to be discovered [254-257].

Several protein-protein interactions have been successfully targeted using chemistry-*in silico-in vitro* screening approaches (and other biophysical methods such X-ray crystallography [258, 259] or NMR spectroscopy [26, 27, 30, 260]) and have been extensively reviewed these last few years [1, 38-40]. Well documented examples include, the Bcl-xL-BH3 protein-protein interaction (cancer, extended compounds mimic the BH3 alpha-helical structure, difficulties come from the flexibility of some regions of the receptor structure, large binding groove) [261], inhibitors of the p53-MDM2 interaction (cancer, relatively small pocket on MDM2, the compounds have to mimic an alpha-helix from p53, it does not seem that important structural changes occur in the binding pocket) [262], modulation of the Keap1 protein-protein interaction (application in, for example, cancer, Parkinson's disease, relatively compact and exposed pocket, somewhat similar to MDM2, the pocket binds to a loop-like structure) [263], the BIR 3 domain of XIAP (cancer, not very pronounced binding groove binding a somewhat extended amino-acid segment from the other protein partner) [264, 265], inhibitors of the IL-2alpha receptor-IL-2 interaction (immunosuppressive agents, not very pronounced binding groove on IL-2 undergoing conformational changes, this region interacts with a somewhat extended amino-acid segment from the receptor) [266, 267], the G protein-coupled receptors (cancer among others, somewhat narrow binding pocket, *in silico* screening usually performed on homology models of not very high "resolution" but integrating numerous ligand-based results biasing the docking experiments and thereby usually leading to relatively high hit rate) [268-271] and the Arf-GDP-Sec7-BFA system, an example of uncompetitive interfacial inhibition in which a drug takes advantage of transient structural conditions created by the macromolecular complex (molecular mechanisms difficult to handle with *in silico* approaches) [272]. In the following, we decided to focus on four recent studies combining *in silico-in vitro* screening that illustrate different strategies and types of PPIs.

### Stat3

Proteins of the STAT family are activated in response to cytokines and growth factors. They promote proliferation, survival and many other biological processes [273]. Upon tyrosine phosphorylation, dimers of STATs form between two monomers and they translocate to the nucleus and bind to specific DNA-response elements in the promoters of target genes to finally induce gene expression. Aberrant activity of one of the member, Stat3, contributes to carcinogenesis and tumor progression. A critical step in STAT activation is the dimerization process. This step relies on the reciprocal binding of the SH2 domain of one monomer to the PpYLKTK sequence of the other Stat3 monomer. The crystal structure of the Stat3 homodimer has been reported [274] and used for *in silico* screening studies in order to search for potential disruptors of the Stat3-Stat3 dimer formation [275-277] (Fig. 7).

In the study of Siddiquee et al., [275], the virtual screening was performed with GLIDE [278] and a compound collection of about 155,000 molecules (prepared with LigPrep) obtained from the NCI's chemical libraries (compound 3D structures generated by Corina). One compound was found to disrupt Stat3-Stat3 complex formation *in vitro* and intact cells (total number of compounds tested experimentally not clearly stated). In the study of Song et al. [277], the same X-ray structure was used and a chemical database containing compounds from Sigma-Aldrich, Merck Index, Ryan Scientific and the NCI was used initially. The four databases offered a final collection of about 429,000 organic molecules (compounds in 3D generated by Corina). The *in silico* screening was performed with DOCK and the top 10% scored compounds from each database were extracted and re-score with X-SCORE [279, 280]. Of the best-scored 200 compounds, 100 molecules were tested experimentally and one molecule showed remarkable inhibition of Stat3 dimerization. From an *in silico* standpoint, the hit rates can be considered as low while from a biological standpoint, these studies can be considered as very successful. Definitely, the pocket tends to be relatively flat, contain a mixture of hydrophobic/aromatic and charged residues and the region

could be flexible (e.g., side chains or main chain) making the docking/scoring difficult (a short and extended peptide segment from the other monomer fits in the pocket and it is thus not an helix as often believed, not all protein interactions involve helices from one partner and a groove in the other partner). Also, the lack of small molecule co-crystallized in the binding pocket does not allow for pose assessment, ligand-based pre-screening and direct calibration of the scoring function according to known binders. These studies reflect a likely situation for most *in silico* screening projects targeting PPIs. It is interesting to note that one package, often considered as one of the best in several benchmark studies, GLIDE, does not perform better than DOCK [281], a program reported to be efficient on some targets but in average less efficient than GLIDE [158]. Of course, direct comparison cannot be performed in this way because the compound collections used are different, but this still illustrates (as suggested by most authors of benchmarking studies) that, although evaluation/comparison studies of *in silico* tools are of utmost importance, one should take the information with cautions. Stat3 is a difficult target and in this case, it seems reasonable to test more molecules experimentally.

### **Antitrypsin polymerization**

$\alpha_1$ -antitrypsin is the archetypal member of the serine proteinase inhibitor or serpin superfamily. Most individuals carry a normal allele but 4% of the Northern European population are heterozygous for the severe Z deficiency variant (this translates to a Glu342Lys substitution and subsequent perturbations of secondary structure elements and of the mobile reactive center loop). The resulting unstable intermediate molecules then link to form loop-sheet polymers in which the reactive loop of one molecule inserts as strand 4 of the so-called  $\beta$ -sheet A of another  $\alpha_1$ -antitrypsin molecule. It is the accumulation of polymers that predisposes the Z  $\alpha_1$ -antitrypsin homozygote to hepatitis, cirrhosis and hepatocellular carcinoma. Other mutations in other serpins also lead to polymer formation leading to thrombosis, angioedema, emphysema or familial encephalopathy [282, 283]. With regard to  $\alpha_1$ -antitrypsin deficiency, the only curative treatment available today is liver transplantation. Recently, Mallya et al. [284] have performed SB-VLS computations with the ICM package against a non-redundant library of about 1.2 million commercial drug-like compounds in a hydrophobic pocket present in this serpin but somewhat remote from this abnormal protein-protein interaction site (Fig. 8). An initial screen was performed on the crystal structure of  $\alpha_1$ -antitrypsin (PDB code 1qlp) [285] but none of the nominated compounds were effective in impeding polymerization. It should be noted that the pocket is challenging: it is located in a flexible region and can be considered as an allosteric cavity. Amino acid side chain simulation was performed to identify the most flexible residues lining the cavity and one Asn residue was found to be very flexible. A low energy conformation for this residue was generated through Monte Carlo simulation. SB-VLS was performed again and 68 compounds were selected for further characterization *in vitro*. One compound completely blocked polymerization while another one presented with interesting activity as it allowed the formation of only dimeric structures. Ligand-induced changes were then modeled for these 2 best compounds prior to an additional round of SB-VLS experiments. These changes were achieved by performing internal coordinate Monte Carlo side chain simulations in the presence of each ligand in its predicted binding pose. Nineteen extra compounds were selected after re-docking on this modified receptor structure, 10 of them were structurally similar to the original two active compounds, suggesting that ligand-based approaches could also have been performed instead of docking-scoring. Some of these molecules were as active as the best molecule identified in the previous screen. Thus, one compound was found with the desired biological activity, opening new ways to block polymer formation and strategies to block protein-protein interaction in the serpin superfamily (i.e., yet the compound appears to be specific of  $\alpha_1$ -antitrypsin and does not act on other members of the family). Many diseases are characterized by protein misfolding, aggregation and polymer formation. In general, drug-like molecules able to interfere with this

process are obtained after HTS, like for examples, for expanded polyglutamine proteins, PrP or tau [222]. It is thus remarkable to be able to use SB-VLS approaches against these complex mechanisms.

### **$\beta$ -Catenin-T-cell factor complex**

$\beta$ -Catenin is a protein with multiple functions that binds to T-cell factor (Tcf) and activates transcription of proteins that regulate the cell cycle. Disruption of this interaction is considered to be a valid anticancer strategy. The complex between  $\beta$ -Catenin and Tcf is characterized by an unusually large interface with a binding constant  $K_d \sim 10$  nM.  $\beta$ -Catenin is a superhelix with a large and shallow positively charged group and the N-terminal residues of Tcf bind along this groove. Virtual ligand screening was performed by Trosset et al. [286] with FLO\_QXP package (4 amino acid side chains were allowed to move during the docking) using an ADME/Tox filtered library from the Pharmacia and Upjohn collection (initial size, 90,000 molecules and 17,000 after filtering). Visual inspection of the top ranking 3000 compounds led to the selection of 42 candidates and finally 22 molecules were tested using NMR WaterLOGSY and isothermal calorimetry experiments leading to 3 confirmed hits and a very high hit rate of 14% given the context and lack of information about the pocket (i.e., the exact binding mode is not yet known nor if the compounds really bind to the predicted pocket). The  $\beta$ -Catenin-Tc3 crystal structure was used for the study (PDB code 1g3j) [287], 6 pockets were predicted at the surface of  $\beta$ -Catenin and one area, around residue Lys435-Arg469 contained several clefts that could potentially anchor a small molecule (this region interacts with a loop and strands from Tcf) (Fig. 9). The authors comment on the obtained hit rate and clearly mentioned that hit rates are not the most critical parameters, rather, the key point is to find active compounds in a reasonable amount of time and reduced cost as compared to massive HTS campaigns.

### **Plant $H^+$ ATPase-14-3-3 protein complex**

Recently Block et al. [160] have been trying to design protein-protein stabilizers, arguing that it could be easier from a thermodynamic standpoint than finding antagonists. They also suggested that to find PPI antagonists via *in silico* approaches, it is necessary to identify a druggable pocket at the interface, a situation that is unlikely to occur. These authors investigated about 200 complexes and found relatively frequently a cavity, close to the interface but not at the interface, suggesting that these sites could be interesting allosteric sites. From our initial investigation of experimental complexes, we, to the opposite, find that a druggable pocket is very often present at the interface although we also noticed cavities near the interface. In order to carry out a proof of concept study, Block et al. [160] decided to search stabilizers of the interaction between plant  $H^+$ -ATPase and 14-3-3 protein. This interaction is boosted by a fungal phytotoxin named Fusicoccin. This small but relatively complex molecule is an herbicide with applications in the agrochemical industry. Yet, it is difficult to synthesize and since even minor chemical modification impairs its biological activity and its potency to stabilize the  $H^+$ -ATPase-14-3-3 complex, finding new chemicals to stabilize this interaction was judged interesting. SB-VLS investigations were performed, to both, find stabilizers of this system and serve as a proof of concept. The experimental structure of the 14-3-3 protein (PDB code 1o9f) [288] is known, this structure contains the co-crystallize Fusicoccin and a short peptide that mimics the ATPase (Fig. 10). The pocket is localized next to one residue of the ATPase peptide and contains four water molecules. Virtual ligand screening was performed using a compound collection of about 2 million compounds that was downsized after some filtering steps (e.g., only molecules with up to 6 rotatable bonds, molecular mass between 250 and 500 Da and some other 3D pharmacophore filtering methods), to about 160,000 molecules. These ones were docked with FLEXX, GOLD and AutoDock and a post-processing step was then performed according to a pharmacophore hypothesis and visual inspection. Seventeen compounds were selected for *in vitro* testing (surface plasmon resonance) and none were found to be efficient in stabilizing the complex.



The authors had to face many difficulties in that study. First, the binding pocket is possibly flexible, it may be different when the complete ATPase is present and it contains water molecules. It is likely that to design protein-protein stabilizers with *in silico* approaches, an experimental structure of the complex and simulations are required and that more compounds would need to be tested. In contrast, it seems possible to design PPI inhibitors even if one structure from only one partner is available, a situation that is more probable at present. Furthermore, while Block et al. [160] suggest that blocking protein-protein interaction should not be really feasible in many cases because of thermodynamic considerations, it does seem that designing stabilizers is even more challenging and most likely equally difficult with regard to energetics. Also, the small compound will have to carry out a very difficult mechanistic task (e.g., stabilize one structure, stabilize the interface, bridge the 2 partners), and, when screening such an exosite next to the interface region, it cannot be known, a priori, if the small molecule will induce some structural changes that stabilize or impede the interaction.

The need of stabilizers is well documented in a wide range of diseases known as amyloidoses. These are also known as protein misfolding diseases since in all cases, a protein normally soluble, becomes insoluble. The situation usually involved some conformational changes, partial unfolding, and protein-protein interaction. About 20 proteins have been related to different forms of amyloidosis, the only common characteristic shared by many of these molecules is their high content in  $\beta$ -structure [289]. The concept of stabilizing a protein complex is clearly of interest although extremely challenging and involving many different underlying mechanisms. For example, stabilization by small molecules was achieved for the transthyretin protein (TTR)[290]. TTR is composed of four identical subunits (homotetrameric protein), each monomer is organized as two sheets of four  $\beta$ -strands. Two monomers associate and the tetramer seems to result from the association of two dimers, leading to a central channel in the protein. This one is mainly hydrophobic and can accommodate two thyroxine molecules. TTR can aggregate, in this case tetramer dissociation must occur prior to aggregation and fibril formation. Some small molecules that bind in the thyroxine-binding channel seem to favor the stabilization of the native tetrameric conformation of TTR and may inhibit the formation of the amyloidogenic intermediate species that lead to amyloid [291-294]. Along the same line and concept of stabilization, superoxide dismutase 1 (SOD1), like transthyretin, can undergo subunit dissociation and aggregation as a result of mutations. SOD1 is a homodimer that displays a small binding pocket at the interface between the two monomers. The structure was screened *in silico* using the Glide package, 100 molecules were selected after screening several commercial databases and 15 compounds apparently stabilized the system, indicating that stabilizers can be obtained using SB-VLS methods [295]. Yet, in this case, the pocket is clearly defined at the interface and does not seem to be an allosteric cavity, the small compounds appear to glue the two monomers, suggesting that an experimental structure of the entire system was required.

## Conclusions and outlook

In spite of present limitations, available *in silico* screening methods can really contribute to the drug discovery process. Appropriate combination/integration of SB-VLS/ligand-based screening and medium to HTS approaches should thus play a major role in the coming years and hopefully contribute to the eradication of complex diseases. A very recent investigation on the HIV-1-Nef complex illustrates further this view [296]. We have noted that *in silico* methods are now established as a powerful alternative and complement to experimental HTS and that they allow for the identification of novel hits with novel chemotypes. Yet SB-VLS methods are complex and as such, a thorough knowledge of the programs, protocols (e.g., multi-step, consensus scoring) is in general a requirement for success in any drug discovery

project. In spite of these strengths, several aspects have to be improved, like the scoring step and a better handling of flexibility, a point that could be critical to the design of PPI modulators. PPIs are usually considered difficult but may in fact be druggable. Hot-spots can be probed and in some cases predicted with combined experimental-*in silico* methods while druggable pockets in the region of the interface can be simulated. The difficulties that are still present to design PPI modulators can be divided into two different categories: (a) appropriate chemical space and (b) right protocol. For category “a”, difficulties with finding PPIs modulators can be due to the fact that appropriate lead compounds may not be present in the libraries, suggesting that it will be necessary to explore new areas of the chemical space. Another reason for the difficulties with PPIs (b) could be that we do not have the right protocols to properly handle this task, and some of the examples discussed in this review support this hypothesis. We have seen that a major bottleneck in some virtual screening projects is that the conformation of the target protein as found in a crystal structure or in theoretical models, be it bound within a macromolecular complex or unbound (apo form), may not be truly representative of the dominant conformation that will bind to the small drug-like ligands. Introduction of flexibility on the receptor side during high-throughput docking computations remain a major problem with no simple or single solutions. Indeed, incorporation of the receptor flexibility is a formidable challenge. Flap motion, side-chain movements, rotation of subdomains and movement of some secondary structure elements (i.e., helices) are well-documented [297, 298]. However, prediction of these movements is extremely difficult in the context of high-throughput ligand docking. There are also other types of flexibility, ranging from Gln/Asn O/N re-orientation upon or prior to ligand binding to loop flexibility that are also very difficult to handle. Water molecules, cofactors and metal ions can also be troublesome. There is room for improvements in the scoring field and possibly better scoring according to the nature of the binding pocket seem feasible. Furthermore, finding the right balance polar and non-polar interactions and better accounting of solvation and entropy should be possible. It is also reasonable to think that a deeper understanding of the existing tools will enhance their possibilities. Several examples of successful modulation of PPIs via combination of *in silico-in vitro* screening have been presented and it is clear that new success stories will be reported in a near future. While the existing *in silico* methods can address some PPIs, it seems also likely that given our lack of knowledge about these interactions, more compounds should be tested experimentally after *in silico* screening as compared to for instance catalytic sites. Although it is often mentioned that PPIs can only be addressed efficiently via NMR and fragment-based approaches, we believe, after reviewing the literature, analysis of our own results and as we witness rapid methodological progresses in the field of computational biology and chemistry, that existing *in silico* tools can already perform such a task but that new protocols will have to be designed. In the examples presented here, in most cases, the hit rates from an *in silico* perspective are not as good as the ones obtained on catalytic sites, a situation that is somewhat expected. Yet, from a biological standpoint, these studies have usually identified very interesting molecules. Overall, in the future, we should get a better knowledge of what organic structural moieties are privileged or preferred at PPI sites and ultimately, we should succeed through targeting PPIs to develop the next generation of therapeutics. By merging the know-how of medicinal chemistry with the new technologies and better knowledge of diseases, by learning from failures and through efforts and tenacity, major accomplishments should be obtained for this target class in a near future.

### **Acknowledgments**

Supports from the Inserm institute and Inserm-Transfert are greatly appreciated.

**Table I: Examples of recently reported tools to predict hot-spots, analyze interfaces and for protein-protein docking**

<b>URLs</b>	<b>Short summary</b>	<b>Ref.</b>
<a href="http://www.molsoft.com/oda.cgi">http://www.molsoft.com/oda.cgi</a>	ODA (Optimal Docking Areas) is a method to predict protein-protein interaction sites on protein surfaces. It identifies optimal surface patches with the lowest docking desolvation energy values	[96]
<a href="http://www.bioinformatics.leeds.ac.uk/ppi_pred/">http://www.bioinformatics.leeds.ac.uk/ppi_pred/</a>	PPI-PRED can be used for protein-protein binding site prediction and DNA binding site prediction	[299]
<a href="http://www.bioinformatics.sussex.ac.uk/SHARP2">http://www.bioinformatics.sussex.ac.uk/SHARP2</a>	SHARP <sup>2</sup> predicts the location of protein interaction sites on the surface of the 3D structure of a protein	[300]
<a href="http://crick.mbu.iisc.ernet.in/~PIC">http://crick.mbu.iisc.ernet.in/~PIC</a>	PIC: Protein Interactions Calculator	[100]
<a href="http://pfp.technion.ac.il">http://pfp.technion.ac.il</a>	Patch Finder Plus (PFplus): a web server for extracting and displaying positive electrostatic patches on protein surfaces	[101]
<a href="http://bip.weizmann.ac.il/promate">http://bip.weizmann.ac.il/promate</a>	ProMateus: protein binding site analysis	[103]
<a href="http://3D-partner.life.nctu.edu.tw">http://3D-partner.life.nctu.edu.tw</a>	3D-partner: a web server to infer interacting partners and binding models	[104]
<a href="http://sage.csb.yale.edu/sitefinder3d">http://sage.csb.yale.edu/sitefinder3d</a>	siteFiNDER3D: a web-based tool for predicting the location of functional sites in proteins	[105]
<a href="http://structure.pitt.edu/servers/fastcontact/">http://structure.pitt.edu/servers/fastcontact/</a>	FastContact: a free energy scoring tool for protein-protein complex structures	[93]
<a href="http://vakser.bioinformatics.ku.edu/resources/gramm/grammx">http://vakser.bioinformatics.ku.edu/resources/gramm/grammx</a>	Gramm: Tools for protein-protein docking GrammX: web interface of Gramm	[301]
<a href="http://bioinfo3d.cs.tau.ac.il/">http://bioinfo3d.cs.tau.ac.il/</a>	Protein docking tools (PatchDock) and related tools. PatchDock, webserver for macromolecules and small molecules docking based on shape complementarity criteria	[120, 302]
<a href="http://mmb.pcb.ub.es/PyDock/">http://mmb.pcb.ub.es/PyDock/</a>	PyDock: tool for protein-protein docking	[114]
<a href="http://nrc.bu.edu/cluster/">http://nrc.bu.edu/cluster/</a>	ClusPro: protein-protein docking webserver using 3 docking programs - DOT, ZDOCK, GRAMM	[303]
<a href="http://www.nmr.chem.uu.nl/haddock/">http://www.nmr.chem.uu.nl/haddock/</a>	HADDOCK software: models any kinds of biomolecular complexes and multi-component (N > 2) systems.	[117, 304]
<a href="http://structure.bu.edu/index.html">http://structure.bu.edu/index.html</a>	PIPER: An FFT-Based Protein Docking Program with Pairwise Potentials	[305]
<a href="http://dags.stanford.edu/InSite/software.html">http://dags.stanford.edu/InSite/software.html</a>	InSite: a computational software that integrates high-throughput protein and sequence data to infer the specific binding regions of interacting protein pairs	[306]

**Table II: Some free tools to predict druggable pockets**

<b>URLs</b>	<b>Short summary</b>	<b>Ref.</b>
<a href="http://www.bioinformatics.leeds.ac.uk/qsitefinder">http://www.bioinformatics.leeds.ac.uk/qsitefinder</a> <a href="http://www.bioinformatics.leeds.ac.uk/pocketfinder">http://www.bioinformatics.leeds.ac.uk/pocketfinder</a>	Server to predict binding sites, Q-site and Pocketfinder	[307]
<a href="http://interface.bioc.columbia.edu/screen">http://interface.bioc.columbia.edu/screen</a>	SCREEN: Server to predict binding site	[159]
<a href="http://medock.csie.ntu.edu.tw/">http://medock.csie.ntu.edu.tw/</a>	MEDock: Online tool to define binding site	[308]
<a href="http://bioserv.rpbs.jussieu.fr/Help/FAFDrugs.html">http://bioserv.rpbs.jussieu.fr/Help/FAFDrugs.html</a>	PASS: Pocket detection method	[206, 309]
<a href="http://firedb.bioinfo.cnio.es/Php/FireStar.php">http://firedb.bioinfo.cnio.es/Php/FireStar.php</a>	FireStar: small molecule binding site prediction	[310]
<a href="http://www.modlab.de">http://www.modlab.de</a>	PocketPicker: tool to predict binding pocket, has to be used with PyMol	[311]
<a href="http://proline.physics.iisc.ernet.in/pocketdepth">http://proline.physics.iisc.ernet.in/pocketdepth</a>	PocketDepth: tool to predict binding pocket	[312]
<a href="http://bioinformatics.cm-uj.krakow.pl/activesite">http://bioinformatics.cm-uj.krakow.pl/activesite</a>	Online recognition of a ligation site in a protein molecule based on the fuzzy-oil-drop model	[313]
<a href="http://scoppi.biotec.tu-dresden.de/pocket/">http://scoppi.biotec.tu-dresden.de/pocket/</a>	LIGSITEcsc: a web server for identification of pockets on protein surfaces	[314]

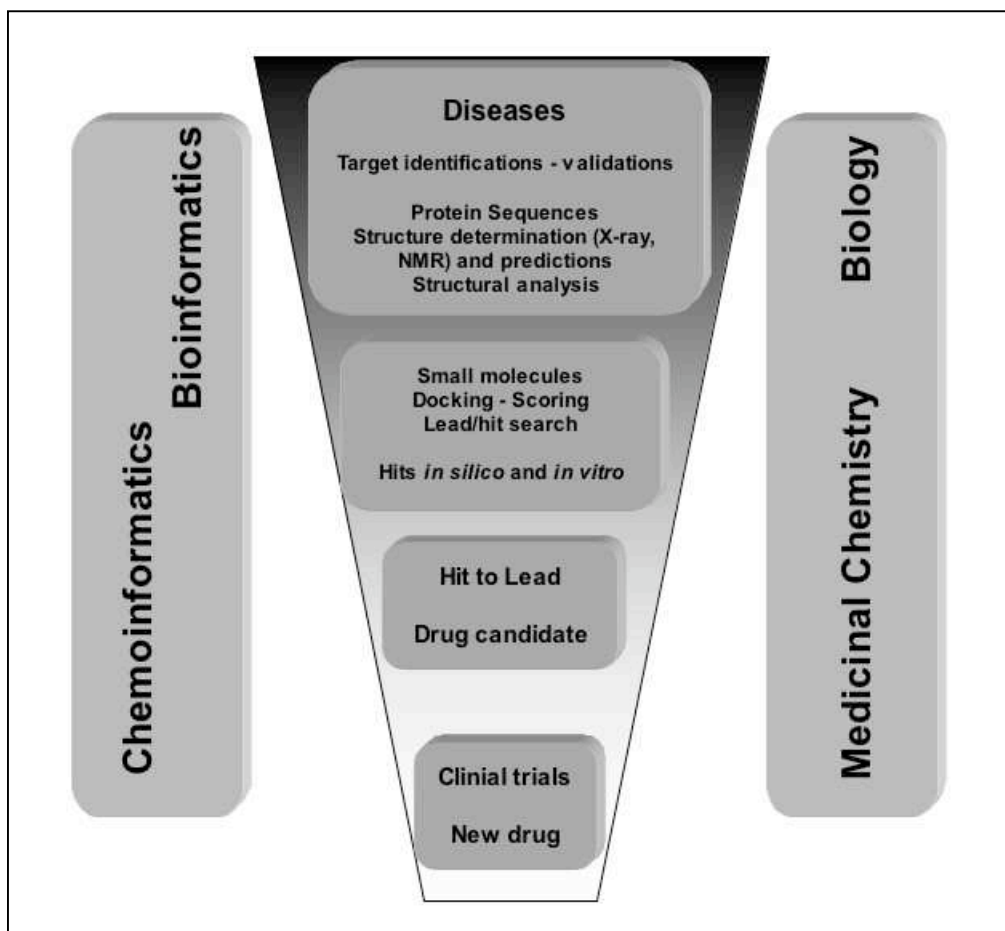
**Table III: Some free compound collections, 2D-to-3D methods and ADME/Tox tools**

URLs	Short summary	Ref.
<a href="http://chembank.med.harvard.edu">http://chembank.med.harvard.edu</a>	ChemBank: Free collections and utilities, known drugs, many annotated molecules, molecules with druglike and non-druglike properties	none
<a href="http://pubchem.ncbi.nlm.nih.gov/">http://pubchem.ncbi.nlm.nih.gov/</a>	PubChem: An information resource linking chemistry and biology	none
<a href="http://www.chemnet.com/">http://www.chemnet.com/</a>	Chemical suppliers and collections	none
<a href="http://www.rxlist.com/02top.htm">http://www.rxlist.com/02top.htm</a>	Top 200 prescriptions in 2002 (structure and name of the compounds)	none
<a href="http://www.bioscreening.com/compound_libraries.htm">http://www.bioscreening.com/compound_libraries.htm</a>	Web directory about compound collections and many related links, database search	none
<a href="http://cdb.ics.uci.edu/CHEM/Web/">http://cdb.ics.uci.edu/CHEM/Web/</a>	ChemDB: Free collections and utilities such as similarity search	none
<a href="http://bioserv.rpbs.jussieu.fr/Help/FAFDrugs.html">http://bioserv.rpbs.jussieu.fr/Help/FAFDrugs.html</a> (see also: <a href="http://www.vls3d.com/">http://www.vls3d.com/</a> )	FAF-Drugs: Free collections in 3D (and ADME/tox) and utilities	[206]
<a href="http://zinc.docking.org">http://zinc.docking.org</a>	ZINC: Free collections in 3D (and ADME/tox)	[208]
<a href="http://bioweb.ucr.edu/ChemMine">http://bioweb.ucr.edu/ChemMine</a>	ChemMine: Free collections and similarity search utilities	[209]
<a href="http://redpoll.pharmacy.ualberta.ca/drugbank/">http://redpoll.pharmacy.ualberta.ca/drugbank/</a>	DrugBank: Numerous data about drugs and targets including drugs already in use	[210]
<a href="http://ntp.niehs.nih.gov/index.html">http://ntp.niehs.nih.gov/index.html</a>	The US National Cancer Institute collections including natural products	none
<a href="http://www.emolecules.com/">http://www.emolecules.com/</a>	With eMolecules you can draw your chemical structure and instantly search millions of molecules from across the Web and from chemical suppliers worldwide	none
<a href="http://chem.sis.nlm.nih.gov/chemidplus/">http://chem.sis.nlm.nih.gov/chemidplus/</a>	ChemID Plus: chemical name, physical and toxicological properties	none
<a href="http://llama.med.harvard.edu/~jklekota/QueryChem.html">http://llama.med.harvard.edu/~jklekota/QueryChem.html</a>	QueryChem: searches public databases using text and structure	[315]
<a href="http://relibase.ebi.ac.uk">http://relibase.ebi.ac.uk</a>	2D to 3D conversion	[316]
<a href="http://www.molecular-networks.com/online_demos/corina_demo.html">http://www.molecular-networks.com/online_demos/corina_demo.html</a>	Corina: 2D to 3D conversion	none
<a href="http://www.eyesopen.com">http://www.eyesopen.com</a>	Omega: 2D to 3D conversion	none
<a href="http://www.molsoft.com">http://www.molsoft.com</a>	ICM: 2D to 3D conversion	none
<a href="http://davapc1.bioch.dundee.ac.uk/programs/prodrg/">http://davapc1.bioch.dundee.ac.uk/programs/prodrg/</a>	Possible 2D to 3D	[317]
<a href="http://bioserv.rpbs.jussieu.fr/Frog.html">http://bioserv.rpbs.jussieu.fr/Frog.html</a>	Frog: a FRee Online druG 3D conformation generator	[318]
<a href="http://edetox.ncl.ac.uk/">http://edetox.ncl.ac.uk/</a>	Find compound properties	none
<a href="http://www.eyesopen.com">http://www.eyesopen.com</a>	Filter: ADME/tox online	none
<a href="http://www.molinspiration.com/">http://www.molinspiration.com/</a>	ADME/tox online	none
<a href="http://www.molsoft.com/mprop">http://www.molsoft.com/mprop</a>	ADME/tox online	none
<a href="http://www.chemaxon.com/products.html">http://www.chemaxon.com/products.html</a>	ADME/tox online	none
<a href="http://sw16.im.med.umich.edu/software/xtool/">http://sw16.im.med.umich.edu/software/xtool/</a>	Xscore : logP computation tool	[319]
<a href="http://www.logp.com/">http://www.logp.com/</a>	Compute logP	none
<a href="http://preadmet.bmdrc.org/preadmet/index.php">http://preadmet.bmdrc.org/preadmet/index.php</a>	PreADMET is a web-based application for predicting ADME data	none
<a href="http://www.niehs.nih.gov/nct/cebs.htm">http://www.niehs.nih.gov/nct/cebs.htm</a>	Chemical Effects in Biological Systems (CEBS) knowledge base (application of systems biology to ADME/Tox)	National Center for Toxicogenomics (none)

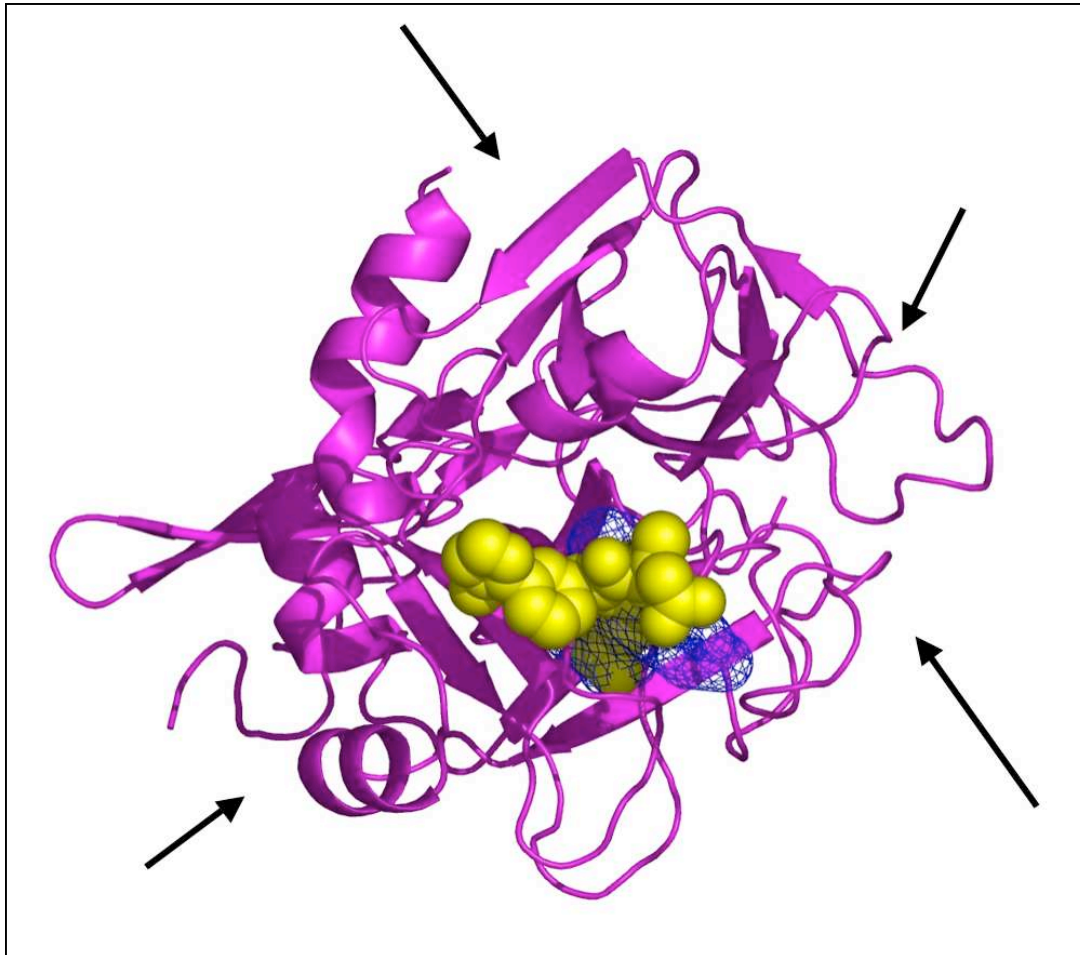
**Table IV: Some free tools for small ligand docking and SB-VLS**

<b>URLs</b>	<b>Short summary</b>	<b>Ref.</b>
<a href="http://www.scripps.edu/mb/olson/doc/autodock">http://www.scripps.edu/mb/olson/doc/autodock</a>	AutoDock, small molecule docking (rigid receptor although some side chains can move, flexible ligand)	[320]
<a href="http://www.simbiosys.ca/ehits/index.html">http://www.simbiosys.ca/ehits/index.html</a>	eHits: small molecule docking (rigid receptor, flexible ligand)	[321]
<a href="http://mgltools.scripps.edu/Members/yongzhao/home-of-flipdock/">http://mgltools.scripps.edu/Members/yongzhao/home-of-flipdock/</a>	FLIPDock: docking flexible ligands into flexible receptors	[247]
<a href="http://www.eyesopen.com">http://www.eyesopen.com</a>	FRED: small molecule rigid docking tool	[322]
<a href="http://www.dddc.ac.cn/tarfisdock/">http://www.dddc.ac.cn/tarfisdock/</a>	tarFisDock docks ligands into the proteins targets in PDTD (Potential Drug Target Database), and outputs the top 2%, 5% or 10% candidates ranked by the energy score, including their binding conformations and a table of the related target information	[323]
<a href="http://dock.compbio.ucsf.edu/">http://dock.compbio.ucsf.edu/</a>	DOCK: small molecule docking tool (rigid receptor, flexible ligand)	[281]
<a href="http://www.biopharmics.com">http://www.biopharmics.com</a>	Surflex: small molecule docking (rigid receptor, flexible ligand)	[324, 325]
<a href="http://www.tcd.uni-konstanz.de/research/plants.php">http://www.tcd.uni-konstanz.de/research/plants.php</a>	Plants: small molecule docking (rigid receptor although some side chains can move, flexible ligand)	[326]
<a href="ftp://ftp2.ipc.pku.edu.cn/pub/software">ftp://ftp2.ipc.pku.edu.cn/pub/software</a>	PSI-DOCK: small molecule docking (rigid receptor although some side chains can move, flexible ligand)	[327]
<a href="http://ang.cz3.nus.edu.sg/cgi-bin/prog/rune.pl">http://ang.cz3.nus.edu.sg/cgi-bin/prog/rune.pl</a>	PEARLS: program for energetic analysis of receptor-ligand system	[328]
<a href="http://www.scfbio-iitd.res.in/dock/pardock.jsp">http://www.scfbio-iitd.res.in/dock/pardock.jsp</a>	ParDOCK is an all-atom energy based Monte Carlo, rigid protein ligand docking tool	[327]
<a href="http://abcis.cbs.cnrs.fr/kindock/">http://abcis.cbs.cnrs.fr/kindock/</a> <a href="http://abcis.cbs.cnrs.fr/LIGBASE_SERV_WEB/PHP/simdock.php">http://abcis.cbs.cnrs.fr/LIGBASE_SERV_WEB/PHP/simdock.php</a>	Kindock and Simdock: tool for comparative docking of protein kinase ligands and ligand transposition server	[156]
<a href="http://indy2.biow.uni-leipzig.de/psoatautodock.html">http://indy2.biow.uni-leipzig.de/psoatautodock.html</a>	Pso@autodock: a fast flexible molecular docking program based on Swarm Intelligence	[329]
<a href="http://www.bio.ifi.lmu.de/~apostola/index.html">http://www.bio.ifi.lmu.de/~apostola/index.html</a>	GlamDock: a new docking tool that could become available	[330]

Fig. 1: Schematic overview of the *in silico-in vitro* drug discovery process

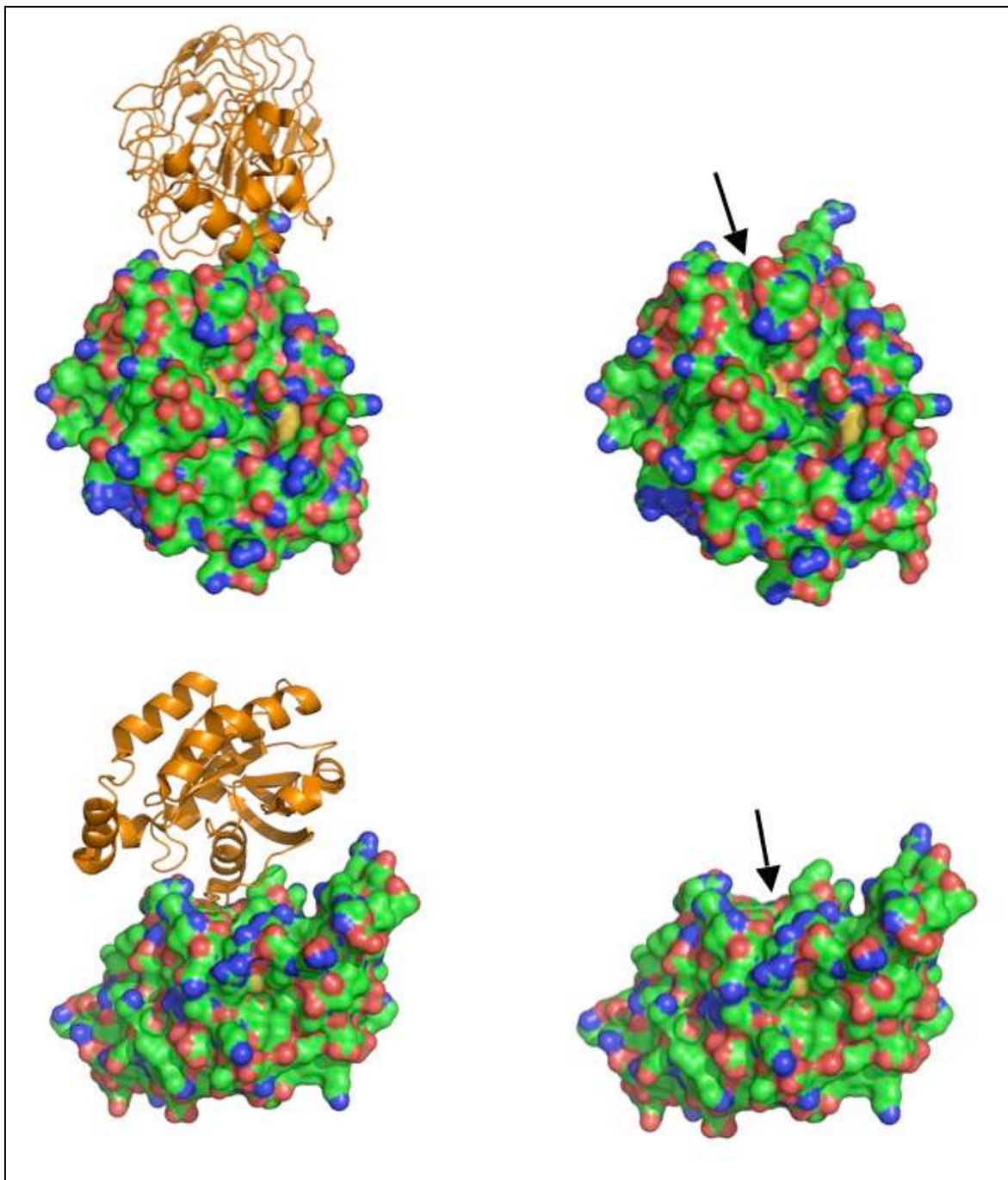


**Fig. 2: Druggable pocket prediction:** The catalytic site of a serine protease is shown with a view down the active site. A crystallized small molecule inhibitor is displayed as CPK sphere (yellow). The predicted druggable pocket performed with Q-Site finder is in navy blue. Several other proposed druggable pockets are highlighted with black arrows but not shown to facilitate the reading of the figure. It is known from site directed mutagenesis and crystallographic studies that serine proteases interact with other proteins in these areas, confirming the overall efficiencies of druggable pocket prediction methods.

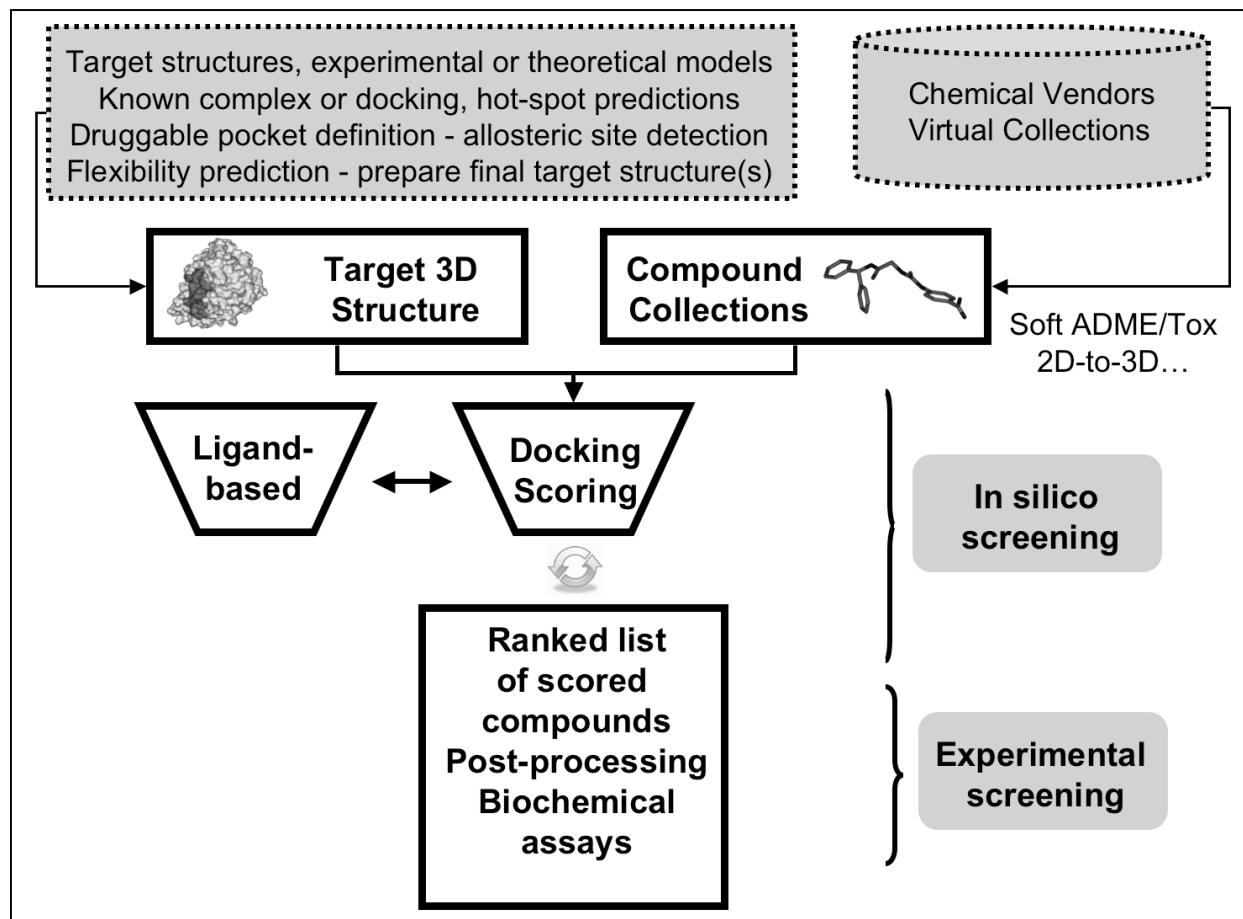




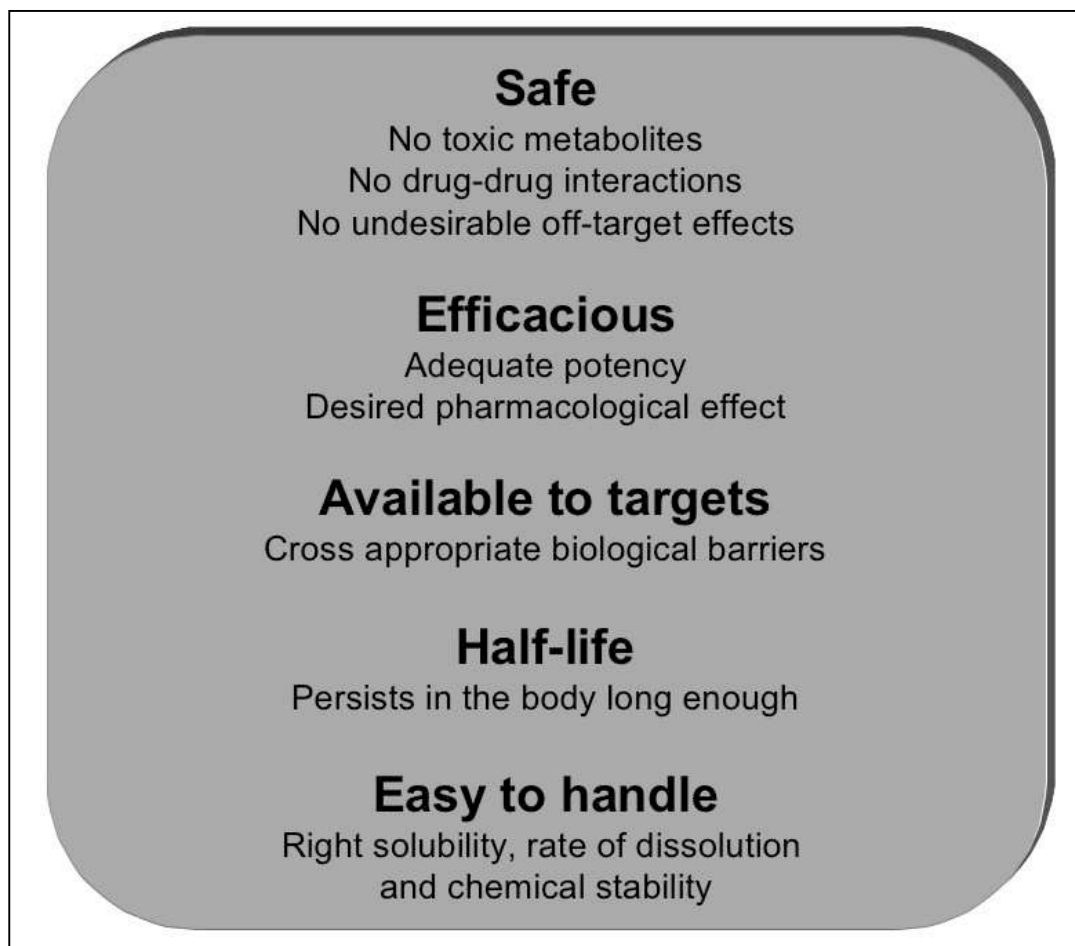
**Fig. 3: Examples of protein complexes with a clear druggable pocket at the interface.** Top left: Thrombin (solid surface with carbon atoms in green, predicted positively charged groups in blue and negatively charged red) is crystallized with platelet glycoprotein Ib-alpha (cartoon representation, orange) (PDB code 1ook) [331]. Top right, thrombin is removed and the arrow highlights a druggable pocket that is more visible in 3D than in a 2D drawing. Bottom left. The Rac1-GTP (cartoon)/p67TPR (solid surface) complex (PDB code 1e96) [332]. Bottom right, one partner is removed and a clear druggable pocket can be seen at the interface.



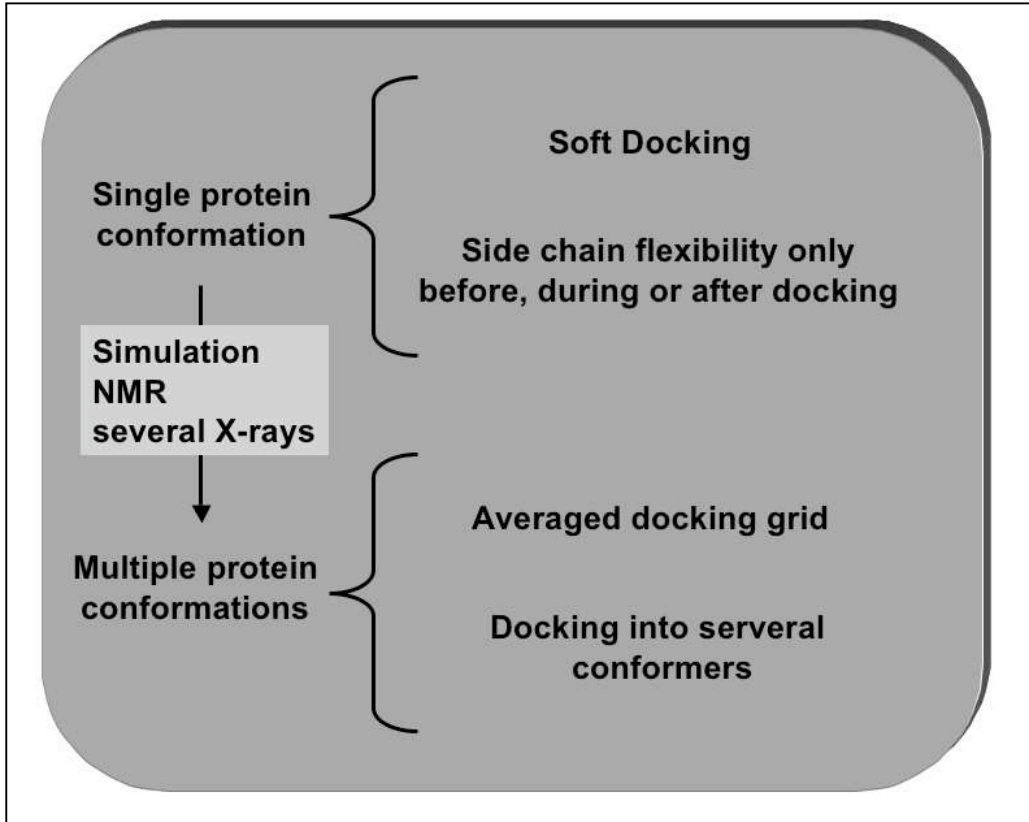
**Fig. 4: Example of *in silico-in vitro* screening flowchart for PPIs**



**Fig. 5: ADME/Tox guidelines:** the ultimate profile that we would like to have for the selected compounds. Many of these parameters cannot be predicted accurately *in silico* with currently available methods and our knowledge on many of the steps is still very limited. The term pharmacokinetics (PK) is often used when dealing with ADME properties, this is because PK can be defined as the study of the time course of a drug within the body and because it incorporates the process of absorption, distribution, metabolism and excretion. For the sake of simplification, pharmacodynamics (PD) is often summarized as the study of what a drug does to the body, whereas pharmacokinetics is the study of what the body does to a drug.

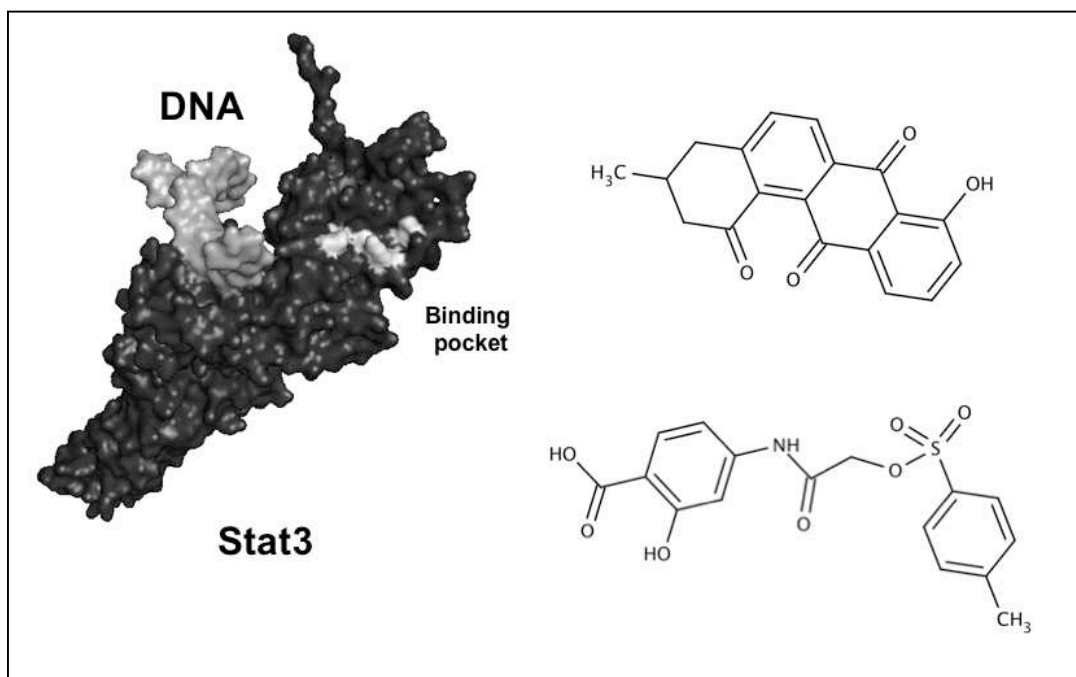


**Fig. 6: Some possible ways for handling receptor flexibility**



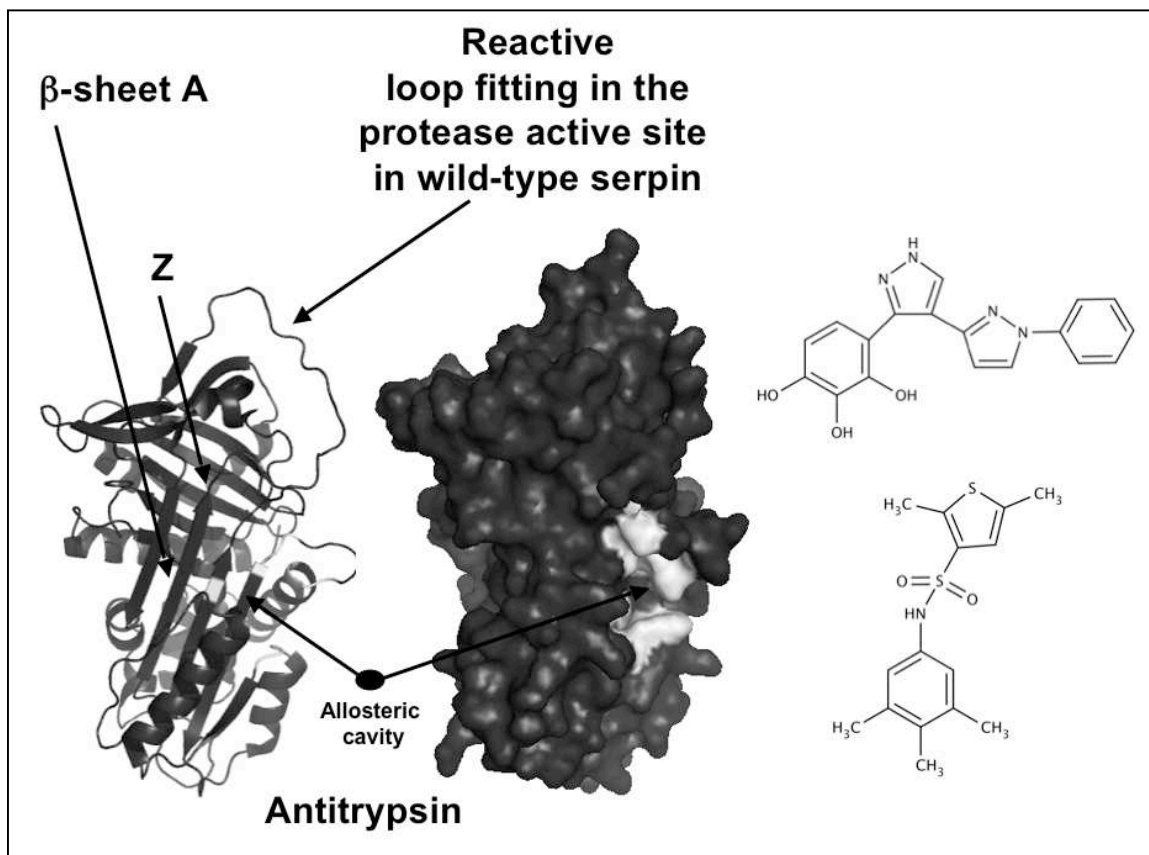
**Fig. 7: Stat3-DNA complex**

The Stat3 monomer structure is shown as a solid surface with the DNA molecule shown in light grey and the binding pocket for association with the other monomer and for the interaction with drug-like compounds identified after *in silico-in vitro* screening shown in white. The top molecule is from Song et al. [277] and the bottom compound is from Siddiquee et al. [275].



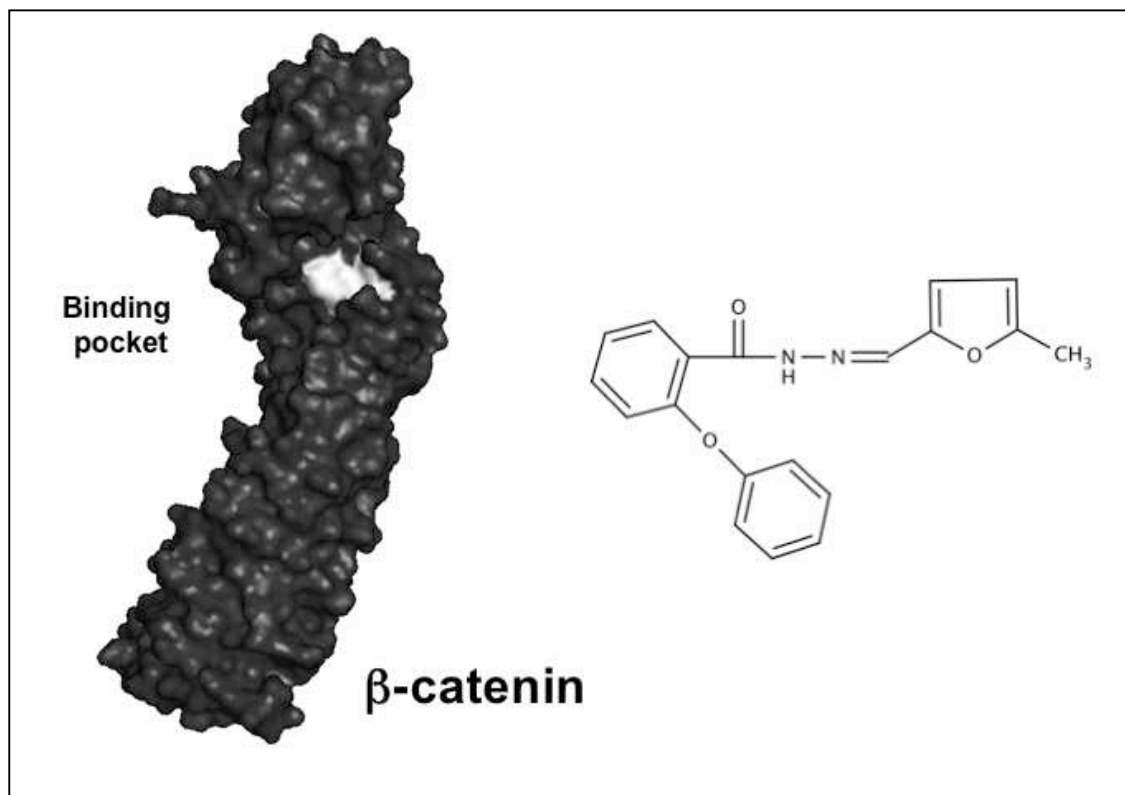
### Fig. 8: Antitrypsin

The Z mutation is highlighted and so are the so-called  $\beta$ -sheet A and reactive center loop. The allosteric binding pocket is colored white and the best compound identified in that study is the bottom compound that block polymerization.



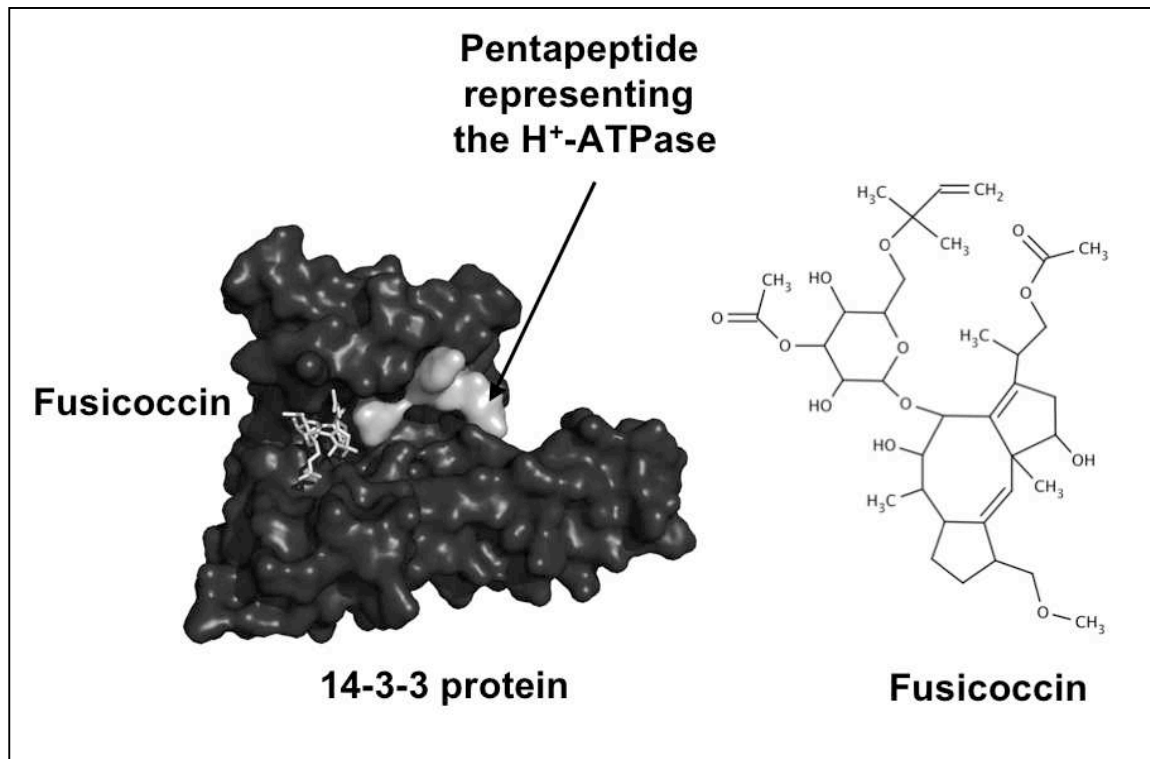
**Fig. 9:  $\beta$ -catenin**

The X-ray structure of  $\beta$ -catenin is shown as a solid surface with the pocket selected for virtual screening highlighted in white. One compound identified by in silico screening followed by biophysical screening is shown. It is not yet fully confirmed if this ligand binds to this pocket.



**Fig. 10: Plant H<sup>+</sup>-ATPase-14-3-3-fusicoccin complex**

SB-VLS was performed on the 14-3-3 protein in order to replace fusicoccin, a small fungal toxin known to stabilize the H<sup>+</sup>-ATPase-14-3-3 protein complex. Only few residues from the ATPase are present in this crystal structure while 4 water molecules (not visible in this figure) are located near the ligand. The toxin is co-crystallized in a small pocket and interacts with at least one residue from the ATPase. The toxin is also drawn to the right to be consistent with the other figures.





## References

- [1] Yin, H. and Hamilton, A. D. (2005) *Angew Chem Int Ed Engl* **44**, 4130-63.
- [2] Lipinski, C. and Hopkins, A. (2004) *Nature* **432**, 855-61.
- [3] Zheng, C.; Han, L.; Yap, C. W.; Xie, B. and Chen, Y. (2006) *Drug Discov Today* **11**, 412-20.
- [4] Harris, C. J. and Stevens, A. P. (2006) *Drug Discov Today* **11**, 880-8.
- [5] Hall, S. E. (2006) *Drug Discov Today* **11**, 495-502.
- [6] Stockwell, B. R. (2004) *Nature* **432**, 846-54.
- [7] Friedrich, U.; Blom, A. M.; Dahlback, B. and Villoutreix, B. O. (2001) *J Biol Chem* **276**, 24122-8.
- [8] Nicolaes, G. A.; Sorensen, K. W.; Friedrich, U.; Tans, G.; Rosing, J.; Autin, L.; Dahlback, B. and Villoutreix, B. O. (2004) *Eur J Biochem* **271**, 2724-36.
- [9] Shen, L.; Villoutreix, B. O. and Dahlback, B. (1999) *Thromb Haemost* **82**, 72-9.
- [10] Mohammadi, M.; Olsen, S. K. and Goetz, R. (2005) *Curr Opin Struct Biol* **15**, 506-16.
- [11] Olson, S. T.; Bjork, I.; Sheffer, R.; Craig, P. A.; Shore, J. D. and Choay, J. (1992) *J Biol Chem* **267**, 12528-38.
- [12] Baines, I. C. and Colas, P. (2006) *Drug Discov Today* **11**, 334-41.
- [13] Ireson, C. R. and Kelland, L. R. (2006) *Mol Cancer Ther* **5**, 2957-62.
- [14] Karamouzis, M. V.; Konstantinopoulos, P. A. and Papavassiliou, A. G. (2007) *J Mol Med* **85**, 427-36.
- [15] Hershberger, S. J.; Lee, S. G. and Chmielewski, J. (2007) *Curr Top Med Chem* **7**, 928-42.
- [16] Dennis, M. S.; Eigenbrot, C.; Skelton, N. J.; Ultsch, M. H.; Santell, L.; Dwyer, M. A.; O'Connell, M. P. and Lazarus, R. A. (2000) *Nature* **404**, 465-70.
- [17] Safholm, A.; Leandersson, K.; Dejmeck, J.; Nielsen, C. K.; Villoutreix, B. O. and Andersson, T. (2006) *J Biol Chem* **281**, 2740-9.
- [18] Saenko, E. L.; Ananyeva, N. M.; Kouliavskaia, D. V.; Khrenov, A. V.; Anderson, J. A.; Shima, M.; Qian, J. and Scott, D. (2002) *Haemophilia* **8**, 1-11.
- [19] Kerschbaumer, R. J.; Riedrich, K.; Kral, M.; Varadi, K.; Dorner, F.; Rosing, J. and Scheifflinger, F. (2004) *J Biol Chem* **279**, 40445-50.
- [20] Clackson, T. and Wells, J. A. (1995) *Science* **267**, 383-6.
- [21] Jacquemin, M.; Lavend'homme, R.; Benhida, A.; Vanzielegem, B.; d'Oiron, R.; Lavergne, J. M.; Brackmann, H. H.; Schwaab, R.; VandenDriessche, T.; Chuah, M. K.; Hoylaerts, M.; Gilles, J. G.; Peerlinck, K.; Vermynen, J. and Saint-Remy, J. M. (2000) *Blood* **96**, 958-65.
- [22] Curat, C. A.; Eck, M.; Dervillez, X. and Vogel, W. F. (2001) *J Biol Chem* **276**, 45952-8.
- [23] Archakov, A. I.; Govorun, V. M.; Dubanov, A. V.; Ivanov, Y. D.; Veselovsky, A. V.; Lewi, P. and Janssen, P. (2003) *Proteomics* **3**, 380-91.
- [24] Nguyen, J. T. and Wells, J. A. (2003) *Proc Natl Acad Sci U S A* **100**, 7533-8.
- [25] Davis, A. M.; Teague, S. J. and Kleywegt, G. J. (2003) *Angew Chem Int Ed Engl* **42**, 2718-36.

- [26] Jahnke, W. (2007) *J Biomol NMR* **39**, 87-90.
- [27] Babaoglu, K. and Shoichet, B. K. (2006) *Nat Chem Biol* **2**, 720-3.
- [28] Anderson, A. C. (2003) *Chem Biol* **10**, 787-97.
- [29] Alvesalo, J. K. O.; Siiskonen, A.; Vainio, M. J.; Tammela, P. S. M. and Vuorela, P. M. (2006) *J Med Chem* **49**, 2353-2356.
- [30] Takeuchi, K. and Wagner, G. (2006) *Curr Opin Struct Biol* **16**, 109-17.
- [31] Davies, J. W.; Glick, M. and Jenkins, J. L. (2006) *Curr Opin Chem Biol* **10**, 343-51.
- [32] Stevens, R. C. (2003) *Drug Discovery World* **4**, 35-48.
- [33] Cheng, A.; Diller, D. J.; Dixon, S. L.; Egan, W. J.; Lauri, G. and Merz, K. M., Jr. (2002) *J Comput Chem* **23**, 172-83.
- [34] Michelson, S.; Sehgal, A. and Friedrich, C. (2006) *Curr Opin Biotechnol* **17**, 666-70.
- [35] Michelson, S. and Joho, K. (2000) *Curr Opin Mol Ther* **2**, 651-4.
- [36] Maiorov, V. and Sheridan, R. P. (2005) *J Chem Inf Model* **45**, 1017-23.
- [37] Miteva, M. A.; Lee, W. H.; Montes, M. O. and Villoutreix, B. O. (2005) *J Med Chem* **48**, 6012-22.
- [38] Whitty, A. and Kumaravel, G. (2006) *Nat Chem Biol* **2**, 112-8.
- [39] Collins, I. and Workman, P. (2006) *Nat Chem Biol* **2**, 689-700.
- [40] Arkin, M. (2005) *Curr Opin Chem Biol* **9**, 317-24.
- [41] Drews, J. (2000) *Science* **287**, 1960-4.
- [42] Jones, S.; Marin, A. and Thornton, J. M. (2000) *Protein Eng* **13**, 77-82.
- [43] Nooren, I. M. and Thornton, J. M. (2003) *J Mol Biol* **325**, 991-1018.
- [44] Keskin, O.; Ma, B. and Nussinov, R. (2005) *J Mol Biol* **345**, 1281-94.
- [45] Keskin, O.; GURSOY, A.; Ma, B. and Nussinov, R. (2007) *Curr Top Med Chem* **7**, 943-51.
- [46] Lee, B. and Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379-380.
- [47] Richards, F. M. (1977) *Annu Rev Biophys Bioeng* **6**, 151-76.
- [48] Cazals, F.; Proust, F.; Bahadur, R. P. and Janin, J. (2006) *Protein Sci* **15**, 2082-92.
- [49] Poupon, A. (2004) *Curr Opin Struct Biol* **14**, 233-41.
- [50] Mitchell, J. C.; Kerr, R. and Ten Eyck, L. F. (2001) *J Mol Graph Model* **19**, 325-30, 388-90.
- [51] Lawrence, M. C. and Colman, P. M. (1993) *J Mol Biol* **234**, 946-50.
- [52] Sinha, N. and Smith-Gill, S. J. (2002) *Curr Protein Pept Sci* **3**, 601-14.
- [53] Jones, S. and Thornton, J. M. (1995) *Prog Biophys Mol Biol* **63**, 31-65.
- [54] Bogan, A. A. and Thorn, K. S. (1998) *J Mol Biol* **280**, 1-9.
- [55] Reichmann, D.; Rahat, O.; Albeck, S.; Meged, R.; Dym, O. and Schreiber, G. (2005) *Proc Natl Acad Sci U S A* **102**, 57-62.
- [56] Ma, B.; Elkayam, T.; Wolfson, H. and Nussinov, R. (2003) *Proc Natl Acad Sci U S A* **100**, 5772-7.
- [57] Gonzalez-Ruiz, D. and Gohlke, H. (2006) *Curr Med Chem* **13**, 2607-25.
- [58] Reichmann, D.; Rahat, O.; Cohen, M.; Neuvirth, H. and Schreiber, G. (2007) *Curr Opin Struct Biol* **17**, 67-76.
- [59] Moreira, I. S.; Fernandes, P. A. and Ramos, M. J. (2007) *Proteins* **68**, 803-12.
- [60] Nye, T. M.; Berzuini, C.; Gilks, W. R.; Babu, M. M. and Teichmann, S. A. (2005) *Bioinformatics* **21**, 993-1001.

- [61] Nye, T. M.; Lio, P. and Gilks, W. R. (2006) *Bioinformatics* **22**, 117-9.
- [62] Lo Conte, L.; Chothia, C. and Janin, J. (1999) *J Mol Biol* **285**, 2177-98.
- [63] Covell, D. G. and Wallqvist, A. (1997) *J Mol Biol* **269**, 281-97.
- [64] Rodier, F.; Bahadur, R. P.; Chakrabarti, P. and Janin, J. (2005) *Proteins* **60**, 36-45.
- [65] Xu, D.; Lin, S. L. and Nussinov, R. (1997) *J Mol Biol* **265**, 68-84.
- [66] Xu, D.; Tsai, C. J. and Nussinov, R. (1997) *Protein Eng* **10**, 999-1012.
- [67] Villoutreix, B. O. (2002) *Curr Protein Pept Sci* **3**, 341-64.
- [68] Kundrotas, P. J. and Alexov, E. (2006) *Biophys J* **91**, 1724-36.
- [69] Chakrabarti, P. and Janin, J. (2002) *Proteins* **47**, 334-43.
- [70] Halperin, I.; Wolfson, H. and Nussinov, R. (2004) *Structure* **12**, 1027-38.
- [71] Caffrey, D. R.; Somaroo, S.; Hughes, J. D.; Mintseris, J. and Huang, E. S. (2004) *Protein Sci* **13**, 190-202.
- [72] Camacho, C. J. (2005) *Proteins* **60**, 245-51.
- [73] Kimura, S. R.; Brower, R. C.; Vajda, S. and Camacho, C. J. (2001) *Biophys J* **80**, 635-42.
- [74] Rajamani, D.; Thiel, S.; Vajda, S. and Camacho, C. J. (2004) *Proc Natl Acad Sci U S A* **101**, 11287-92.
- [75] Preissner, R.; Goede, A. and Frommel, C. (1998) *J Mol Biol* **280**, 535-50.
- [76] Fischer, T. B.; Arunachalam, K. V.; Bailey, D.; Mangual, V.; Bakhru, S.; Russo, R.; Huang, D.; Paczkowski, M.; Lalchandani, V.; Ramachandra, C.; Ellison, B.; Galer, S.; Shapley, J.; Fuentes, E. and Tsai, J. (2003) *Bioinformatics* **19**, 1453-4.
- [77] Bader, G. D.; Betel, D. and Hogue, C. W. (2003) *Nucleic Acids Res* **31**, 248-50.
- [78] Henrick, K. and Thornton, J. M. (1998) *Trends Biochem Sci* **23**, 358-61.
- [79] Stein, A.; Russell, R. B. and Aloy, P. (2005) *Nucleic Acids Res* **33**, D413-7.
- [80] Davis, F. P. and Sali, A. (2005) *Bioinformatics* **21**, 1901-7.
- [81] Finn, R. D.; Marshall, M. and Bateman, A. (2005) *Bioinformatics* **21**, 410-2.
- [82] Ng, S. K.; Zhang, Z.; Tan, S. H. and Lin, K. (2003) *Nucleic Acids Res* **31**, 251-4.
- [83] Gong, S.; Park, C.; Choi, H.; Ko, J.; Jang, I.; Lee, J.; Bolser, D. M.; Oh, D.; Kim, D. S. and Bhak, J. (2005) *BMC Bioinformatics* **6**, 207.
- [84] Levy, E. D.; Pereira-Leal, J. B.; Chothia, C. and Teichmann, S. A. (2006) *PLoS Comput Biol* **2**, e155.
- [85] Kundrotas, P. J. and Alexov, E. (2007) *Nucleic Acids Res* **35**, D575-9.
- [86] Gao, Y.; Douguet, D.; Tovchigrechko, A. and Vakser, I. A. (2007) *Proteins* **69**, 845-851.
- [87] Winter, C.; Henschel, A.; Kim, W. K. and Schroeder, M. (2006) *Nucleic Acids Res* **34**, D310-4.
- [88] Mintz, S.; Shulman-Peleg, A.; Wolfson, H. J. and Nussinov, R. (2005) *Proteins* **61**, 6-20.
- [89] Riley, R.; Lee, C.; Sabatti, C. and Eisenberg, D. (2005) *Genome Biol* **6**, R89.
- [90] Gilson, M. K.; Given, J. A.; Bush, B. L. and McCammon, J. A. (1997) *Biophys J* **72**, 1047-69.
- [91] Noskov, S. Y. and Lim, C. (2001) *Biophys J* **81**, 737-50.
- [92] Camacho, C. J.; Ma, H. and Champ, P. C. (2006) *Proteins* **63**, 868-77.
- [93] Champ, P. C. and Camacho, C. J. (2007) *Nucleic Acids Res* **35**, W556-60.

- [94] Villoutreix, B. O.; Renault, N.; Lagorce, D.; Sperandio, O.; Montes, M. and Miteva, M. A. (2007) *Curr Protein Pept Sci* **8**, 381-411.
- [95] Jones, S. and Thornton, J. M. (2004) *Curr Opin Chem Biol* **8**, 3-7.
- [96] Fernandez-Recio, J.; Totrov, M.; Skorodumov, C. and Abagyan, R. (2005) *Proteins* **58**, 134-43.
- [97] Kufareva, I.; Budagyan, L.; Raush, E.; Totrov, M. and Abagyan, R. (2007) *Proteins* **67**, 400-17.
- [98] Kortemme, T. and Baker, D. (2002) *Proc Natl Acad Sci U S A* **99**, 14116-21.
- [99] Bordner, A. J. and Abagyan, R. (2005) *Proteins* **60**, 353-66.
- [100] Tina, K. G.; Bhadra, R. and Srinivasan, N. (2007) *Nucleic Acids Res* **35**, W473-6.
- [101] Shazman, S.; Celniker, G.; Haber, O.; Glaser, F. and Mandel-Gutfreund, Y. (2007) *Nucleic Acids Res* **35**, W526-30.
- [102] Liu, S.; Zhang, C.; Zhou, H. and Zhou, Y. (2004) *Proteins* **56**, 93-101.
- [103] Neuvirth, H.; Heinemann, U.; Birnbaum, D.; Tishby, N. and Schreiber, G. (2007) *Nucleic Acids Res* **35**, W543-8.
- [104] Chen, Y. C.; Lo, Y. S.; Hsu, W. C. and Yang, J. M. (2007) *Nucleic Acids Res* **35**, W561-7.
- [105] Innis, C. A. (2007) *Nucleic Acids Res* **35**, W489-94.
- [106] Burgoyne, N. J. and Jackson, R. M. (2006) *Bioinformatics* **22**, 1335-42.
- [107] Grosdidier, S.; Pons, C.; Solernou, A. and Fernandez-Recio, J. (2007) *Proteins* **69**, 852-8.
- [108] Bonvin, A. M. (2006) *Curr Opin Struct Biol* **16**, 194-200.
- [109] Gray, J. J. (2006) *Curr Opin Struct Biol* **16**, 183-93.
- [110] Fernandez-Recio, J.; Totrov, M. and Abagyan, R. (2004) *J Mol Biol* **335**, 843-65.
- [111] Fernandez-Recio, J.; Totrov, M. and Abagyan, R. (2003) *Proteins* **52**, 113-7.
- [112] Fernandez-Recio, J.; Totrov, M. and Abagyan, R. (2002) *Pac Symp Biocomput*, 552-63.
- [113] Fernandez-Recio, J.; Totrov, M. and Abagyan, R. (2002) *Protein Sci* **11**, 280-91.
- [114] Chelliah, V.; Blundell, T. L. and Fernandez-Recio, J. (2006) *J Mol Biol* **357**, 1669-82.
- [115] Camacho, C. J. and Zhang, C. (2005) *Bioinformatics* **21**, 2534-6.
- [116] Bastard, K.; Prevost, C. and Zacharias, M. (2006) *Proteins* **62**, 956-69.
- [117] Dominguez, C.; Boelens, R. and Bonvin, A. M. (2003) *J Am Chem Soc* **125**, 1731-7.
- [118] Palma, P. N.; Krippahl, L.; Wampler, J. E. and Moura, J. J. (2000) *Proteins* **39**, 372-84.
- [119] Jiang, L.; Kuhlman, B.; Kortemme, T. and Baker, D. (2005) *Proteins* **58**, 893-904.
- [120] Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R. and Wolfson, H. J. (2005) *Nucleic Acids Res* **33**, W363-7.
- [121] Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R. and Wolfson, H. J. (2005) *Proteins* **60**, 224-31.
- [122] Mustard, D. and Ritchie, D. W. (2005) *Proteins* **60**, 269-74.
- [123] Dominguez, C.; Bonvin, A. M.; Winkler, G. S.; van Schaik, F. M.; Timmers, H. T. and Boelens, R. (2004) *Structure* **12**, 633-44.
- [124] Comeau, S. R.; Gatchell, D. W.; Vajda, S. and Camacho, C. J. (2004) *Bioinformatics* **20**, 45-50.

- [125] Autin, L.; Miteva, M. A.; Lee, W. H.; Mertens, K.; Radtke, K. P. and Villoutreix, B. O. (2005) *J Thromb Haemost* **3**, 2044-56.
- [126] Autin, L.; Steen, M.; Dahlback, B. and Villoutreix, B. O. (2006) *Proteins* **63**, 440-50.
- [127] Toby, G. G. and Golemis, E. A. (2001) *Methods* **24**, 201-17.
- [128] Pflieger, K. D.; Seeber, R. M. and Eidne, K. A. (2006) *Nat Protoc* **1**, 337-45.
- [129] Karlsson, R. and Falt, A. (1997) *J Immunol Methods* **200**, 121-33.
- [130] Pierce, M. M.; Raman, C. S. and Nall, B. T. (1999) *Methods* **19**, 213-21.
- [131] Lipfert, J. and Doniach, S. (2007) *Annu Rev Biophys Biomol Struct* **36**, 307-27.
- [132] Russell, R. B.; Alber, F.; Aloy, P.; Davis, F. P.; Korkin, D.; Pichaud, M.; Topf, M. and Sali, A. (2004) *Curr Opin Struct Biol* **14**, 313-24.
- [133] Vasilescu, J. and Figeys, D. (2006) *Curr Opin Biotechnol* **17**, 394-9.
- [134] Hultschig, C.; Kreutzberger, J.; Seitz, H.; Konthur, Z.; Bussow, K. and Lehrach, H. (2006) *Curr Opin Chem Biol* **10**, 4-10.
- [135] Zhong, S.; Macias, A. T. and MacKerell, A. D., Jr. (2007) *Curr Top Med Chem* **7**, 63-82.
- [136] Hardy, J. A. and Wells, J. A. (2004) *Curr Opin Struct Biol* **14**, 706-15.
- [137] Freire, E. (2000) *Proc Natl Acad Sci U S A* **97**, 11680-2.
- [138] Hajduk, P. J.; Huth, J. R. and Tse, C. (2005) *Drug Discov Today* **10**, 1675-82.
- [139] Hajduk, P. J.; Huth, J. R. and Fesik, S. W. (2005) *J Med Chem* **48**, 2518-25.
- [140] Schulz-Gasch, T. and Stahl, M. (2003) *J Mol Model (Online)* **9**, 47-57.
- [141] Coleman, R. G.; Salzberg, A. C. and Cheng, A. C. (2006) *J Chem Inf Model* **46**, 2631-7.
- [142] Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C. and Huang, E. S. (2007) *Nat Biotechnol* **25**, 71-5.
- [143] Halgren, T. (2007) *Chem Biol Drug Des* **69**, 146-8.
- [144] An, J.; Totrov, M. and Abagyan, R. (2004) *Genome Inform Ser Workshop Genome Inform* **15**, 31-41.
- [145] An, J.; Totrov, M. and Abagyan, R. (2005) *Mol Cell Proteomics* **4**, 752-61.
- [146] Niwa, T. (2006) *J Chem Inf Model* **46**, 2158-66.
- [147] Vajda, S. and Guarnieri, F. (2006) *Curr Opin Drug Discov Devel* **9**, 354-62.
- [148] Laurie, A. T. and Jackson, R. M. (2006) *Curr Protein Pept Sci* **7**, 395-406.
- [149] Noble, W. S. (2006) *Nat Biotechnol* **24**, 1565-7.
- [150] Han, L. Y.; Zheng, C. J.; Xie, B.; Jia, J.; Ma, X. H.; Zhu, F.; Lin, H. H.; Chen, X. and Chen, Y. Z. (2007) *Drug Discov Today* **12**, 304-13.
- [151] Zhang, X. and Bajaj, C. (2007) *Comput Syst Bioinformatics Conf* **6**, 275-86.
- [152] Li, B.; Turuvekere, S.; Agrawal, M.; La, D.; Ramani, K. and Kihara, D. (2007) *Proteins*.
- [153] Gold, N. D. and Jackson, R. M. (2006) *J Chem Inf Model* **46**, 736-42.
- [154] Kuhn, D.; Weskamp, N.; Schmitt, S.; Hullermeier, E. and Klebe, G. (2006) *J Mol Biol* **359**, 1023-44.
- [155] Dessailly, B. H.; Lensink, M. F.; Orenco, C. A. and Wodak, S. J. (2007) *Nucleic Acids Res*.
- [156] Martin, L.; Catherinot, V. and Labesse, G. (2006) *Nucleic Acids Res* **34**, W325-9.
- [157] Kellenberger, E.; Muller, P.; Schalón, C.; Bret, G.; Foata, N. and Rognan, D. (2006) *J Chem Inf Model* **46**, 717-27.

- [158] Kellenberger, E.; Rodrigo, J.; Muller, P. and Rognan, D. (2004) *Proteins* **57**, 225-42.
- [159] Nayal, M. and Honig, B. (2006) *Proteins*.
- [160] Block, P.; Weskamp, N.; Wolf, A. and Klebe, G. (2007) *Proteins* **68**, 170-86.
- [161] Sperandio, O.; Miteva, M. A.; Delfaud, F. and Villoutreix, B. O. (2006) *Curr Protein Pept Sci*, 369-393.
- [162] Jain, A. N. (2004) *Curr Opin Drug Discov Devel* **7**, 396-403.
- [163] Bajorath, J. (2001) *Drug Discov Today* **6**, 989-995.
- [164] Leach, A. R.; Shoichet, B. K. and Peishoff, C. E. (2006) *J Med Chem* **49**, 5851-5.
- [165] Klebe, G. (2006) *Drug Discov Today* **11**, 580-94.
- [166] Jalaie, M. and Shanmugasundaram, V. (2006) *Mini Rev Med Chem* **6**, 1159-67.
- [167] McInnes, C. (2007) *Curr Opin Chem Biol* **11**, 494-502.
- [168] Krovat, E. M.; Steindl, T. and Langer, T. (2005) *Current Computer-Aided Drug Design* **1**, 93-102.
- [169] Shoichet, B. K. (2004) *Nature* **432**, 862-5.
- [170] Abagyan, R. and Totrov, M. (2001) *Curr Opin Chem Biol* **5**, 375-82.
- [171] Kitchen, D. B.; Decornez, H.; Furr, J. R. and Bajorath, J. (2004) *Nat Rev Drug Discov* **3**, 935-49.
- [172] Prathipati, P.; Dixit, A. and Saxena, A. K. (2007) *Curr Computer-Aided Drug Design* **3**, 341-352.
- [173] Stahl, M.; Guba, W. and Kansy, M. (2006) *Drug Discov Today* **11**, 326-33.
- [174] Oprea, T. I. and Matter, H. (2004) *Curr Opin Chem Biol* **8**, 349-58.
- [175] Green, D. V. (2003) *Prog Med Chem* **41**, 61-97.
- [176] Brown, N. and Jacoby, E. (2006) *Mini Rev Med Chem* **6**, 1217-29.
- [177] Zhao, H. (2007) *Drug Discov Today* **12**, 149-55.
- [178] Hristozov, D. P.; Oprea, T. I. and Gasteiger, J. (2007) *J Comput Aided Mol Des.*
- [179] Nayeem, A.; Sitkoff, D. and Krystek, S., Jr. (2006) *Protein Sci* **15**, 808-24.
- [180] Ferrara, P. and Jacoby, E. (2007) *J Mol Model* **13**, 897-905.
- [181] Rockey, W. M. and Elcock, A. H. (2006) *Curr Protein Pept Sci* **7**, 437-57.
- [182] Jenwitheesuk, E.; Horst, J. A.; Rivas, K. L.; Van Voorhis, W. C. and Samudrala, R. (2008) *Trends Pharmacol Sci*.
- [183] Geromichalos, G. D. (2007) *J Buon* **12 Suppl 1**, S101-18.
- [184] Newman, D. J. and Cragg, G. M. (2007) *J Nat Prod* **70**, 461-77.
- [185] Hergenrother, P. J. (2006) *Curr Opin Chem Biol* **10**, 213-8.
- [186] Krier, M.; Bret, G. and Rognan, D. (2006) *J Chem Inf Model* **46**, 512-24.
- [187] Monge, A.; Arrault, A.; Marot, C. and Morin-Allory, L. (2006) *Mol Divers* **10**, 389-403.
- [188] Sirois, S.; Hatzakis, G.; Wei, D.; Du, Q. and Chou, K. C. (2005) *Comput Biol Chem* **29**, 55-67.
- [189] Baurin, N.; Baker, R.; Richardson, C.; Chen, I.; Foloppe, N.; Potter, A.; Jordan, A.; Roughley, S.; Parratt, M.; Greaney, P.; Morley, D. and Hubbard, R. E. (2004) *J Chem Inf Comput Sci* **44**, 643-651.
- [190] Saxena, A. K. and Prathipati, P. (2006) *SAR QSAR Environ Res* **17**, 371-92.
- [191] Lipinski, C. A. (2000) *J Pharmacol Toxicol Methods* **44**, 235-49.
- [192] Muegge, I. (2003) *Med Res Rev* **23**, 302-21.

- [193] Oprea, T. I.; Davis, A. M.; Teague, S. J. and Leeson, P. D. (2001) *J Chem Inf Comput Sci* **41**, 1308-15.
- [194] Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E. M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G. and Schneider, G. (2002) *J Med Chem* **45**, 137-42.
- [195] Rishton, G. M. (2003) *Drug Discov Today* **8**, 86-96.
- [196] Vedani, A.; Dobler, M. and Lill, M. A. (2006) *Basic Clin Pharmacol Toxicol* **99**, 187-94.
- [197] Blake, J. F. (2005) *Med Chem* **1**, 649-55.
- [198] Beresford, A. P.; Segall, M. and Tarbit, M. H. (2004) *Curr Opin Drug Discov Devel* **7**, 36-42.
- [199] Good, A. C. and Hermsmeier, M. A. (2007) *J Chem Inf Model* **47**, 110-4.
- [200] Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; Taylor, R. D. and Watson, P. (2004) *J Chem Inf Comput Sci* **44**, 793-806.
- [201] Knox, A. J.; Meegan, M. J.; Carta, G. and Lloyd, D. G. (2005) *J Chem Inf Model* **45**, 1908-19.
- [202] Orry, A. J.; Abagyan, R. A. and Cavasotto, C. N. (2006) *Drug Discov Today* **11**, 261-6.
- [203] Olah, M. M.; Bologna, C. G. and Oprea, T. I. (2004) *Curr Drug Discov Technol* **1**, 211-20.
- [204] Congreve, M.; Carr, R.; Murray, C. and Jhoti, H. (2003) *Drug Discov Today* **8**, 876-7.
- [205] Rees, D. C.; Congreve, M.; Murray, C. W. and Carr, R. (2004) *Nat Rev Drug Discov* **3**, 660-72.
- [206] Miteva, M. A.; Violas, S.; Montes, M.; Gomez, D.; Tuffery, P. and Villoutreix, B. O. (2006) *Nucleic Acids Research* **34**, W738-44.
- [207] Wheeler, D. L.; Barrett, T.; Benson, D. A.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; DiCuccio, M.; Edgar, R.; Federhen, S.; Geer, L. Y.; Helmberg, W.; Kapustin, Y.; Kenton, D. L.; Khovayko, O.; Lipman, D. J.; Madden, T. L.; Maglott, D. R.; Ostell, J.; Pruitt, K. D.; Schuler, G. D.; Schriml, L. M.; Sequeira, E.; Sherry, S. T.; Sirotkin, K.; Souvorov, A.; Starchenko, G.; Suzek, T. O.; Tatusov, R.; Tatusova, T. A.; Wagner, L. and Yaschenko, E. (2006) *Nucleic Acids Res* **34**, D173-80.
- [208] Irwin, J. J. and Shoichet, B. K. (2005) *J Chem Inf Model* **45**, 177-82.
- [209] Girke, T.; Cheng, L. C. and Raikhel, N. (2005) *Plant Physiol* **138**, 573-577.
- [210] Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z. and Woolsey, J. (2006) *Nucleic Acids Res* **34**, D668-72.
- [211] Polgar, T.; Magyar, C.; Simon, I. and Keseru, G. M. (2007) *J Chem Inf Model*.
- [212] Milletti, F.; Storchi, L.; Sforza, G. and Cruciani, G. (2007) *J Chem Inf Model*.
- [213] Lu, Y.; Nikolovska-Coleska, Z.; Fang, X.; Gao, W.; Shangary, S.; Qiu, S.; Qin, D. and Wang, S. (2006) *J Med Chem* **49**, 3759-62.
- [214] Neugebauer, A.; Hartmann, R. W. and Klein, C. D. (2007) *J Med Chem* **50**, 4665-8.
- [215] Alvarez, J. C. (2004) *Curr Opin Chem Biol* **8**, 365-70.
- [216] Cavasotto, C. N. and Orry, A. J. (2007) *Curr Top Med Chem* **7**, 1006-14.

- [217] Fletcher, S. and Hamilton, A. D. (2007) *Curr Top Med Chem* **7**, 922-7.
- [218] Segers, K.; Sperandio, O.; Sack, M.; Fischer, R.; Miteva, M. A.; Rosing, J.; Nicolaes, G. A. and Villoutreix, B. O. (2007) *Proc Natl Acad Sci U S A* **104**, 12697-702.
- [219] Huth, J. R.; Yu, L.; Collins, I.; Mack, J.; Mendoza, R.; Isaac, B.; Braddock, D. T.; Muchmore, S. W.; Comess, K. M.; Fesik, S. W.; Clore, G. M.; Levens, D. and Hajduk, P. J. (2004) *J Med Chem* **47**, 4851-7.
- [220] Foloppe, N.; Matassova, N. and Aboul-Ela, F. (2006) *Drug Discov Today* **11**, 1019-27.
- [221] Raghuraman, A.; Mosier, P. D. and Desai, U. R. (2006) *J Med Chem* **49**, 3553-62.
- [222] Rochet, J. C. (2007) *Expert Rev Mol Med* **9**, 1-34.
- [223] Brooijmans, N. and Kuntz, I. D. (2003) *Annu Rev Biophys Biomol Struct* **32**, 335-73.
- [224] Jain, A. N. (2006) *Curr Protein Pept Sci* **7**, 407-420.
- [225] Gilson, M. K. and Zhou, H. X. (2007) *Annu Rev Biophys Biomol Struct* **36**, 21-42.
- [226] Huang, N. and Jacobson, M. P. (2007) *Curr Opin Drug Discov Devel* **10**, 325-31.
- [227] Rajamani, R. and Good, A. C. (2007) *Curr Opin Drug Discov Devel* **10**, 308-15.
- [228] Feher, M. (2006) *Drug Discov Today* **11**, 421-8.
- [229] Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M. and Merz, K. M., Jr. (2007) *Drug Discov Today* **12**, 725-31.
- [230] Springer, C.; Adalsteinsson, H.; Young, M. M.; Kegelmeyer, P. W. and Roe, D. C. (2005) *J Med Chem* **48**, 6821-31.
- [231] Wang, J.; Kang, X.; Kuntz, I. D. and Kollman, P. A. (2005) *J Med Chem* **48**, 2432-44.
- [232] Floriano, W. B.; Vaidehi, N.; Zamanakos, G. and Goddard, W. A., 3rd (2004) *J Med Chem* **47**, 56-71.
- [233] Huang, N.; Kalyanaraman, C.; Irwin, J. J. and Jacobson, M. P. (2006) *J Chem Inf Model* **46**, 243-53.
- [234] Deng, Z.; Chuaqui, C. and Singh, J. (2004) *J Med Chem* **47**, 337-44.
- [235] Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F. and Mason, J. S. (2007) *J Chem Inf Model* **47**, 279-94.
- [236] Wei, B. Q.; Baase, W. A.; Weaver, L. H.; Matthews, B. W. and Shoichet, B. K. (2002) *J Mol Biol* **322**, 339-55.
- [237] Triballeau, N.; Acher, F.; Brabet, I.; Pin, J. P. and Bertrand, H. O. (2005) *J Med Chem* **48**, 2534-47.
- [238] Truchon, J. F. and Bayly, C. I. (2007) *J Chem Inf Model* **47**, 488-508.
- [239] Kroemer, R. T.; Vulpetti, A.; McDonald, J. J.; Rohrer, D. C.; Trosset, J. Y.; Giordanetto, F.; Cotesta, S.; McMartin, C.; Kihlen, M. and Stouten, P. F. (2004) *J Chem Inf Comput Sci* **44**, 871-81.
- [240] Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J. and Corbeil, C. R. (2007) *Br J Pharmacol*.
- [241] Teague, S. J. (2003) *Nat Rev Drug Discov* **2**, 527-41.
- [242] Mintseris, J.; Wiehe, K.; Pierce, B.; Anderson, R.; Chen, R.; Janin, J. and Weng, Z. (2005) *Proteins* **60**, 214-6.
- [243] Marco, E. and Gago, F. (2007) *ChemMedChem* **2**, 1388-1401.
- [244] Carlson, H. A. (2002) *Curr Opin Chem Biol* **6**, 447-52.
- [245] Alonso, H.; Bliznyuk, A. A. and Gready, J. E. (2006) *Med Res Rev* **26**, 531-68.



- [246] Huang, Z.; Wong, C. F. and Wheeler, R. A. (2007) *Proteins*.
- [247] Zhao, Y. and Sanner, M. F. (2007) *Proteins* **68**, 726-737.
- [248] Cavasotto, C. N. and Abagyan, R. A. (2004) *J Mol Biol* **337**, 209-25.
- [249] Barril, X. and Morley, S. D. (2005) *J Med Chem* **48**, 4432-43.
- [250] Nabuurs, S. B.; Wagener, M. and Vlieg, J. (in press) *J Med Chem*.
- [251] Corbeil, C. R.; Englebienne, P. and Moitessier, N. (2007) *J Chem Inf Model* **47**, 435-49.
- [252] Eyrisch, S. and Helms, V. (2007) *J Med Chem*.
- [253] Luque, I. and Freire, E. (2000) *Proteins Suppl* **4**, 63-71.
- [254] Sugaya, N.; Ikeda, K.; Tashiro, T.; Takeda, S.; Otomo, J.; Ishida, Y.; Shiratori, A.; Toyoda, A.; Noguchi, H.; Takeda, T.; Kuhara, S.; Sakaki, Y. and Iwayanagi, T. (2007) *BMC Pharmacol* **7**, 10.
- [255] Ruffner, H.; Bauer, A. and Bouwmeester, T. (2007) *Drug Discov Today* **12**, 709-16.
- [256] Peri, S.; Navarro, J. D.; Amanchy, R.; Kristiansen, T. Z.; Jonnalagadda, C. K.; Surendranath, V.; Niranjan, V.; Muthusamy, B.; Gandhi, T. K.; Gronborg, M.; Ibarrola, N.; Deshpande, N.; Shanker, K.; Shivashankar, H. N.; Rashmi, B. P.; Ramya, M. A.; Zhao, Z.; Chandrika, K. N.; Padma, N.; Harsha, H. C.; Yatish, A. J.; Kavitha, M. P.; Menezes, M.; Choudhury, D. R.; Suresh, S.; Ghosh, N.; Saravana, R.; Chandran, S.; Krishna, S.; Joy, M.; Anand, S. K.; Madavan, V.; Joseph, A.; Wong, G. W.; Schiemann, W. P.; Constantinescu, S. N.; Huang, L.; Khosravi-Far, R.; Steen, H.; Tewari, M.; Ghaffari, S.; Blobbe, G. C.; Dang, C. V.; Garcia, J. G.; Pevsner, J.; Jensen, O. N.; Roepstorff, P.; Deshpande, K. S.; Chinnaiyan, A. M.; Hamosh, A.; Chakravarti, A. and Pandey, A. (2003) *Genome Res* **13**, 2363-71.
- [257] Bork, P.; Jensen, L. J.; von Mering, C.; Ramani, A. K.; Lee, I. and Marcotte, E. M. (2004) *Curr Opin Struct Biol* **14**, 292-9.
- [258] Congreve, M.; Murray, C. W. and Blundell, T. L. (2005) *Drug Discov Today* **10**, 895-907.
- [259] Marsden, B. D.; Sundstrom, M. and Knapp, S. (2006) *Expert Opin. Drug Discov.* **1**, 123-136.
- [260] Morelli, X. and Rigby, A. C. (2007) *Curr Computer-Aided Drug Design* **3**, 33-49.
- [261] Stauffer, S. R. (2007) *Curr Top Med Chem* **7**, 961-5.
- [262] Dudkina, A. S. and Lindsley, C. W. (2007) *Curr Top Med Chem* **7**, 952-60.
- [263] Kern, J. T.; Hannink, M. and Hess, J. F. (2007) *Curr Top Med Chem* **7**, 972-8.
- [264] Schimmer, A. D.; Dalili, S.; Batey, R. A. and Riedl, S. J. (2006) *Cell Death Differ* **13**, 179-88.
- [265] Rajapakse, H. A. (2007) *Curr Top Med Chem* **7**, 966-71.
- [266] Arkin, M. R. and Wells, J. A. (2004) *Nat Rev Drug Discov* **3**, 301-17.
- [267] Thanos, C. D.; DeLano, W. L. and Wells, J. A. (2006) *Proc Natl Acad Sci U S A* **103**, 15422-7.
- [268] Moro, S.; Spalluto, G. and Jacobson, K. A. (2005) *Trends Pharmacol Sci* **26**, 44-51.
- [269] Klabunde, T. and Hessler, G. (2002) *Chembiochem* **3**, 928-44.
- [270] Becker, O. M.; Marantz, Y.; Shacham, S.; Inbal, B.; Heifetz, A.; Kalid, O.; Bar-Haim, S.; Warshaviak, D.; Fichman, M. and Noiman, S. (2004) *Proc Natl Acad Sci U S A* **101**, 11304-9.

- [271] Becker, O. M.; Shacham, S.; Marantz, Y. and Noiman, S. (2003) *Curr Opin Drug Discov Devel* **6**, 353-61.
- [272] Pommier, Y. and Cherfils, J. (2005) *Trends Pharmacol Sci* **26**, 138-45.
- [273] Yu, H. and Jove, R. (2004) *Nat Rev Cancer* **4**, 97-105.
- [274] Becker, S.; Groner, B. and Muller, C. W. (1998) *Nature* **394**, 145-51.
- [275] Siddiquee, K.; Zhang, S.; Guida, W. C.; Blaskovich, M. A.; Greedy, B.; Lawrence, H. R.; Yip, M. L.; Jove, R.; McLaughlin, M. M.; Lawrence, N. J.; Sebti, S. M. and Turkson, J. (2007) *Proc Natl Acad Sci U S A* **104**, 7391-6.
- [276] Shao, H.; Xu, X.; Mastrangelo, M. A.; Jing, N.; Cook, R. G.; Legge, G. B. and Tweardy, D. J. (2004) *J Biol Chem* **279**, 18967-73.
- [277] Song, H.; Wang, R.; Wang, S. and Lin, J. (2005) *Proc Natl Acad Sci U S A* **102**, 4700-5.
- [278] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P. and Shenkin, P. S. (2004) *J Med Chem* **47**, 1739-49.
- [279] Wang, R.; Lu, Y.; Fang, X. and Wang, S. (2004) *J Chem Inf Comput Sci* **44**, 2114-25.
- [280] Wang, R.; Lai, L. and Wang, S. (2002) *J Comput Aided Mol Des* **16**, 11-26.
- [281] Kuntz, I. D. (1992) *Science* **257**, 1078-82.
- [282] Picard, V.; Dautzenberg, M. D.; Villoutreix, B. O.; Orliaguet, G.; Alhenc-Gelas, M. and Aiach, M. (2003) *Blood* **102**, 919-25.
- [283] Lomas, D. A. and Mahadeva, R. (2002) *J Clin Invest* **110**, 1585-90.
- [284] Mallya, M.; Phillips, R. L.; Saldanha, S. A.; Goptu, B.; Brown, S. C.; Termine, D. J.; Shirvani, A. M.; Wu, Y.; Sifers, R. N.; Abagyan, R. and Lomas, D. A. (2007) *J Med Chem* **50**, 5357-63.
- [285] Elliott, P. R.; Pei, X. Y.; Dafforn, T. R. and Lomas, D. A. (2000) *Protein Sci* **9**, 1274-81.
- [286] Trosset, J. Y.; Dalvit, C.; Knapp, S.; Fasolini, M.; Veronesi, M.; Mantegani, S.; Gianellini, L. M.; Catana, C.; Sundstrom, M.; Stouten, P. F. and Moll, J. K. (2006) *Proteins* **64**, 60-7.
- [287] Graham, T. A.; Weaver, C.; Mao, F.; Kimelman, D. and Xu, W. (2000) *Cell* **103**, 885-96.
- [288] Wurtele, M.; Jelich-Ottmann, C.; Wittinghofer, A. and Oecking, C. (2003) *Embo J* **22**, 987-94.
- [289] Merlini, G. and Bellotti, V. (2003) *N Engl J Med* **349**, 583-96.
- [290] Hammarstrom, P.; Wiseman, R. L.; Powers, E. T. and Kelly, J. W. (2003) *Science* **299**, 713-6.
- [291] Klabunde, T.; Petrassi, H. M.; Oza, V. B.; Raman, P.; Kelly, J. W. and Sacchettini, J. C. (2000) *Nat Struct Biol* **7**, 312-21.
- [292] Adamski-Werner, S. L.; Palaninathan, S. K.; Sacchettini, J. C. and Kelly, J. W. (2004) *J Med Chem* **47**, 355-74.
- [293] Maia, F.; Almeida Mdo, R.; Gales, L.; Kijjoa, A.; Pinto, M. M.; Saraiva, M. J. and Damas, A. M. (2005) *Biochem Pharmacol* **70**, 1861-9.
- [294] Almeida, M. R.; Gales, L.; Damas, A. M.; Cardoso, I. and Saraiva, M. J. (2005) *Curr Drug Targets CNS Neurol Disord* **4**, 587-96.

- [295] Ray, S. S.; Nowak, R. J.; Brown, R. H., Jr. and Lansbury, P. T., Jr. (2005) *Proc Natl Acad Sci U S A* **102**, 3639-44.
- [296] Betzi, S.; Restouin, A.; Opi, S.; Arold, S. T.; Parrot, I.; Guerlesquin, F.; Morelli, X. and Collette, Y. (2007) *Proc Natl Acad Sci U S A*.
- [297] Brylinski, M. and Skolnick, J. (2007) *Proteins*.
- [298] Najmanovich, R.; Kuttner, J.; Sobolev, V. and Edelman, M. (2000) *Proteins* **39**, 261-8.
- [299] Bradford, J. R. and Westhead, D. R. (2005) *Bioinformatics* **21**, 1487-94.
- [300] Murakami, Y. and Jones, S. (2006) *Bioinformatics* **22**, 1794-5.
- [301] Tovchigrechko, A. and Vakser, I. A. (2006) *Nucleic Acids Res* **34**, W310-4.
- [302] Shatsky, M.; Dror, O.; Schneidman-Duhovny, D.; Nussinov, R. and Wolfson, H. J. (2004) *Nucleic Acids Res* **32**, W503-7.
- [303] Comeau, S. R.; Gatchell, D. W.; Vajda, S. and Camacho, C. J. (2004) *Nucleic Acids Res* **32**, W96-9.
- [304] de Vries, S. J.; van Dijk, A. D.; Krzeminski, M.; van Dijk, M.; Thureau, A.; Hsu, V.; Wassenaar, T. and Bonvin, A. M. (2007) *Proteins* **69**, 726-733.
- [305] Kozakov, D.; Brenke, R.; Comeau, S. R. and Vajda, S. (2006) *Proteins* **65**, 392-406.
- [306] Wang, H.; Segal, E.; Ben-Hur, A.; Li, Q. R.; Vidal, M. and Koller, D. (2007) *Genome Biol* **8**, R192.
- [307] Laurie, A. T. and Jackson, R. M. (2005) *Bioinformatics* **21**, 1908-16.
- [308] Chang, D. T.; Oyang, Y. J. and Lin, J. H. (2005) *Nucleic Acids Res* **33**, W233-8.
- [309] Brady, G. P. and Stouten, P. F. W. (2000) *J Computer-Aided Molecular Design* **14**, 383-401.
- [310] Lopez, G.; Valencia, A. and Tress, M. L. (2007) *Nucleic Acids Res* **35**, W573-7.
- [311] Weisel, M.; Proschak, E. and Schneider, G. (2007) *Chem Cent J* **1**, 7.
- [312] Kalidas, Y. and Chandra, N. (2007) *J Struct Biol*.
- [313] Brylinski, M.; Konieczny, L. and Roterman, I. (2006) *Bioinformation* **1**, 127-9.
- [314] Huang, B. and Schroeder, M. (2006) *BMC Struct Biol* **6**, 19.
- [315] Klekota, J.; Roth, F. P. and Schreiber, S. L. (2006) *Bioinformatics* **22**, 1670-3.
- [316] Bergner, A.; Gunther, J.; Hendlich, M.; Klebe, G. and Verdonk, M. (2001) *Biopolymers* **61**, 99-110.
- [317] Schuttelkopf, A. W. and van Aalten, D. M. (2004) *Acta Crystallogr D Biol Crystallogr* **60**, 1355-63.
- [318] Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O. and Tuffery, P. (2007) *Nucleic Acids Res* **35**, W568-72.
- [319] Wang, R.; Lu, Y. and Wang, S. (2003) *J Med Chem* **46**, 2287-303.
- [320] Morris, G. M.; Goodsell, D. S.; Huey, R. and Olson, A. J. (1996) *J Comput Aided Mol Des* **10**, 293-304.
- [321] Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S. and Johnson, A. P. (2006) *Curr Protein Pept Sci* **7**, 421-35.
- [322] McGann, M. R.; Almond, H. R.; Nicholls, A.; Grant, J. A. and Brown, F. K. (2003) *Biopolymers* **68**, 76-90.
- [323] Li, H.; Gao, Z.; Kang, L.; Zhang, H.; Yang, K.; Yu, K.; Luo, X.; Zhu, W.; Chen, K.; Shen, J.; Wang, X. and Jiang, H. (2006) *Nucleic Acids Res* **34**, W219-24.
- [324] Jain, A. N. (2007) *J Comput Aided Mol Des* **21**, 281-306.

- [325] Jain, A. N. (2003) *J Med Chem* **46**, 499-511.
- [326] Korb, O.; Stutzle, T. and Exner, T. E. (2006) *Ant Colony Optimization and Swarm Intelligence, 5th International Workshop*, 247-258.
- [327] Pei, J.; Wang, Q.; Liu, Z.; Li, Q.; Yang, K. and Lai, L. (2006) *Proteins* **62**, 934-46.
- [328] Han, L. Y.; Lin, H. H.; Li, Z. R.; Zheng, C. J.; Cao, Z. W.; Xie, B. and Chen, Y. Z. (2006) *J Chem Inf Model* **46**, 445-450.
- [329] Namasivayam, V. and Gunther, R. (2007) *Chem Biol Drug Des.*
- [330] Tietze, S. and Apostolakis, J. (2007) *J Chem Inf Model* **47**, 1657-72.
- [331] Celikel, R.; McClintock, R. A.; Roberts, J. R.; Mendolicchio, G. L.; Ware, J.; Varughese, K. I. and Ruggeri, Z. M. (2003) *Science* **301**, 218-21.
- [332] Lapouge, K.; Smith, S. J.; Walker, P. A.; Gamblin, S. J.; Smerdon, S. J. and Rittinger, K. (2000) *Mol Cell* **6**, 899-907.