Molecules **2015**, *20*, 13384-13421; doi:10.3390/molecules200713384

OPEN ACCESS

molecules

ISSN 1420-3049

www.mdpi.com/journal/molecules

Review

Molecular Docking and Structure-Based Drug Design Strategies

Leonardo G. Ferreira *, Ricardo N. dos Santos, Glaucius Oliva and Adriano D. Andricopulo *

Laboratório de Química Medicinal e Computacional, Centro de Pesquisa e Inovação em Biodiversidade e Fármacos, Instituto de Física de São Carlos, Universidade de São Paulo, Av. João Dagnone 1100, São Carlos-SP 13563-120, Brazil; E-Mails: rnsantos@ursa.ifsc.usp.br (R.N.S.); oliva@ifsc.usp.br (G.O.)

* Authors to whom correspondence should be addressed; E-Mails: leonardo@ifsc.usp.br (L.G.F.); aandrico@ifsc.usp.br (A.D.A.); Tel.: +55-163-373-8095 (A.D.A.); Fax: +55-163-373-9881 (A.D.A.).

Academic Editor: Rino Ragno

Received: 13 May 2015 / Accepted: 20 July 2015 / Published: 22 July 2015

Abstract: Pharmaceutical research has successfully incorporated a wealth of molecular modeling methods, within a variety of drug discovery programs, to study complex biological and chemical systems. The integration of computational and experimental strategies has been of great value in the identification and development of novel promising compounds. Broadly used in modern drug design, molecular docking methods explore the ligand conformations adopted within the binding sites of macromolecular targets. This approach also estimates the ligand-receptor binding free energy by evaluating critical phenomena involved in the intermolecular recognition process. Today, as a variety of docking algorithms are available, an understanding of the advantages and limitations of each method is of fundamental importance in the development of effective strategies and the generation of relevant results. The purpose of this review is to examine current molecular docking strategies used in drug discovery and medicinal chemistry, exploring the advances in the field and the role played by the integration of structure- and ligand-based methods.

Keywords: molecular modeling; drug discovery; molecular target; molecular interaction; pharmacophore; virtual screening; SBDD; SBVS

---

1. Introduction

The research-based pharmaceutical industry has increasingly employed modern medicinal chemistry methods, including molecular modeling, as powerful tools for the study of structure-activity relationships (SAR) [1]. In addition to pharmacodynamics data (e.g., potency, affinity, efficacy, selectivity), pharmacokinetic properties (ADMET: absorption, distribution, metabolism, excretion and toxicity) have also been studied through the application of these methodologies [2]. The field has progressed hand-in-hand with advances in biomolecular spectroscopic methods such as X-ray crystallography and nuclear magnetic resonance (NMR), which have enabled striking progress in molecular and structural biology. These techniques have allowed the resolution of more than 100,000 three-dimensional protein structures, providing vital structural information about key macromolecular drug targets [3]. Efforts in storing, organizing and exploring such information have generated a growing demand for robust and sophisticated computational tools. Based on this perspective, the accurate integration of *in silico* and experimental methods has provided the up-to-date understanding of the intricate aspects of intermolecular recognition [4].

Within this framework, structure-based drug design (SBDD) methods (*i.e.*, the use of three-dimensional structural information gathered from biological targets) are a prominent component of modern medicinal chemistry [5]. Molecular docking, structure-based virtual screening (SBVS) and molecular dynamics (MD) are among the most frequently used SBDD strategies due to their wide range of applications in the analysis of molecular recognition events such as binding energetics, molecular interactions and induced conformational changes [6]. A distinct approach in drug design comprises the use of bioactive small-molecule libraries. The unique chemical diversity available in these libraries represents the space occupied by ligands known to interact with a specific target. This type of information is used in ligand-based drug design (LBDD) methods [7]. Ligand-based virtual screening (LBVS), similarity searching, QSAR modeling and pharmacophore generation are some of the most useful LBDD methods [8].

SBDD and LBDD approaches have been applied as valuable drug discovery tools both in academia and industry [9], owing to their versatility and synergistic character. The integration of these approaches has been successfully employed in a number of investigations of structural, chemical and biological data [10,11].

2. Structure-Based Drug Design (SBDD)

Understanding the principles by which small-molecule ligands recognize and interact with macromolecules is of great importance in pharmaceutical research and development (R & D) [12]. SBDD refers to the systematic use of structural data (e.g., macromolecular targets, also called receptors), which are usually obtained experimentally or through computational homology modeling [13]. The purpose is to conceive ligands with specific electrostatic and stereochemical attributes to achieve high receptor binding affinity. The availability of three-dimensional macromolecular structures enables a diligent inspection of the binding site topology, including the presence of clefts, cavities and sub-pockets. Electrostatic properties, such as charge distribution, can also be carefully examined. Current SBDD methods allow for the design of ligands containing the necessary features for efficient modulation of the target receptor [12,13]. Selective modulation of a validated drug target by high affinity ligands

interferes with specific cellular processes, ultimately leading to the desired pharmacological and therapeutic effects [14].

SBDD is a cyclic process consisting of stepwise knowledge acquisition (Figure 1). Starting from a known target structure, *in silico* studies are conducted to identify potential ligands. These molecular modeling procedures are followed by the synthesis of the most promising compounds [15]. Next, evaluations of biological properties, such as potency, affinity and efficacy, are carried out using diverse experimental platforms [16]. Provided that active compounds are identified, the three-dimensional structure of the ligand-receptor complex can be solved. The available structure allows the observation of several intermolecular features supporting the process of molecular recognition. Structural descriptions of ligand-receptor complexes are useful for the investigation of binding conformations, characterization of key intermolecular interactions, characterization of unknown binding sites, mechanistic studies and the elucidation of ligand-induced conformational changes [17].

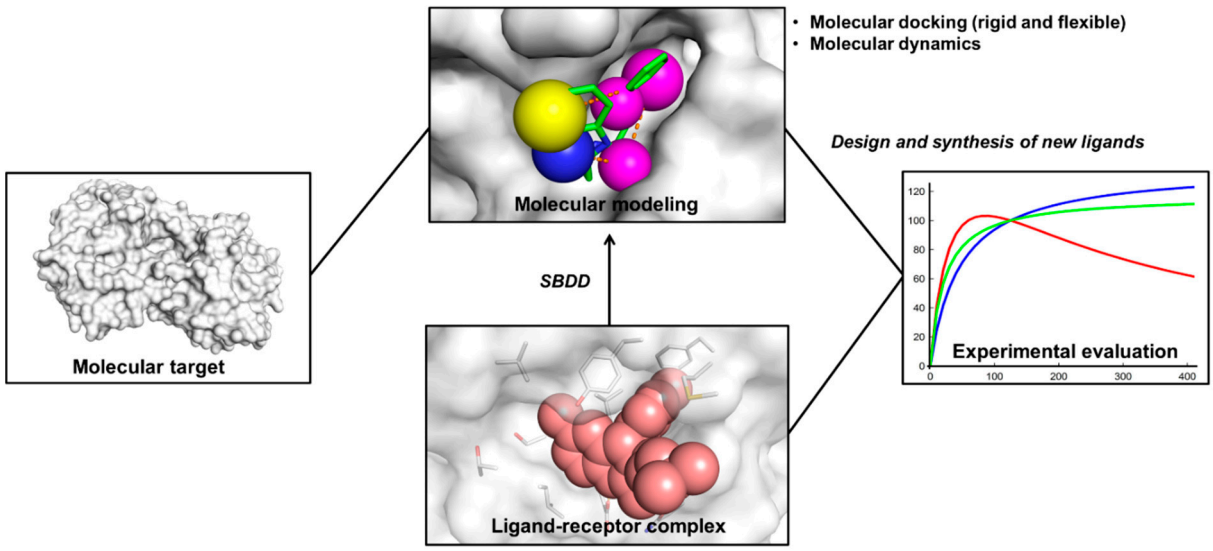

Figure 1. Outline of SBDD. The three-dimensional structure of the molecular target is employed in molecular modeling studies. Promising compounds are synthesized and then experimentally evaluated. Given that bioactive small-molecules are discovered, the structure of a ligand-receptor complex can be obtained. The binding complex is used in molecular modeling studies and novel compounds are designed.

Once a ligand-receptor complex has been determined, biological activity data are correlated to the structural information [18]. In this way, the SBDD process starts over with new steps to incorporate molecular modifications with the potential to increase the affinity of new ligands for the binding site. The flexibility of the target receptor is an essential aspect that must be considered throughout the modeling phase, bearing in mind that substantial conformational change can occur upon ligand binding. The use of techniques such as flexible docking and MD are useful in addressing the flexibility issue [19,20].

3. Molecular Docking

Molecular docking is one of the most frequently used methods in SBDD because of its ability to predict, with a substantial degree of accuracy, the conformation of small-molecule ligands within the

appropriate target binding site (Figure 2) [21]. Following the development of the first algorithms in the 1980s, molecular docking became an essential tool in drug discovery [22]. For example, investigations involving crucial molecular events, including ligand binding modes and the corresponding intermolecular interactions that stabilize the ligand-receptor complex, can be conveniently performed [23]. Furthermore, molecular docking algorithms execute quantitative predictions of binding energetics, providing rankings of docked compounds based on the binding affinity of ligand-receptor complexes [22,23].

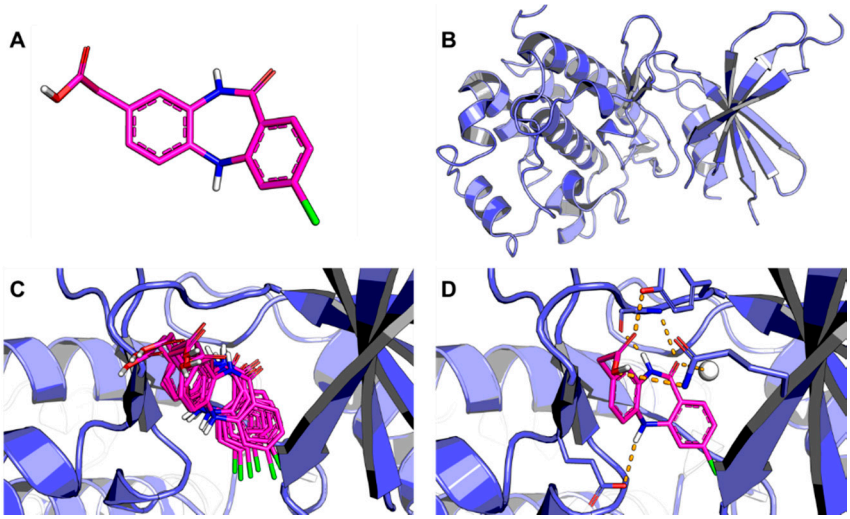

Figure 2. Outline of the molecular docking process. (**A**) Three-dimensional structure of the ligand; (**B**) Three-dimensional structure of the receptor; (**C**) The ligand is docked into the binding cavity of the receptor and the putative conformations are explored; (**D**) The most likely binding conformation and the corresponding intermolecular interactions are identified. The protein backbone is represented as a cartoon. The ligand (carbon in magenta) and active site residues (carbon in blue) are shown in stick representation. Water is shown as a white sphere and hydrogen bonds are indicated as dashed lines.

The identification of the most likely binding conformations requires two steps: (i) exploration of a large conformational space representing various potential binding modes; (ii) accurate prediction of the interaction energy associated with each of the predicted binding conformations [24]. Molecular docking programs perform these tasks through a cyclical process, in which the ligand conformation is evaluated by specific scoring functions. This process is carried out recursively until converging to a solution of minimum energy [23–25].

3.1. Conformational Search

In the conformational search stage, structural parameters of the ligands, such as torsional (dihedral), translational and rotational degrees of freedom, are incrementally modified (Figure 3A). Conformational search algorithms perform this task by applying systematic and stochastic search methods [25,26].

Systematic search methods promote slight variations in the structural parameters, gradually changing the conformation of the ligands [27]. The algorithm probes the energy landscape of the conformational space and, after numerous search and evaluation cycles, converges to the minimum energy solution corresponding to the most likely binding mode (Figure 3B). Although the method is effective in

exploring the conformational space, it can converge to a local minimum rather than the global minimum. This drawback can be overcome by performing simultaneous searches starting from different points of the energy landscape (*i.e.*, distinct conformations) [28].

Stochastic methods carry out the conformational search by randomly modifying the structural parameters of the ligands [29]. For this, the algorithm generates ensembles of molecular conformations and populates a wide range of the energy landscape (Figure 3C). This strategy avoids trapping the final solution at a local energy minimum and increases the probability of finding a global minimum. As the algorithm promotes a broad coverage of the energy landscape, the computational cost associated with this procedure is an important limitation [28,29].

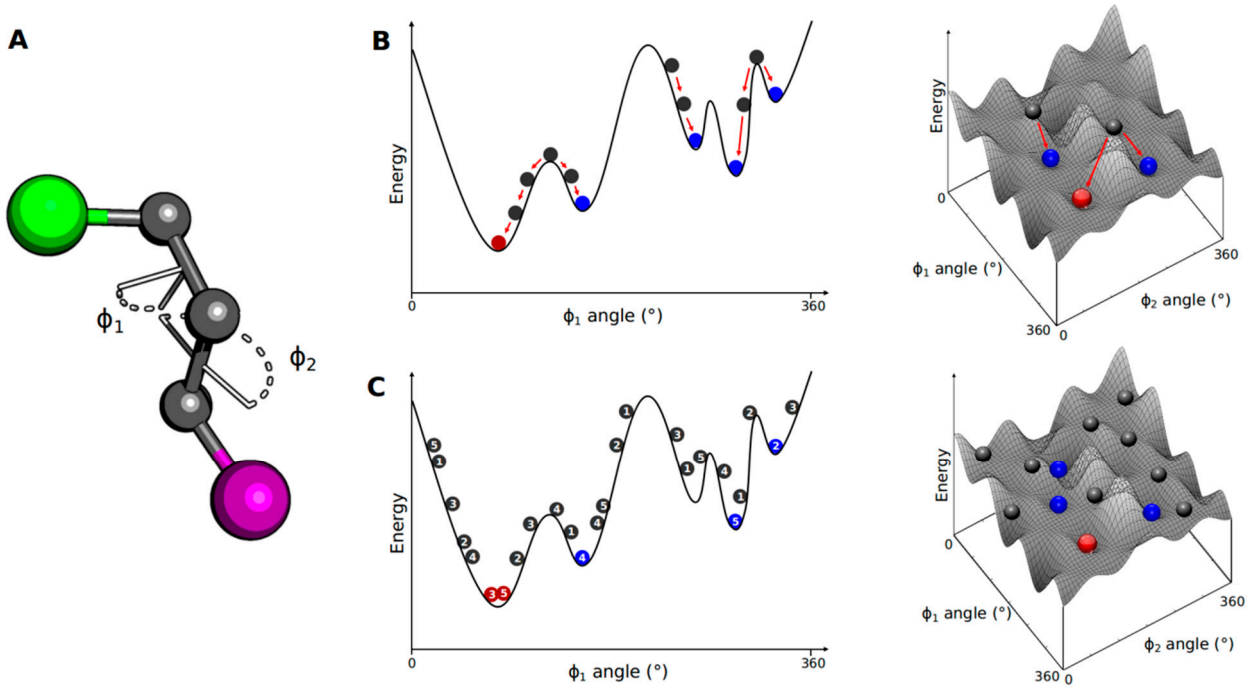

Figure 3. Small-molecule conformational search methods. (**A**) A molecule containing two bulky groups (green and purple spheres) has its conformation defined by two internal dihedrals Φ_1 and Φ_2; (**B**) Considering Φ_2 as a frozen dihedral, the energy variation due to rotation of Φ_1 is plotted in a 1D energy landscape. The initial structure (grey spheres) is modified by changing Φ_1, leading to a decrease in energy. The systematic search algorithm changes all structural parameters until a local (blue spheres) or global (red sphere) energy minimum is reached; (**C**) The stochastic search explores the conformational space by randomly generating distinct conformations, populating a broad range of the energy landscape. This procedure increases the probability of finding a global energy minimum.

Systematic and stochastic methods are included in widely used molecular docking programs, which have specific approaches to address their respective problems [27]. For instance, systematic search methods explore all combinations of the structural parameters. The number of possible combinations grows exponentially as the degrees of freedom associated with the ligand increase, resulting in a phenomenon known as combinatorial explosion. Docking programs such as FRED, Surflex and DOCK solve this problem by applying an incremental construction algorithm in which the ligand is gradually

built in the binding site (Figure 4) [30–32]. In this strategy, the chemical structure is initially broken into several fragments (Figure 4A). Next, one of these parts is selected as an anchor fragment and is docked in a complementary region of the binding site (Figure 4B) while the remaining fragments are sequentially added (Figure 4C–E). The process continues until the entire ligand has been constructed. The algorithm performs the conformational search only for the fragments being added, reducing the degrees of freedom to be explored, and thereby avoiding combinatorial explosion [33].

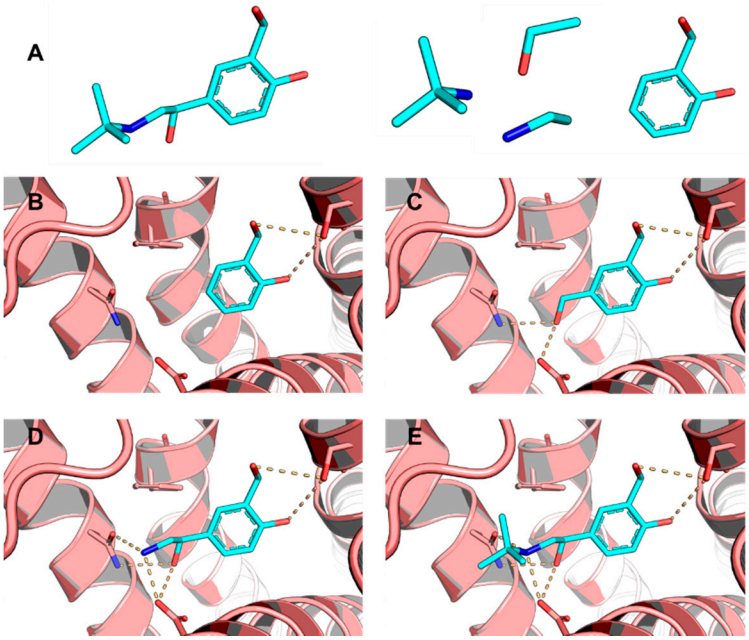

Figure 4. The incremental construction method. (**A**) The ligand (stick representation, carbon in cyan) is broken into several fragments; (**B**) The anchor fragment is docked in the binding site of the molecular target (cartoon representation, carbon in salmon); (**C**) The next fragment is docked after the anchor fragment; (**D** and **E**) The other fragments are docked sequentially to construct the entire ligand in its binding conformation. Residues in the active site are shown in stick representation (carbon in salmon). Hydrogen bonds are indicated as dashed lines.

Genetic algorithms (GA) are an interesting application of the stochastic search, which have been successfully used in molecular docking programs such as AutoDock and Gold [34,35]. The GA algorithm addresses the high computational cost associated with stochastic methods by applying concepts of the theory of evolution and natural selection. As a first step, the algorithm encodes all of the structural parameters of the initial structure in a chromosome, which is represented by a vector. Starting from this chromosome, the random search algorithm generates an initial population of chromosomes covering a wide area of the energy landscape. This population is evaluated and the most adapted chromosomes (*i.e.*, those with the lowest energy values) are selected as templates for the generation of the next population. This procedure decreases the average energy of the chromosome ensemble by transmitting favorable structural characteristics from one population to another, reducing therefore, the conformational space to be explored. The GA routine is recursively executed and, after a reasonable number of conformational search-and-evaluation cycles, converges to a conformation (chromosome) corresponding to the global energy minimum [36].

Regardless the specifics of each method, any conformational search algorithm should be able to explore a wide range of the energy landscape in a reasonable amount of time. Ideally, the evaluation of a modest set of molecules needs to be concluded in a few minutes. A list of widely used molecular docking algorithms categorized according to the conformational search methodology is provided in Table 1.

Table 1. Examples of conformational search algorithms.

Systematic Search	Random/Stochastic Search
eHiTS [28]	AutoDock [34]
FRED [30]	Gold [35]
Surflex-Dock [31]	PRO_LEADS [44]
DOCK [32]	EADock [45]
GLIDE [37]	ICM [46]
EUDOC [38]	LigandFit [47]
FlexX [39]	Molegro Virtual Docker [48]
Hammerhead [40]	CDocker [49]
Flog [41]	GlamDock [50]
SLIDE [42]	PLANTS [51]
ADAM [43]	MolDock [52]
	MOE_Dock [53]

3.2. Evaluation of Binding Energetics

Molecular docking programs use scoring functions to estimate the binding energetics of the predicted ligand-receptor complexes. The energy variation, due to the formation of the ligand-receptor structure, is given by the binding constant (K_d) and the Gibbs free energy (ΔG_L) [54]. Prediction of the binding energy is performed by evaluating the most important physical-chemical phenomena involved in ligand-receptor binding, including intermolecular interactions, desolvation and entropic effects [55]. Therefore, the greater the number of physical-chemical parameters evaluated, the greater the accuracy of the scoring function. However, the computational cost increases in proportion to the number of variables included in the function, a shortcoming that reduces the productivity of the docking algorithm. Ideally, efficient scoring functions should offer a balance between accuracy and speed, which is a critical aspect when working with large ligand sets.

Scoring functions are categorized in the three following groups: force-field-based, empirical, and knowledge-based functions [56]. Force-field-based scoring functions estimate the binding energy by summing the contributions of bonded (bond stretching, angle bending, and dihedral variation) and non-bonded terms (electrostatic and van der Waals interactions) in a general master function. This type of scoring function applies an *ab initio* method to calculate the energy associated with each term of the function using the equations of classical mechanics [57]. A major limitation of force-field-based methods is their inaccuracy in estimating entropic contributions. This shortcoming is due to the lack of a reasonable physical model to describe this phenomenon. Furthermore, the solvent is not explicitly considered, hindering the estimation of desolvation energies [58].

Empirical scoring functions are another type of evaluation method. Each term of the function describes one type of physical event involved in the formation of the ligand-receptor complex. These include hydrogen-bonding, ionic and apolar interactions, as well as desolvation and entropic effects [59]. As a

first step in the development of an empirical function, a series of protein-ligand complexes with known binding affinities is used as a training set to perform a multiple linear regression analysis. Then, the weight constants generated by the statistical model are used as coefficients that adjust the terms of the equation. A drawback of empirical scoring functions is their dependence on the accuracy of the data used to develop the model [60]. However, because of the simplicity of the employed energy terms, empirical functions are faster than force-field-based methods. Surflex and FlexX are broadly used molecular docking programs using empirical scoring functions [31,39].

A third approach used to evaluate ligand-receptor binding energy is the knowledge-based scoring functions. The method uses pairwise energy potentials extracted from known ligand-receptor complexes to obtain a general function [61]. These potentials are constructed by taking into account the frequency with which two different atoms are found within a given distance in the structural dataset. The different types of interactions observed in the dataset are classified and weighted according to their frequency of occurrence. The final score is given as a sum of these individual interactions. As knowledge-based functions do not rely on reproducing binding affinities (empirical methods) or *ab initio* calculations (force-field methods), they offer a suitable balance between accuracy and speed [62].

Every scoring function has its virtues and limitations. Therefore, the simultaneous use of different scoring methodologies has been increasingly employed as a way to obtain a consensus scoring [63]. This can be very useful, as it combines the advantages and simultaneously attenuates the shortcomings of each method [64]. Examples of consensus scoring functions are MultiScore, X-Cscore, GFscore, SCS, SeleX-CS and CONSENSUS-DOCK [65–70]. Table 2 provides a list of several scoring functions implemented in the most frequently used molecular docking programs.

Most docking programs are able to successfully predict the conformation of the ligand within the target binding site, as can be confirmed by comparison of predicted complexes with their corresponding crystallographic data. However, most programs do not reproduce the absolute interaction energy of the ligand-receptor complex with satisfactory agreement. Issues such as desolvation and entropic effects are examples of the challenges to be overcome by the current docking algorithms [71,72].

Table 2. Examples of scoring functions implemented in widely used molecular docking programs.

Force-Field-Based	Empirical	Knowledge-Based
DOCK [32]	AutoDock [34]	SMoG [82]
AutoDock [34]	GlideScore [37]	DrugScore [62]
GoldScore [35]	ChemScore [60]	PMF_Score [83]
ICM [46]	X_Score [66]	MotifScore [84]
LigandFit [47]	F_Score [73]	RF_Score [85]
Molegro Virtual Docker [48]	Fresno [75]	PESD_SVM [86]
SYBYL_G-Score [73]	SCORE [76]	PoseScore [87]
SYBYL_D-Score [73]	LUDI [77]	
MedusaScore [74]	SFCscore [78]	
	HYDE [79]	
	LigScore [80]	
	PLP [81]	

3.3. Covalent Bonds in Molecular Docking

Covalent drugs have demonstrated to be opportune alternatives in several therapeutic areas such as cancer, diabetes, and infectious, cardio-vascular, gastro-intestinal and neurologic diseases. Recent reports have claimed that approximately one-third of the currently marketed enzyme modulators are covalent inhibitors [88]. Covalent ligands act by irreversibly inactivating their targets; consequently, recovery of the inhibited biological function involves re-synthesis of the targeted protein. Usually, covalent inhibitors bind to their molecular targets with high affinity, leading to a long-lasting pharmacological response, and consequently requiring less frequent administration [89]. Well-known drawbacks of covalent drugs such as toxicity, lack of specificity and high reactivity, have led most R & D programs to avoid such compounds [90]. This conception has been reconsidered and an increased interest in covalent inhibitors has been reported recently. As a result, diverse strategies have been developed to approach the binding of covalent small-molecule inhibitors. Covalent docking algorithms are aimed to explore the energy landscape available to the ligand when it is covalently linked to the receptor, as well as evaluate the binding energetics of the interaction [91]. Despite the recent resurgence of covalent drugs, molecular modeling methods devised to address the problem of covalent docking are not as developed as those dedicated to noncovalent docking [92].

Binding of covalent drugs has some differences from noncovalent molecular interaction, especially with respect to binding thermodynamics. Current molecular mechanics (MM) algorithms are able to predict with good accuracy noncovalent binding events. However, the formation of covalent bonds is not satisfactorily approached by these methods [93]. The issue of covalent-bond formation can be appropriately handled by quantum mechanical methods (QM), which are able to explore the whole reaction mechanism [92].

The problem of modeling covalent bonds in molecular docking has been targeted by widely used molecular docking programs such as DOCK [32], AutoDock [34] and Gold [35]. Each of these programs employs a particular approach to manage covalent docking. Gold, for instance, attempts to mimic the covalent bond formation by defining an atom in both the ligand and the receptor to play the role of "link atoms" [94]. Subsequently, the ligand link atom is overlaid on the protein link atom and the geometry of the covalent bond is evaluated by specific terms of the scoring function (clash, torsion and valence-angle bending terms). Another program—DOCKovalent—is an adaptation of DOCK3.6 aimed to perform large-scale, covalent virtual screening [95]. The algorithm defines *a priori* a covalent attachment point and systematically explores the ligand conformational space around the modeled covalent bond. Each conformation is ranked with the default scoring function implemented in DOCK3.6. Another approach is a recent adaptation of AutoDock4, which proposes the so-called two-point attractor method for covalent docking [96]. The default AutoDock routine consists of the calculation of an interaction energy map, constructed by using several probe atoms; and a subsequent conformational search that uses these maps as reference tables to evaluate the binding energetics. The two-point attractor approach works as follows: first, the two terminal atoms of the residue covalently bound to the ligand are removed. Next, this fragment is attached to the correct atom of the ligand, and labeled with two specific atoms types (A and B). Then, a Gaussian function is employed to generate modified interaction maps for these atoms, centered on their original location in the covalently bound amino acid residue. These interaction energy maps penalize ligand conformations in which A or B are not properly placed in their original positions.