MolDock Applied to Structure-Based Virtual Screening

Article *in* Current Drug Targets · March 2010

DOI: 10.2174/138945010790711941 · Source: PubMed

1 author:

Walter Filgueira De Azevedo Jr.
Pontifícia Universidade Católica do Rio Grande do Sul
200 PUBLICATIONS **7,161** CITATIONS

SEE PROFILE

MolDock Applied to Structure-Based Virtual Screening

Walter Filgueira De Azevedo Jr. [*,a,b]

[a]*Faculdade de Biociências, Instituto Nacional de Ciência e Tecnologia em Tuberculose-CNPq, Laboratório de Bioquímica Estrutural (LaBioQuest), Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS, Porto Alegre, Brazil*

[b]*Programa de Pós-Graduação em Medicina e Ciências da Saúde. Programa de Pós-Graduação em Biologia Celular e Molecular. Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre – RS, Brazil*

Abstract: Molecular docking is a simulation process where the binding of a small molecule is identified in the structure of a protein target. There are several different computational approaches to solve this problem. Here it is described recent developments in application of evolutionary algorithms to molecular docking simulations. Evolutionary algorithms are classified as a group of computational techniques based on the concepts of Darwin's theory of evolution that are designed to find the best possible solution to optimization problems. A successfully implementation of this algorithm can be found in the program MolDock. The main features of MolDock are reviewed here. We also describe application of MolDock to purine nucleoside phosphorylase, shikimate kinase and cyclin-dependent kinase 2.

Keywords: Evolutionary algorithms, molecular docking, structure-based virtual screening, protein-ligand, docking, CDK2, shikimate kinase, PNP.

INTRODUCTION

Molecular docking is a simulation process to predict the conformation of a receptor-ligand complex, where the receptor can be a protein and the ligand a small molecule. It can also be defined as a simulation process where a ligand position is estimated in a predicted or pre-defined binding site in the receptor molecule. All structure-based virtual screening projects are based on the hypothesis that it is possible to computationally determine the three-dimensional structure of binary complexes involving a protein and a ligand. For recent reviews see [1-5].

In the last decade, data concerning biological activity, structure and enzyme inhibition available have increased considerably. Structure databases, such as Protein Data Bank (PDB) [6, 7] has over 61,085 protein structures (October, 31st 2009), many of these structures may be considered as potential drug targets. There are also specific databases where structures of binary complexes are available as well as information about their binding affinities, such as in PDBBIND [8, 9], PLD [10], AffinDB [11], and BindDB [12]. This wealth of data about three-dimensional experimental structure and affinity data has been employed as source of information for development of docking algorithms and validation.

In addition, there are many small molecule databases, such as ZINC [13], PubChem [14], ChemDB [15], and DrugBank [16, 17], which together surpass one million of deposited structures of potential small-molecule ligands. The procedure of structure-based virtual screening (SBVS) through docking has become crucial when it is necessary to test a database of thousands (or even millions) of compounds against one or more protein targets in a feasible computational time.

Speed and accuracy are key features for obtaining a successful result in molecular docking simulations. The main objective in development of a docking algorithm is to obtain a fast method which is able to discover the novel lead compound (in SBVS) or reproduce experimental conformation (for validation with experimental data) at higher accuracy as possible. There are several molecular docking programs, such as DOCK [18], AUTODOCK [19, 20], GOLD [21, 22], FLEXX [23, 24], ZDOCK [25], M-ZDOCK [26], MS-DOCK [27], Surflex [28], MCDOCK [29], MolDock [30], and GemDock [31] and others. Recent comparison of the major docking programs indicates that programs based on evolutionary algorithms, such as GemDock and MolDock present better overall performance when compared with Flexx, GOLD and Surflex [30, 31].

The present review is focused on the application of the recently development program, MolDock, to SBVS initiatives. We briefly discuss the main feature of MolDock algorithm and also application of MolDock to identify new inhibitors for protein targets.

REDOCKING

Every application of docking simulations to SBVS needs a preliminary stage where the molecular docking protocol is tested. In order to check whether a docking strategy is capable of recover a crystallographic structure, we compare the structure obtained from docking simulation (pose) against the ligand position obtained from the crystallographic structure of a binary complex involving protein and ligand. This process is called redocking. Once the poses are generated each position is compared with the reference ligand (crystallographic structure) with RMSD. This RMSD is calculated between two sets of atomic coordinates, in this

*Address correspondence to this author at the Av. Ipiranga, 6681, CEP 90619-900, Porto Alegre, Rio Grande do Sul, Brazil; Tel:+55 51 3353-4529; Fax:+55 51 3320-3629; E-mail: walter@azevedolab.net; Web: http://azevedolab.net

1389-4501/10 $55.00+.00

case, one for the crystallographic structure (x_c,y_c,z_c) and another for the atomic coordinates obtained from the docking simulations (x_d,y_d,z_d), the summation is taken over all N atoms being compared, the equation is as follows:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ci} - x_{di})^2 + (y_{ci} - y_{di})^2 + (z_{ci} - z_{di})^2}$$

In docking simulations we expect that the best results generate RMSD values below 1.5 Å, when compared to crystallographic structures. Nevertheless, this cutoff value may vary, depending on the number of rotatable bonds present in the ligand to be docked.

Evolutionary Algorithms Applied to Molecular Docking

We may define evolutionary algorithms (EA) as a group of computational approaches based on the concepts of Darwin's theory of evolution that are designed to find optimal solution to problems [32]. Regardless of its definition, EAs are heuristic algorithms. They tend to find the best or one of the best solutions, but also they can be trapped in the local optimal solutions, unable to find the global best result. We can say that in EAs, the evolutionary course is simplified, and consequently it has very little in common with genuine world evolution.

Essentially an EA consists of a population of individuals (candidate solutions) exposed to random variation by means of variation operators, such as mutation and recombination. The individual being changed is frequently referred to as the parent and the resultant solution after modification is called the offspring. Occasionally, more than one parent is employed to make the offspring by recombination of solutions, which is referred to as crossover. Many authors arrange these algorithms in three classes [33]: evolution strategies (ES), evolutionary programming, and genetic algorithms (GA). Being GA the first to be proposed and applied to molecular docking simulations [19, 20]. There are key points common to all algorithms but some huge differences in the implementation of evolutionary concepts. Common to all of them is the basic idea of creating a population of candidate solutions to the problem. The members of the population are scored applying a fitness function that measures the excellence of these candidates. The population is modified over the time and may evolve towards fitter candidates. This procedure of producing new candidates is generally referred as 'breeding', with new candidates being the 'offspring' that are generated from the 'parents" of the previous iteration. Here we need to make a small note to explain how all algorithms will be described in the present review. All evolutionary algorithms described here are iterative; that is, we begin from some candidates, and then proceed further in cycles, according to an algorithm hopefully towards a better solution. Each cycle is known as iteration. Each EA will be described here are employed to the general problem of finding the global minimum energy, in docking applications the best binary complex closer to the crystallographic structure.

Perhaps the best-known of the three classes is the genetic algorithm (GA) [32]. The main idea behind a GA is to follow the biological process of evolution in choosing the path to arrive at an optimum configuration of a given complex system. For instance, for an interacting many-body system, the equilibrium is reached by moving the system to the configuration that is at the global minimum on its potential energy surface [34]. As we can see in Fig (**1**) where we have a rugged-potential energy surface, finding the global minimum may demand intense computing if we are going to perform an exhaustive search for the global minimum. Therefore bio-inspired strategies to evaluate the landscape may reduce the CPU time among these bio-inspired strategies we have GAs.

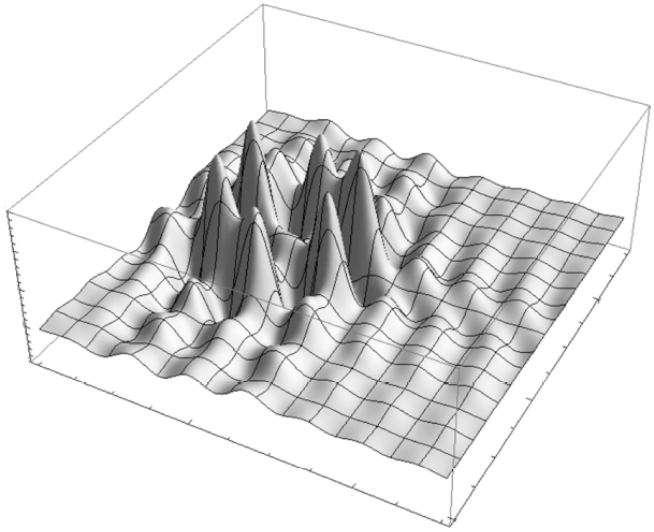

Fig. (1). Rugged-potential energy surface.

Several steps are involved in a canonical GA. First we need to generate an initial population of configurations, which is called the initial gene pool of N possible solutions. In conformational analysis, this initial population would match to a set of randomly generated conformations of the molecule. More specifically, for molecular docking simulations this would be a set of conformations of the ligand to be docked in the protein target. Each member of the gene pool (population) is called by a 'chromosome', which is commonly stored as a binary function. The chromosomes codes for the values of the torsion angles of the rotatable bonds in the ligand, as we can see in Fig. (**2**). Initial gene pool is most easily generated by randomly setting bits to 0 or 1 in the chromosomes. After decoding each chromosome and assigning the torsion angles to the appropriate values in the ligand, the fitness of each member of the population can be determined. This fitness function may be an internal potential energy function. Then we have to choose some members of the gene pool to be parents for reproduction. The operator that can mix the genes of the two parents is called crossover, which reflects how genetics attributes are passed on. In order to create true offspring, each of the parent chromosomes is cut into segments that are exchanged and attached together to generate the new chromosomes of the offspring. There is yet another operator, called mutation operator that carry out mutations on selected chromosomes. Afterward we permit a certain percentage of bits in the chromosome to mutate. This completes one cycle of the GA. The new gene pool then becomes the current population ready for a new cycle. The GA repeatedly applies this sequence for a predetermined number of iterations and/or until in converges.

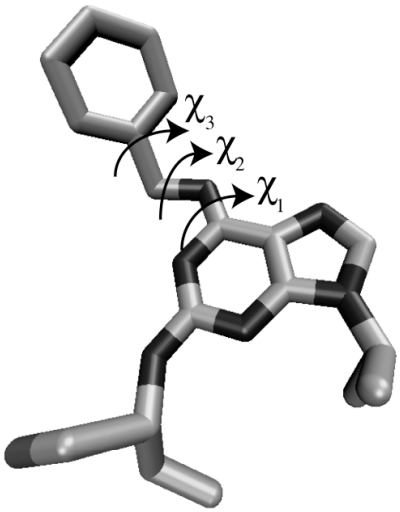

Fig. (2). Torsion angles of the rotatable bonds in the ligand

The main difference between EP and GA is that the former does not employ the crossover operator. It applies exclusively the mutation operator to generate new individuals from parents. Furthermore, individuals in EP are usually represented by means of a sequence of real numbers stored in a vector, instead of employing a binary function (zero and one). At the beginning of each iteration of the EP algorithm one child is bred from each of the members of the current gene pool, applying the mutation operator. During mutation the real number in the chromosome is changed to a new real number generally taken from a Gaussian distribution. We have then N children and N parents to be evaluated using the fitness function, all of them compete for survival into the next generation.

To select the individuals for the next generation tournaments are carried out. This tournament is a session where each individual is compared to a number of Z opponents selected at random from the 2N population of offspring and parents. The individuals are then classified according to the number of victories they obtained in the tournaments and the suitable number is chosen from the top of the set to generate the population to be employed in the next iteration. One key point in this methodology is related to the compromise between premature convergence and the CPU time spent to find a solution, therefore special care should be taken to select an appropriate number of opponents (Z). This can be done selecting a training set, where several attempts are performed until an optimal number Z is reached. This number is kept as a default parameter to be used in a new docking simulation.

Evolutionary strategies (ES) are very similar to EP, but diverge in two key features. First, the tournament is replaced by a direct ranking and, second, the crossover operators introduced in the GAs are now included in algorithm.

Overview of MolDock Algorithm

MolDock [30] is an implementation of EA, focused on molecular docking simulations [32]. Computational approximations of an evolution process, called genetic operators, are applied to simulate the permanence of the most favorable features. In a sample space, where there is a problem or a search routine and many different possible solutions (candidates), each option is ranked based on a set of parameters (scoring function, or fitness function), and only the best ranked solutions are kept for the next iteration. This process is repeated until an optimal solution can be found.

The program MolDock makes use of a slight variation of the EA, which is called guided differential evolution algorithm. This methodology is based on an EA modification called differential evolution (DE), which brings a different method to choose and alter candidate solutions (individuals). The major original idea in DE is to generate offspring from a weighted difference of parent solutions. The DE works as follows. In the first step, all individuals are initialized and evaluated according to the fitness function. Afterward, the following process will be carried out if the termination condition is not satisfied. For each individual in the population, an offspring is created by adding a weighted difference of the parent solutions, which are randomly chosen from the population. After that the offspring replaces the parent, if and only if it is fitter. Otherwise, the parent survives and is passed on to the next generation (iteration of the algorithm). The termination condition is reached when the current number of fitness (energy) evaluations performed exceeded the maximum number of evaluations allowed (max evaluations parameter setting). Furthermore, early termination was permitted if the variance of the population was below a certain threshold (0.01 here). Moreover guided differential evolution employs a cavity prediction algorithm to limit predicted conformations (poses) during the search procedure. More specifically, if a candidate solution is placed outside the cavity, it is translated so that a randomly chosen ligand atom will be located within the region spanned by the cavity. Fig. (**3**) shows the cavities predicted by MolDock. Obviously this strategy is only employed if a cavity has been found. If no cavities are reported, the search process does not limit the candidate solutions.

Fig. (3). Graphical interface with the cavities (indicated by ellipses) identified by MolDock.

In MolDock, only the ligand properties are represented in the individuals since the protein remains rigid during the docking simulation. Consequently a candidate solution is determined by a vector of real-valued numbers representing ligand position, orientation, and conformation as Cartesian coordinates for the ligand translation, four variables specifying the ligand orientation (encoded as a rotation vector and a rotation angle), and one angle for each flexible torsion angle

in the ligand (if present). For each individual in the initial population, each of the three translational parameters (encoded as a position relative to the crystallographic native ligand) for *x*, *y*, and *z* is assigned an evenly distributed random number between -15.0 and 15.0 Å, which is added to the center of the crystallographic reference ligand. Initializing the orientation is carried out the Shoemake [35] for generating uniform random quaternions and converted these quaternions to their rotation axis/rotation angle representation. The flexible torsion angles (if present) are given a random angle ranging from -180° to +180°.

The scoring function used by MolDock is derived from the piecewise linear potential PLP scoring functions [31]. The scoring function used by MolDock further improves these scoring functions with a new hydrogen bonding term and new charge schemes [30]. Based on above described EA classification MolDock algorithm may be classified as an ES, since it employs direct ranking of the solutions and the crossover operators. MolDock showed better overall performance in docking simulations when compared with Surflex [28], Flexx [23, 24] and GOLD [21, 22]. MolDock presents a very friendly interface, which facilitates its use. In addition, MolDock is available for linux, windows and MAC OSX. MolDock graphical interface with the reference ligand can be seen in Fig. (**4**). Fig. (**5**) shows the results obtained from the docking simulations, five best results are shown in the binding pocket.

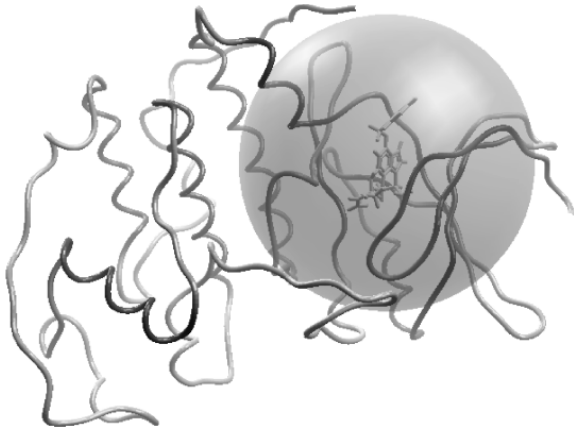

Fig. (4). MolDock graphical interface with the reference ligand centered at binding site indicated by gray sphere.

Fig. (5). MolDock graphical interface with docking results (5 results).

In the next sections we describe application of this docking algorithm to three important protein targets that have been used in SBVS. The binary complexes involving CDK2, PNP and SK were submitted to redocking in order to identify the best docking strategy.

Human Cyclin-Dependent Kinase 2

CDK2 (EC 2.7.11.22) is a key enzyme, which is responsible for the control of cell cycle progress. This enzyme is inactive as monomer as its activation requires binding to cyclin a protein which level oscillates during cell cycle [36-48]. Since deregulation of cyclins and/or modification or absence of CDK inhibitors (CKIs) have been connected with many cancers, there is strong interest in chemical inhibitors of CDKs that could play a central role in the discovery of new family of antitumor agents [41]. The elucidation of the crystallographic structure of human CDK opens the possibility for SBVS. Analysis of CDK2 structure indicated the presence of a hydrophobic pocket, between the two domains. This site allows a surprising wide range of different molecules to bind to CDK2. Fig. (**6**) indicates the surface of the CDK2 generated by MolDock.

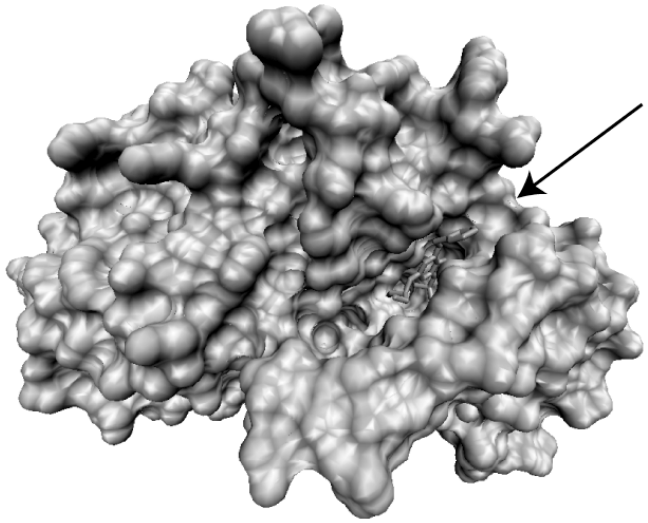

Fig. (6). Molecular surface of human CDK2. ATP-binding pocket is indicated by an arrow.

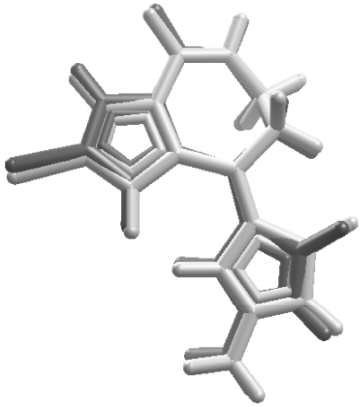

Fig. (7). Docking simulation of the structure of CDK2 in complex with hymenialdisine. Docking result (dark gray) and crystallographic structure (light gray).

We applied the above described molecular docking protocol to the structure of CDK2 in complex with hymenialdisine (PDB access code: 1dm2) [48]. We used the default protocol available in the program MolDock (Table **1**), and the lowest energy result is shown in Fig. (**7**). The RMSD is 0.48 Å, which indicate that the present molecular docking protocol is adequate for this protein target and can be used in SBVS.

Especially interesting is the analysis of the intermolecular hydrogen bonds. Crystallographic structures CDK2s present at least two intermolecular hydrogen bonds. These bonds involve main-chain atoms of Glu 81 and Leu 83. Analysis of the prevalence of the hydrogen bonds in the docked structure shows that intermolecular interactions with the CDK molecular fork are preserved.

Human Purine Nucleoside Phosphorylase

PNP is a protein target for drug development, which could provoke immune suppression to treat, for example autoimmune diseases, T-cell leukemia, lymphoma and organ transplantation rejection. In addition, PNP inhibitors can also be employed to prevent cleavage of anticancer and antiviral drugs, since many of these drugs imitate natural purine nucleosides and can thereby be cleaved by PNP prior to accomplishing their therapeutic function [49-50]. This enzyme catalyzes the reversible phosphorolysis of N-ribosidic bonds of both purine nucleosides and deoxynucleosides, except adenosine, generating purine base and ribose (or deoxyribose) 1-phosphate [49-66]. Furthermore, PNP cleaves glycosidic bond with inversion of configuration to generate alpha-ribose 1-phosphate.

We used the default protocol available in the program MolDock (Table **1**), to perform redocking of the inhibitor immucillin-H against the structure of human PNP. The lowest energy result is shown in Fig. (**8**). The RMSD 1.5 Å, which indicate that the protocol is fair for this protein target. This result can be significantly improved increasing the number of individual in the population (Max population size)(data not shown).

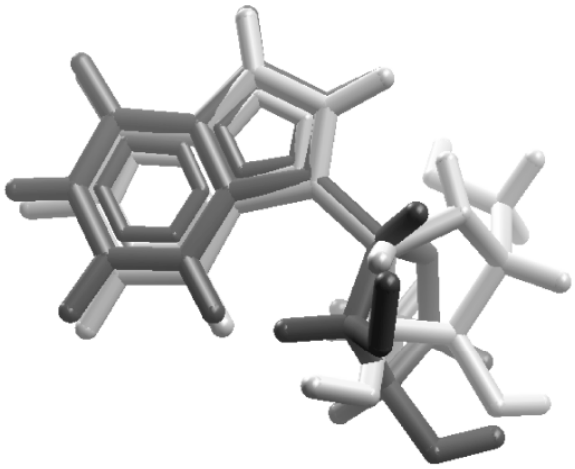

Fig. (8). Docking simulation of the structure of PNP in complex with immucillin-H. Docking result (dark gray) and crystallographic structure (light gray).

Analysis of the intermolecular hydrogen bonds present in the docking structure indicated the participation of residues His 86, Tyr88, Glu201, Met219, Thr242, Asn243 and His257. Detailed inspection of the intermolecular hydrogen bonds present in the docked structure strongly indicates the prevalence of most of the intermolecular hydrogen bonds identified in the crystallographic structure.

Shikimate Kinase

Enzymes of shikimate pathway are appealing targets for development of antibacterial drugs [67], since this pathway is vital for bacteria, whereas it is not present in mammals [68-80]. Therefore in bacterial diseases, inhibition of any of shikimate pathway enzymes is improbable to cause toxic side effects on the host. Furthermore, the value of shikimate pathway can be indicated by the finding that deletion of the *aroA* gene, which codes EPSPS, causes *Streptomyces pneumoniae* and *Bordetella bronchiseptica* strains to be attenuated for virulence [72, 74]. The shikimate route is composed of seven enzymatic steps, the fifth enzyme of this pathway is shikimate kinase (EC 2.7.1.71), which catalyses the phosphorylation of the 3- hydroxyl group of shikimic acid (shikimate) using ATP as a co-substrate.

Shikimate kinase (SK) is a member of nucleoside monophosphate kinases (NMP kinase), and a typical feature of the NMP kinases is that they undergo large conformational changes during catalysis, typified by adenylate kinase [72]. Members of this class of enzyme are responsible for catalysis of the transfer reaction of the terminal phosphoryl group from a nucleoside triphosphate, typically ATP, to the phosphoryl group on a nucleoside monophosphate. In the case of SK the substrate is shikimate.

Fig. (**9**) shows the structure of MtSK complexed with shikimate, where we can clearly see the LID domain. Previously published analysis indicated that the LID domain

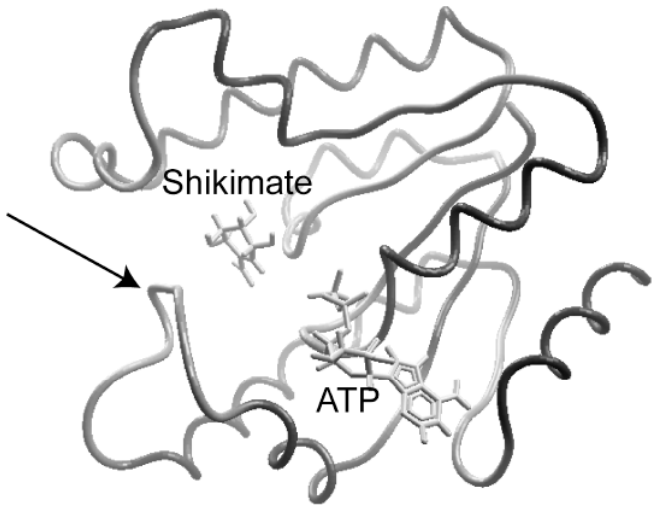

Fig. (9). Structure of MtSK. The lid domain is indicated by an arrow.

undergoes conformation change in the structure without shikimate [72]. The structure of MtSK without the molecule of shikimate presents high flexibility for the residues involved in the LID domain. Analysis of the residues participating in the LID domain in the complex MtSK-shikimate indicates the presence of a hydrogen bond network involving residues 112- 124 and the shikimate, which promotes the stabilization

Table 1. Default Parameters Used in the Docking Simulations with MolDock [31]

Scoring function
Score: MolDock Score(GRID)
Grid resolution (Å): 0.30
Ligand evaluation:
Binding site
Origin: Reference ligand
Center:
Radius:
Search algorithm
Algorithm: MolDock SE
Number of runs: 10
Constrain poses to cavity: yes
After docking: Optimize H-bonds
Parameter settings
Max iterations: 1500
Max population size: 50
Pose generation
Energy threshold: 100.0
Tries. Min: 10 Quick: 10 Max: 30
Simplex evolution
Max steps: 300
Neighbor distance factor: 1.00
Return multiple poses for each run
Max number of poses returned: 5
Enable energy threshold: No
Cluster similar poses. RMSD threshold: 1.00
Ignore similar poses (for multiple runs only). RMSD threshold: 1.00

of the LID domain in the closed conformation. Molecular dynamics simulations with shikimate showed the movement of the LID domain in the MtSK structure. The LID domain is moved towards the structure at the final of the molecular dynamics simulation, demonstrating that the molecular dynamics simulation results are in agreement with experimental X-ray crystallographic data, confirming that the binding of the shikimate to the SK structure is enough to promote the closure of the LID domain. On the other hand, the lacking of the shikimate in the MtSK promotes the opening of the LID domain. This structural feature should be considered in any molecular docking simulations against SK inhibitors.

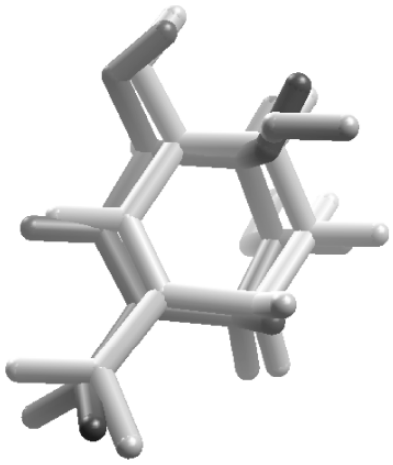

Fig. (10). Docking simulation of the structure of SK in complex with shikimate. Docking result (dark gray) and crystallographic structure (light gray).

The atomic coordinates of MtSK in complex with shikimate were used in the redocking simulations (PDB access code: 1u8a). We used the default protocol available in the program MolDock (Table **1**), to perform redocking of the shikimate against the structure of MtSK. The lowest energy result is shown in Fig. (**10**). The RMSD 0.7 Å, which indicate that the protocol is adequate for this protein target.

FINAL REMARKS

One key point in the development of docking algorithms is the accuracy of the docking simulation. The accuracy may vary depending on what target is being tested and what kind of molecules composes the screening library. Highest speed and highest accuracy are ideal, although opposite features for virtual screening through docking simulations. Methods which are more complex, considering many physicochemical and thermodynamic properties tend to present higher accuracy. However these methods consume more CPU time. This is exactly the case when we use molecular dynamics simulations to evaluate ligand-biding affinity [81-83]. Likewise, methods which take into account simpler parameters, as shape matching algorithms, are able to predict docking conformations in fast speed, however at lower accuracy rate. Nevertheless, molecular docking simulations based on EA have shown to be capable to generate poses with low RMSD. In the structures analyzed in the present review they present RMSD ranging from 0.48 to 1.5 Å, indicating the predicting power of the MolDock.

One of the reasons why MolDock works properly is that the variation operator exploits the population variety in the following manner: at first when the candidate solutions in the population are randomly generated, the diversity is large. Consequently when offspring are created, the differences between parental solutions are high, resulting in large step sizes being employed. As the algorithm converges to superior solutions, the population diversity is reduced and the step sizes used to create offspring are correspondingly lowered. Thus, by using the differences between other individuals in the population, DE automatically adapts the step sizes employed to create offspring as the search procedure converges toward good solutions.

ACKNOWLEDGEMENTS

The author would like to thank Ms. Saima Nayab (Senior Manager for Publications of Current Drug Targets) for her help on preparation of this issue. This work was supported by grants from CNPq and Instituto Nacional de Ciência e Tecnologia do Conselho Nacional de Desenvolvimento Científico e Tecnológico-Ministério de Ciência e Tecnologia (INCT-Tuberculose, CNPq-MCT, Brazil). WFA is senior researcher for CNPq (Brazil).

ABBREVIATIONS

CDK2 = Cyclin-dependent kinase 2

DE = differential evolution

EA = Evolutionary algorithm

EP = Evolutionary programming

GA = Genetic algorithms

PDB = Protein Data Bank

PNP = Purine nucleoside phosphorylase

RMSD = Root mean square deviation

SK = Shikimate kinase

REFERENCES

[1] De Azevedo Jr WF, Dias R. Computational methods for calculation ligand-binding affinity. Curr Drug Targets 2008; 9: 1031-39.

[2] Dias R, De Azevedo Jr WF. Molecular docking algorithms. Curr Drug Targets 2008; 9: 1040-47.

[3] Dias R, Timmers LFSM, Caceres RA, De Azevedo Jr WF. Evaluation of molecular docking using polynomial empirical scoring functions. Curr Drug Targets 2008; 9: 1062-70.

[4] Canduri F, De Azevedo Jr WF. Protein crystallography in drug discovery. Curr Drug Targets 2008; 9: 1048-53.

[5] De Azevedo Jr WF, Dias R. Experimental approaches to evaluate the thermodynamics of protein-drug interactions. Curr Drug Targets 2008; 9: 1071-76.

[6] Berman HM. The Protein Data Bank: a historical perspective. Acta Crystallogr A 2008; 64: 88-95.

[7] Westbrook J, Ito N, Nakamura H, Henrick K, Berman HM. PDBML: the representation of archival macromolecular structure data in XML. Bioinform 2005; 21: 988-92.

[8] Wang R, Fang X, Lu Y, Yang CY, Wang S. The PDBbind database: methodologies and updates. J Med Chem 2005; 48: 4111-19.

[9] Wang R, Fang X, Lu Y, Wang S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem 2004; 47: 2977-80.

[10] Puvanendrampillai D, Mitchell JB. L/D Protein Ligand Database (PLD): additional understanding of the nature and specificity of protein-ligand complexes. Bioinform 2003; 19: 1856-7.

[11] Block P, Sotriffer CA, Dramburg I, Klebe G. AffinDB: a freely accessible database of affinities for protein-ligand complexes from the PDB. Nucleic Acids Res 2006; 34: D522-6.

[12] Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. Nucleic Acids Res 2007; 35(D): 198-201.

[13] Irwin JJ, Shoichet BK. ZINC--a free database of commercially available compounds for virtual screening. J Chem Inf Model 2005; 45: 177-82.

[14] Rosania GR, Crippen G, Woolf P, States D, Shedden K. A cheminformatic toolkit for mining biomedical knowledge. Pharmac Res 2007; 24: 1791-802.

[15] Chen JH, Linstead E, Swamidass SJ, Wang D, Baldi P. ChemDB update--full-text search and virtual chemical space. Bioinform 2007; 23: 2348-51.

[16] Wishart DS, Knox C, Guo AC, *et al*. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 2006; 34: 668-72.

[17] Wishart DS. DrugBank and its relevance to pharmacogenomics. Pharmacogenetics 2008; 9: 1155-62.

[18] Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 2001; 15: 411-28.

[19] Morris GM, Goodsell DS, Halliday RS, *et al*. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem 1998; 19: 1639-62.

[20] Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 1996; 9: 1-5.

[21] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins 2003; 52: 609-23.

[22] Joy S, Nair PS, Hariharan R, Pillai MR. Detailed comparison of the protein-ligand docking efficiencies of GOLD, a commercial package and ArgusLab, a licensable freeware. In Silico Biol 2006; 6: 601-5.

[23] Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996; 261: 470-89.

[24] Kramer B, Rarey M, Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins 1999; 37: 228-41.

[25] Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins 2003; 52: 80-7.

[26] Pierce B, Tong W, Weng Z. M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. Bioinform 2005; 21: 1472-8.

[27] Sauton N, Lagorce D, Villoutreix BO, Miteva MA. MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. BMC Bioinform 2008; 9: 184-96.

[28] Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 2003; 46: 499-511.

[29] Liu M, Wang S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. J Comput Aided Mol Des 1999; 13: 435-51.

[30] Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. J Med Chem 2006; 49: 3315-21.

[31] Yang JM, Chen CC. GEMDOCK: a generic evolutionary method for molecular docking. Proteins 2004; 55: 288-304.

[32] Goldberg DE. Genetic Algorithms in Search, Optimization, and Machine Learning, New York: Addilson-Wiley 1989.

[33] Leach AR. Molecular Modelling. Principles and Applications, 2nd ed. Essex: Pearson Prentice Hall 2001.

[34] Pang T. An Introduction to Computational Physics, 2nd ed. New York: Cambridge University Press 2006.

[35] Shoemake K. In Graphics Gems III; Kirk D, Ed. Boston: AP Professional (Academic Press) 1992; pp 124-32.

[36] De Azevedo Jr WF, Mueller-Dieckmann HJ, Schulze-Gahmen U, Worland PJ, Sausville E, Kim SH. Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase. Proc Natl Acad Sci USA 1996; 93: 2735-40.

[37] De Azevedo Jr WF, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim SH. Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. Eur J Biochem 1997; 243: 518-26.

[38] De Azevedo Jr WF, Canduri F, da Silveira NJ. Structural basis for inhibition of Cyclin-Dependent Kinase 9 by flavopiridol. Biochem Biophys Res Commun 2002; 293: 566-71.

[39] De Azevedo Jr WF, Gaspar RT, Canduri F, Camera Jr JC, Silveira NJF. Molecular model of cyclin-dependent kinase 5 complexed with roscovitine. Biochem Biophys Res Commun 2002; **297**:1154-8.

[40] Canduri F, Uchoa HB, De Azevedo Jr WF. Molecular models of cyclin-dependent kinase 1 complexed with inhibitors. Biochem Biophys Res Commun 2004; 324: 661-6.

[41] Canduri F, De Azevedo Jr WF. Structural basis for interaction of inhibitors with Cyclin-Dependent Kinase 2. Curr Computer-Aided Drug Des 2005; 1: 53-64.

[42] Manhani KK, Arcuri HA, Da Silveira NJ, Uchôa HB, De Azevedo Jr WF, Canduri F. Molecular models of protein kinase 6 from *Plasmodium falciparum*. J Mol Mod 2005. **12**: 42-8.

[43] Krystof V, Cankar P, Hajduch M, *et al*. 4-Arylazo-3,5-diamino-1H-pyrazole CDK Inhibitors: SAR Study, Crystal Structure in Complex with CDK2, Selectivity, and Cellular Effects. J Med Chem 2006; 49: 6500-9.

[44] Leopoldino AM, Canduri F, Cabral H, *et al*. Expression, purification, and circular dichroism analysis of human CDK9. Protein Expr Purif 2006; **47**: 614-20.

[45] Canduri F, Perez PC, Caceres RA, De Azevedo Jr WF. Protein kinases as targets for antiparasitic chemotherapy drugs. Curr Drug Targets 2007; 8: 389-98.

[46] Canduri F, Perez PC, Caceres RA, De Azevedo Jr WF. CDK9 a Potential Target for Drug Development. Med Chem 2008; 4: 210-8.

[47] Perez PC, Caceres RA, Canduri F, De Azevedo Jr WF. Molecular modeling and dynamics simulation of human cyclin-dependent kinase 3 complexed with inhibitors. Comput Biol Med 2009; 39: 130-40.

[48] Gray NS, Wodicka L, Thunnissen AM, *et al*. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. Science 1998; 281: 533-8.

[49] De Azevedo Jr. WF, Canduri F, Santos DM, *et al*. Crystal structure of human purine nucleoside phosphorylase at 2.3 A resolution. Biochem Biophys Res Commun 2003; 308: 545-52.

[50] Santos DM, Canduri F, Pereira JH, *et al*. Crystal structure of human purine nucleoside phosphorylase complexed with acyclovir. Biochem Biophys Res Commun 2003; 308: 553-9.

[51] De Azevedo Jr WF, Canduri F, Dos Santos DM, *et al*. Structural basis for inhibition of human PNP by immucillin-H. Biochem Biophys Res Commun 2003; 309: 917-22.

[52] De Azevedo Jr WF, Santos GC, Santos DM, *et al*. Docking and small angle X-ray scattering studies of purine nucleoside phosphorylase. Biochem Biophys Res Commun 2003; **309:** 923-8.

[53] De Azevedo Jr WF, Canduri F, dos Santos DM, *et al*. Crystal structure of human PNP complexed with guanine. Biochem Biophys Res Commun 2003; 312: 767-72.

[54] Canduri F, dos Santos DM, Silva RG, *et al*. Structures of human Purine Nucleoside Phosphorylase complexed with Inosine and ddI. Biochem Biophys Res Commun 2004; 313: 907-14.

[55] Da Silveira NJF, Uchoa HB, Canduri F, *et al*. Structural Bioinformatics Study of PNP from *Schistosoma mansoni*. Biochem Biophys Res Commun 2004; 322: 100-4.

[56] Nolasco DO, Canduri F, Pereira JH, *et al*. Crystallographic structure of PNP from *Mycobacterium tuberculosis* at 1.9 A resolution. Biochem Biophys Res Commun 2004; **324**: 789-94.

[57] Canduri F, Fadel V, Dias MVB, *et al*. Crystal structure of human PNP complexed with hypoxanthine and sulfate ion. Biochem Biophys Res Commun 2005; 326: 335-8.

[58] Canduri F, Fadel V, Basso LA ,Palma MS, Santos DS, De Azevedo Jr WF. New catalytic mechanism for human purine nucleoside phosphorylase. Biochem Biophys Res Commun 2005; 327: 646-9.

[59] Canduri F, Silva RG, dos Santos DM, *et al*. Structure of human PNP complexed with ligands. Acta Crystallogr D Biol Crystallogr 2005; 61: 856-62.

[60] Silva RG, Pereira JH, Canduri F, De Azevedo Jr WF, Basso LA, Santos DS. Kinetics and crystal structure of human purine nucleoside phosphorylase in complex with 7-methyl-6-thio-guanosine. Arch Biochem Biophys 2005; 442: 49-58.

[61] De Azevedo Jr WF, Canduri F, Basso LA, Palma MS, Santos DS. Determining the Structural Basis for Specificity of Ligands Using Crystallographic Screening. Cell Biochem Biophys 2006; 44: 405-11.

[62] Timmers LFSM, Caceres RA, Vivan, AL, *et al*. Structural studies of human purine nucleoside phosphorylase: towards a new specific empirical scoring function. Arch Biochem Biophys 2008; 479: 28-38.

[63] Caceres RA, Saraiva Timmers LF, Dias R, Basso LA, Santos DS, De Azevedo Jr WF. Molecular modeling and dynamics simulations of PNP from *Streptococcus agalactiae*. Bioorg. Med Chem 2008; 16: 4984-93.

[64] Timmers LFSM, Caceres RA, Basso LA, Santos DS, De Azevedo Jr WF. Structural Bioinformatics Study of PNP from *Listeria monocytogenes*. Prot Pept Lett 2008; **15:** 843-9.

[65] Pauli I, Timmers LFSM, Caceres RA, Basso LA, Santos DS, De Azevedo Jr WF. Molecular modeling and dynamics studies of Purine Nucleoside Phosphorylase from *Bacteroides fragilis*. J Mol Model 2009; 15(8): 913-22.

[66] Timmers LFSM, Caceres RA, Dias R, Basso LA, Santos DS, De Azevedo Jr WF. Molecular modeling, dynamics and docking studies of Purine Nucleoside Phosphorylase from ***Streptococcus pyogenes.*** Biophys Chem 2009; 142: 7-16.

[67] De Azevedo Jr WF, Canduri F, Simões de Oliveira J, *et al*. Molecular model of shikimate kinase from *Mycobacterium tuberculosis*. Biochem Biophys Res Commun 2002; **295**: 142-8.

[68] Pereira JH, Canduri F, de Oliveira JS, *et al*. Structural Bioinformatics Study of EPSP synthase from *Mycobacterium tuberculosis*. Biochem Biophys Res Commun 2003; 312: 608-14.

[69] Pereira JH, de Oliveira JS, Canduri F, *et al*. Structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimic acid. Acta Crystallogr D Biol Crystallogr 2004; 60: 2310-9.

[70] Arcuri HA, Canduri F, Pereira JH, *et al*. Molecular models for shikimate pathway enzymes of *Xylella fastidiosa*. Biochem Biophys Res Commun 2004; **320:** 979-91.

[71] Borges JC, Pereira JH, Vasconcelos IB, *et al*. Phosphate closes the solution structure of the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) from *Mycobacterium tuberculosis*. Arch Biochem Biophys 2006; **452**: 156-64.

[72] Pereira JH, Vasconcelos IB, Oliveira JS, *et al*. Shikimate kinase: A potential target for development of novel antitubercular agents. Curr Drug Targets 2007; 8: 459-68.

[73] Dias MVB, Faim LM, Vasconcelos IB, *et al*. Effects of magnesium and chloride ions and shikimate on the structure of shikimate kinase from *Mycobacterium tuberculosis*. Acta Crystallograph Sect F Struct Biol Cryst Commun 2007; **63**: 1-6.

[74] Dias MVB, Ely F, Palma MS, De Azevedo Jr WF, Basso LA, Santos DS. Chorismate synthase: An attractive target for drug development against orphan diseases. Curr Drug Targets 2007; 8: 437-44.

[75] Pauli I, Caceres RA, De Azevedo Jr WF. Molecular modeling and dynamics studies of shikimate kinase from Bacillus anthracis. Bioorg Med Chem 2008; 16: 8098-108.

[76] Arcuri HA, Borges JC, Fonseca IO, *et al*. Structural studies of shikimate 5-dehydrogenase from *Mycobacterium tuberculosis.* Proteins 2008; 72: 720-30.

[77] Dias MV, Borges JC, Ely F, *et al*. Structure of chorismate synthase from *Mycobacterium tuberculosis*. J Struct Biol 2006; **154:** 130-43.

[78] Marques MR, Vaso A, Neto JR, *et al*. Dynamics of glyphosate-induced conformational changes of *Mycobacterium tuberculosis* 5-enolpyruvylshikimate-3-phosphate synthase (EC 2.5.1.19) determined by hydrogen-deuterium exchange and electrospray mass spectrometry. Biochem 2008; 47: 7509-22.

[79] Marques MR, Pereira JH, Oliveira JS, *et al*. The inhibition of 5-enolpyruvylshikimate-3-phosphate synthase as a model for development of novel antimicrobials. Curr Drug Targets 2007; 8: 445-57.

[80] Barcellos GB, Caceres RA, De Azevedo Jr WF. Structural studies of shikimate dehydrogenase from *Bacillus anthracis* complexed with cofactor NADP. J Mol Model 2009; **15**: 147-55.

[81] De Amorim HL, Caceres RA, Netz PA. Linear interaction energy (LIE) method in lead discovery and optimization. Curr Drug Targets 2008; 9: 1100-5.

[82] Okimoto N, Futatsugi N, Fuji H, *et al*. High-performance drug discovery: computational screening by combining docking and molecular dynamics simulations. PLoS Comput Biol 2009; 5: e1000528.

[83] Canduri F, Teodoro LG, Fadel V, *et al*. Structure of human uropepsin at 2.45 A resolution. Acta Crystallogr D Biol Crystallogr 2001; 57: 1560-70.

Received: May 19, 2009 Revised: June 19, 2009 Accepted: August 24, 2009