

Octopus: a platform for the virtual high-throughput screening of a pool of compounds against a set of molecular targets

Eduardo Habib Bechelane Maia^{1,2} · Vinícius Alves Campos^{1,2} · Bianca dos Reis Santos^{1,2} · Marina Santos Costa^{1,2} · Iann Gabriel Lima^{1,2} · Sandro J. Greco³ · Rosy I. M. A. Ribeiro¹ · Felipe M. Munayer³ · Alisson Marques da Silva³ · Alex Gutterres Taranto¹

Received: 31 July 2016 / Accepted: 6 December 2016 / Published online: 7 January 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract Octopus is an automated workflow management tool that is scalable for virtual high-throughput screening (vHTS). It integrates MOPAC2016, MGLTools, PyMOL, and AutoDock Vina. In contrast to other platforms, Octopus can perform docking simulations of an unlimited number of compounds into a set of molecular targets. After generating the ligands in a drawing package in the Protein Data Bank (PDB) format, Octopus can carry out geometry refinement using the semi-empirical method PM7 implemented in MOPAC2016. Docking simulations can be performed using AutoDock Vina and can utilize the Our Own Molecular Targets (OOMT) databank. Finally, the proposed software compiles the best binding energies into a standard table. Here, we describe two successful case studies that were verified by biological assay. In the first case study, the vHTS process was carried out for 22 (phenylamino)urea derivatives. The vHTS process identified a metalloprotease with the PDB code 1GKC as a molecular target for derivative LE&007. In a biological assay, compound LE&007 was found to inhibit 80% of the activity of this enzyme. In the second case study, compound Tx001 was submitted to the Octopus routine, and the results suggested that *Plasmodium falciparum*

ATP6 (PfATP6) as a molecular target for this compound. Following an antimalarial assay, Tx001 was found to have an inhibitory concentration (IC₅₀) of 8.2 μM against PfATP6. These successful examples illustrate the utility of this software for finding appropriate molecular targets for compounds. Hits can then be identified and optimized as new antineoplastic and antimalarial drugs. Finally, Octopus has a friendly Linux-based user interface, and is available at www.drugdiscovery.com.br.

Keywords Structure-based drug design · Docking · Virtual screening

Introduction

The innovation process in the pharmaceutical industry is driven by the release of new drugs onto the market, as this process often involves modifying the structure of a known drug in an attempt to produce a new drug that is as active or more active towards a target receptor. [1]. First, pharmaceutical companies search for “hits;” a hit is a compound that shows activity towards a specific target under study, which is evaluated by performing biological and toxicological assays and structure–activity relationship (SAR) studies. The hit then becomes a lead compound. In this context, even though investment in the innovation process has increased, the availability of new drugs on the market has not followed suit. This is partly because the regulatory requirements associated with the approval of a new drug have also been increased in order to prevent pharmacological accidents (such as the birth defects resulting from thalidomide use by pregnant women). Furthermore, the high cost of the biological assays and the other methodologies used are limiting factors, especially for startups. This makes it critical to develop new approaches to identifying lead compounds [2].

This paper belongs to Topical Collection Brazilian Symposium of Theoretical Chemistry (SBQT 2015)

✉ Alex Gutterres Taranto
labqf_ufsj@ufsj.edu.br

¹ Universidade Federal de São João del Rei—Campus Centro-Oeste, Divinópolis, MG, Brazil

² Centro Federal de Educação Tecnológica de Minas Gerais—Campus Divinópolis, Divinópolis, MG, Brazil

³ Universidade Federal do Espírito Santo—UFES, Vitória, ES, Brazil

The most common strategies used to identify active compounds are analog design and systematic screening [3–6]. Analog design is a strategy widely used by research groups. It involves synthesizing analogs of active compounds that are currently on the market; these analogs are known as “me-too compounds.” Since proposed me-too compounds have very similar structures to active compounds, they have a good chance of also being active towards the desired target. This biological activity can be improved by optimizing the compound. However, analog design is generally considered to produce only incremental innovation. Amoxicillin is a good example of a drug obtained using this approach—it shows improved bioavailability compared to penicillin, which permits it to be administered orally. In other words, the analog design approach is ligand-focused.

On the other hand, systematic screening involves searching for evidence that a particular molecule or set of molecules, which may be natural or synthetic, has/have significant biological activity [7]. This is an exhaustive and time-consuming pharmacological investigative process. It is repeated until a compound with biological activity is identified. Recently, this methodology was mechanized. Such a high-throughput screening (HTS) process can screen several thousand compounds simultaneously using 30–50 different biochemical assays. When a hit compound is found, it can be submitted to a lead optimization process in the hope that it can ultimately be used as a lead compound. This approach has a good rate of success. However, this methodology can only be implemented by pharmaceutical companies or consolidated research groups. Efavirenz, delavirdine, and nevirapine are examples of antiviral drugs obtained by HTS that inhibit the reverse transcriptase of HIV. In this approach, drug discovery is target-focused [8, 9].

Even though HTS helps to get new drugs on the market, this approach is limited by its high cost, which has motivated the development of new technologies based on high-throughput screening and combinatorial synthesis. Considering the large number of biological targets available in the Protein Data Bank (PDB) [10] and the diverse libraries of compounds such as ZINC [11] that can be used to generate new drugs, structure-based virtual screening has shown itself to be a useful new strategy for identifying novel bioactive substances via molecular docking [12, 13]. Docking is an *in silico* method that is employed to identify hit compounds for a three-dimensional structure-of-interest receptor. Docking programs measure the affinities of small molecules (ligands) for a molecular target to determine their interaction energies with the target [14]. In addition, visualization software can be used to show the intermolecular interactions responsible for molecular recognition, such as those associated with the complex between the ligand and the receptor. As a result, docking can identify the most promising hits for biological assays and decrease the cost of drug development.

Docking approaches can be applied in two different contexts. First, a library of ligands normally obtained from analog design can be submitted to docking simulation to find the best hit for a specific molecular target. This approach is known as virtual screening (VS) [15]. In contrast, the pharmacological activity of a specific ligand such as a new natural compound can be searched for by performing docking simulations with a set of targets. This approach is called inverse virtual screening (IVS) [16, 17]. IVS, for instance, helped identify dorzolamide as an anhydrase carbonic inhibitor that could be used in cases of glaucoma [9].

Molecular docking programs were first used in the early 1990s. Back then, there were high expectations that this approach would support the development of new drugs. However, after a few years of use, the community noticed that the ranking functions used in these programs did not accurately predict the free energy of binding. In the last decade, these tools have made use of advances in technology and changes in docking methodologies [18]. These improvements in technology, including better processors and software, have permitted the implementation of IVS with a staggered docking methodology and a set of molecular targets. This approach is called virtual high-throughput screening (vHTS), and it simulates HTS experiments but is faster and more affordable [1, 7]. vHTS motivated the development of the Octopus software described in the present paper.

Octopus is an in-house automated workflow management tool that performs vHTS. It integrates MOPAC2016 [19], MGLTools [13], PyMOL [20], and AutoDock Vina [21] in order to perform molecular docking through a user-friendly interface. Unlike other platforms, such as Raccoon2 [22] and PyRx (<http://pyrx.scripps.edu>), Octopus can simulate the molecular docking of an unlimited number of ligands against an unlimited number of molecular targets. Further, neither Raccoon2 nor PyRx permit the refinement of ligands using MOPAC2016. In addition, Octopus includes a databank of 42 molecular targets (called the Our Own Molecular Targets Data Bank, OOMT [23]) against malaria, dengue, and cancer. These targets have been parameterized in the Protein Data Bank Partial Charge (Q) & Atom Type (T) (PDBQT) format [23]. OOMT was validated based on the root-mean-square deviation (RMSD) and the area under the ROC curve (AUC) using two different docking methodologies: AutoDock Vina and DOCK 6 [24].

Search algorithms in virtual screening software

Most of the software used for molecular docking can be categorized based on the analysis of ligand flexibility and the search process strategy used (including systematic searches or random searches or those based on simulation [12]).

In a systematic search or incremental algorithm, a set of values is determined for each degree of freedom. The goal is to

apply a combinatorial method for all molecular degrees of freedom through incremental ligand construction at the receptor site. Thus, the algorithm probes for different conformations of the same molecule [25]. In incremental algorithms, the ligand is fragmented and one of its fragments is positioned at the binding site of the molecular target during docking. The fragments are successively added until the molecule is completely rebuilt. Conformational ensembles comprise tools that use a molecular motion database which stores a set of conformations for each molecule and submits it to the docking process. During docking, each conformation is considered static. These approaches are effective at exploring the conformational space, but they can converge to a local minimum rather than the global minimum [21].

In contrast, deterministic algorithms do not rely on random data. Thus, the result is predetermined by the input data. The simulation methods of molecular dynamics and energy minimization are examples of the deterministic search algorithms used in docking. These methods have a high computational cost. In molecular dynamics simulations, atoms and molecules interact for a predetermined time, and we observe if they continue to interact after a particular time has elapsed or if the interaction is lost. Energy minimization algorithms apply an energy minimization strategy to an initial conformation of a molecule to find its minimum-energy conformation during the docking process.

Other random strategies used in molecular docking include generic algorithms and Monte Carlo methods. Genetic algorithms are implemented as a computer simulation where a population of abstract representations is mutated to search for better solutions. Each individual represents a possible solution to the problem. For each new generation, the adaptation of each solution is evaluated. Thus, some individuals are selected for the next generation, and they are recombined or mutated to generate new individuals. This process is repeated to find better solutions until it is finalized. A Monte Carlo method uses a statistical methodology based on a large set of random samples to get results that approximate reality [26]. Thus, Monte Carlo methods perform a sufficiently high number of successive simulations to allow probabilities to be calculated heuristically. When used as docking methods, Monte Carlo methods randomly generate an initial conformation of the ligand and calculate its binding energy. Based on this initial conformation, a new configuration is generated. If the binding energy for the new configuration is less (i.e., more negative) than that for the initial conformation, then it is automatically accepted as the reference for the next iteration. Otherwise, another evaluation is performed to check whether it should be used as the reference. This process is repeated until the desired number of iterations is reached.

In general, however, most virtual screening software packages utilize a combination of these approaches. Table 1 summarizes the strategies used for selected docking tools.

Table 1 Search algorithms used in docking software (adapted from [12] and [25])

Search algorithm	Method	Software
Random	Genetic algorithm	AutoDock Vina [21] GOLD [27] DARWIN [28] rDOCK [29]
	Monte Carlo	Autodock Vina [21] MOE-Dock [30] ICM [31] MC DOCK [32] DOCK VISION AFFINITY QXP [33] GLIDE [34] PharmDOCK [35] Biovia Discovery Studio [36]
Systematic	Deterministic	FRED [37]
	Conformational ensembles	Flexibase/FLOG [38] FRED [39] SLIDE [40] eHITS [41] Surflex-Dock [42] DOCK [43] GLIDE [34] EUDOC [44] FlexX [45]
	Incremental construction	DOCK [46] FlexX [47] Glide [34] Surflex-Dock [42] Hammerhead [48] HOOK [49] SYBYL-X [50] eHITS [41] EUDOC [44]
Simulation	Molecular dynamics	DynaDock [51]
	Energy minimization	DOCK [46] AutoDock Vina [21]

Scoring functions in virtual screening software

Docking software packages use scoring functions to estimate the strength of noncovalent interactions between a ligand and a molecular target via mathematical methods [52]. Scoring functions are one of the most important elements of structure-based drug design. However, despite their widespread use, estimating the strength of interaction between a ligand and a molecular target remains a major challenge in docking methods.

There are three basic important applications of scoring functions in molecular docking. The first is the determination of the binding site and the binding conformation for a molecular target and a ligand. Another is the prediction of the binding affinity between a protein and a ligand. Finally, they can

also be used to identify potential drugs for a given protein from large databases.

There are three types of scoring function [12, 25, 52, 53]: force-field, empirical, and knowledge-based. Force-field (FF) scoring functions are calculated based on the intermolecular interactions between the atoms of the ligand and those of the molecular target, such as van der Waals, electrostatic, and bond stretching/bending/torsional forces. FF scoring functions are usually based on experimental data and follow the principles of quantum physics [12]. However, these methods do not consider the solvent in their calculations. They also lack a physical model that describes entropic contributions, which leads to imprecision in the results generated by the scoring function.

Empirical scoring functions estimate the binding free energy based on weighted structural parameters obtained after adjusting the scoring functions based on the experimentally determined binding constants for a set of complexes [53]. This creates a training dataset of some protein–ligand complexes with known affinities [12]. Thus, linear regression is performed to predict the values of some variables [52]. Constants known as weights are then generated using the empirical function to use as coefficients to adjust the terms of the equation. Each term of the function describes a type of physical event involved in the formation of the ligand–receptor complex. Thus, hydrogen bonding as well as ionic, nonpolar, desolvation, and entropic effects are all considered.

In knowledge-based scoring functions, the binding affinity is calculated using the sum of the interactions between the ligand atoms and target atoms [53]. These functions obtain statistical data (i.e., the frequencies of specific intermolecular ligand–receptor interactions) on large databases (such as the Protein Data Bank). For example, if a hydrogen bond is present in 90% of the relevant cases, this bond is weighted more heavily in the equation of the force field. They use pairwise energy potentials extracted from known target–ligand complexes to obtain a generic scoring function, and generally assume that intermolecular interactions occur near atoms or functional groups, as such intermolecular interactions occur more frequently and are more likely to favorably contribute to the binding affinity. The final score is given as a sum of the scores of

all individual interactions. Table 2 summarizes the types of scoring functions used in various docking tools.

Octopus

Octopus is software for virtual high-throughput screening (vHTS) developed in Shell Script, Python, HTML, and CSS. It offers fast and user-friendly docking simulations. It integrates MOPAC2016 [54], PyMOL [20], MGLTools [13], and AutoDock Vina [21] via an interface that is intuitive and self-assessing (i.e., Octopus takes the output of Mopac and prepares it automatically for use as input to other programs).

In general, docking software is suitable for carrying out a simulation of one ligand docking into a specific molecular target. However, Octopus can automatically perform virtual high-throughput screening (vHTS) of N ligands docking into M molecular targets, i.e., it can perform simulations of an unlimited number of compounds docking into a set of molecular targets.

The main advantages of Octopus relative to MOPAC2016, PyMOL, MGLTools, and AutoDock Vina are its automation, ease of use, speed, and error reduction. If Octopus is not used, each of the four programs mentioned must be managed by a human operator. Also, the output of one program must be used as input for the next program, which often requires user action, introducing delays into the screening process and the possibility of user-generated errors. Therefore, there is also a need to check for human error at each step that requires user action. Also, the steps necessitating user action must be performed for each ligand–target combination. However, in Octopus, as soon as one of the programs is completed, the next is executed automatically without user intervention. Consequently, Octopus reduces the possibility of user-generated error because it reduces human interactions.

In the Octopus protocol, MOPAC2016 refines the ligands, PyMOL visualizes the ligand geometry, MGLTools determines the rotatable bonds and assigns net atomic Gasteiger–Marsili charges, and AutoDock Vina performs the molecular docking. Finally, the results are compiled and presented as binding energies for ligand–receptor combinations (Fig. 1). The protocol used by Octopus is summarized in the “Methods” section.

Table 2 Scoring functions used in docking software (adapted from [12, 25, 52, 53])

Scoring function	Software
Force-field	GOLD [27], DARWIN [28], MOE-Dock [30], MC DOCK [32], QXP [33], GLIDE [34], FRED [37], DOCK [43], EUDOC [44]
Empirical	AutoDock Vina [21], rDOCK [29], ICM [31], GLIDE [34], PharmDOCK [35], Biovia Discovery Studio [36], FRED [37], SLIDE [40], eHITS [41], FlexX [45], Hammerhead [48], HOOK [49], Surflex-Dock [42]
Knowledge-based	AutoDock Vina [21], GOLD [27]

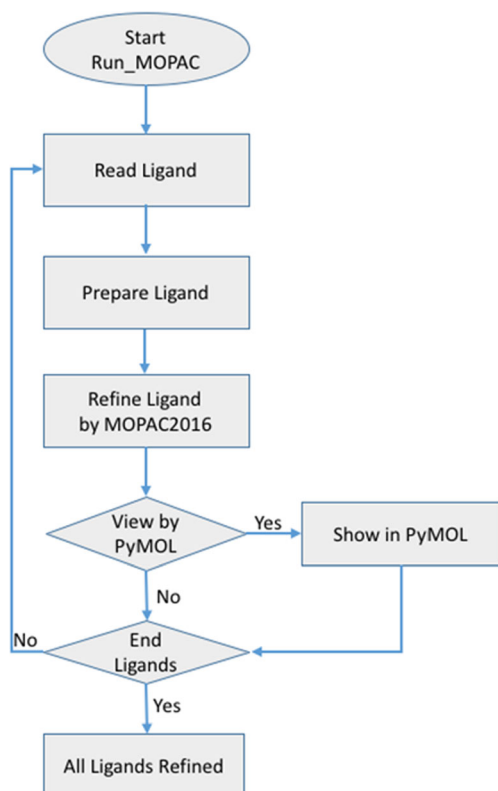


Fig. 1 Octopus vHTS results in html format. The *yellow row* shows reference values obtained from the redocking of crystallographic ligands. All values are in kcal/mol

Methods

In this section, we describe the steps performed in the Octopus protocol:

1. First, directories of ligands and targets are chosen (all the ligands and all the targets must be placed in separate directories). For instance, the ligand directory could be derived from the ZINC platform, as shown in Fig. 2.
2. When choosing the target directory (Fig. 3), a previously parameterized molecular target databank called Our Own Molecular Targets (OOMT) [23] that is included in Octopus can be utilized. The OOMT databank comprises various receptors from the Protein Data Bank (PDB), and it includes specific targets for cancer, dengue, and malaria. The main objective of the OOMT databank is to facilitate virtual screening studies using molecular docking at specific molecular targets. Appropriate biological assays can then be performed based on the results of the molecular docking. The OOMT databank has a configuration file with *X*, *Y*, and *Z* coordinates, and a grid box size delimiting the region for molecular docking simulations and the reference binding energy according to the crystallographic ligand.

3. As mentioned before, the 3D structures of ligands can be obtained from ZINC [11]. If the ligands come from a known public database, then we can proceed to step 4. Otherwise, if the ligand has been generated using the MarvinSketch program [56] or another application, then Octopus will carry out out ligand refinement using the run_MOPAC software developed in Python. This software reads the net atomic charges of the atoms of all the ligands in PDB format into the ligand folder. Next, all of the ligands are refined by the semi-empirical parametric method 7 (PM7) [55] implemented in MOPAC2016 using a routine minimum search (EF) [19]. The user is asked to check how many alpha and beta electrons are present in each molecular orbital after energy minimization of the ligands. This reduces the likelihood of accepting incorrect structures (i.e., free radicals) for subsequent calculations. This process is can be applied for ZINC databank structures converting from smile format (only) to pdb format through babel software using the keyword gen3d. The automated workflow of run_MOPAC is shown in Fig. 4.
4. In this step, ligands are converted from PDB to PDBQT format while assigning the rotatable bonds, the Gasteiger–Marsili net atomic charges [56], and only the hydrogens on polar atoms (oxygen and nitrogen) are retained; the other hydrogens are removed [13].
5. Visual inspection of the geometries of the ligands is then performed using PyMOL [20].
6. In this step, the ligands in PDBQT format are submitted to molecular docking by AutoDock Vina [21], which executes until all of the ligands have been docked into the targets. Configuration files follow the AutoDock Vina protocol, with exhaustiveness set to 24 [57].
7. Finally, the binding energy results for each molecular target are generated in CSV or HTML format. This makes it simple for the user to determine whether the ligand is capable of interacting with a specific molecular target. Figure 1 shows an example of the results obtained by Octopus in HTML. First, complementary information about the experiments (number of ligands, number of targets, date and time of experiment) is shown. The default crystallographic values for the binding energies between the ligands and targets are also displayed.

In addition, the entire process can be repeated while storing the previous results. A summary of the Octopus algorithm is presented as a six-step workflow in Fig. 5.

The interface of Octopus

Octopus has a user-friendly interface. Figure 6 shows the start interface of Octopus. Five selection options are available: (1) inverse virtual screening without run_MOPAC; (2) inverse

IVS - Inverse Virtual Screening

Octopus Result

NUMBER OF LIGANDS: 8

NUMBER OF TARGETS: 10

DATE: 2016/07/13

HOUR: 09:08:34

TARGETS	1AGW	1DDX	1GKC	1GMY	1LD8	1LF3	1LRH	1QIB	1QJA	1R6A
Lig_Crist	-9.6	-5.6	-8.3	-7.7	-8.0	-9.3	-8.1	-10.7	-9.1	-9.3
zinc1.pdbqt	-8.3	-9.2	-8.7	-7.4	-7.4	-7.4	-5.2	-9.0	-6.9	-8.0
zinc2.pdbqt	-5.4	-7.1	-5.1	-5.6	-5.6	-5.0	-5.4	-5.6	-4.8	-5.5
zinc3.pdbqt	-7.3	-10.0	-8.4	-6.8	-7.6	-7.3	-5.5	-8.6	-5.8	-8.1
zinc4.pdbqt	-7.9	-8.3	-8.6	-7.9	-7.8	-7.3	-4.9	-8.5	-7.0	-8.3
zinc5.pdbqt	-6.5	-8.9	-8.1	-6.9	-7.1	-6.2	-6.9	-7.7	-5.3	-6.6
zinc6.pdbqt	-8.3	-10.6	-8.1	-7.4	-8.7	-7.7	-5.9	-7.8	-6.5	-8.0
zinc7.pdbqt	-4.9	-6.1	-6.1	-5.2	-5.1	-4.8	-5.4	-5.9	-4.4	-5.2
zinc8.pdbqt	-8.4	-10.1	-8.1	-6.9	-8.3	-7.6	-6.1	-8.6	-6.5	-8.2

LIST LIGANDS ERROR: 0

Fig. 2 Library of ligands obtained from the ZINC platform

virtual screening with run_MOPAC; (3) run_MOPAC; (4) tutorials; and (5) install software.

Inverse virtual screening without run_MOPAC must be used when the PDB file is downloaded from a public databank. Steps 1, 3, 4, 5, and 6 of Octopus are performed in this protocol (Fig. 5). *Inverse virtual screening with*

run_MOPAC must be used when the PDB file is generated with the MarvinSketch program. In this case, all six steps of Octopus presented in Figure 5 are performed (*run_MOPAC* refines a set of ligands when they are generated by the user using a tool such as MarvinSketch). *Tutorials* on manual installation and the use of all applications associated with

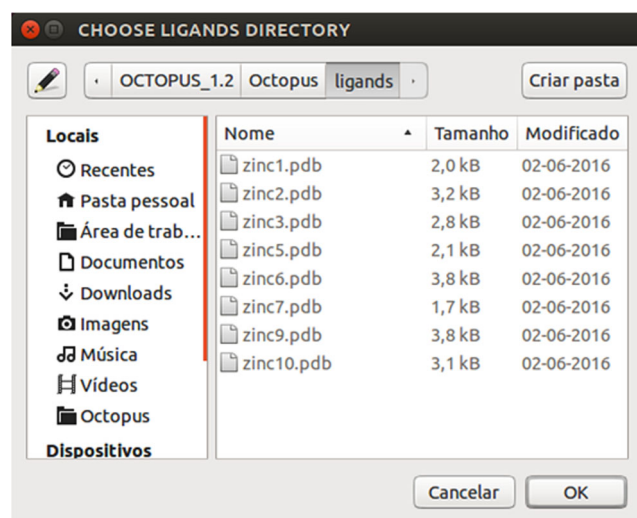


Fig. 3 Select molecular targets from the OOMT

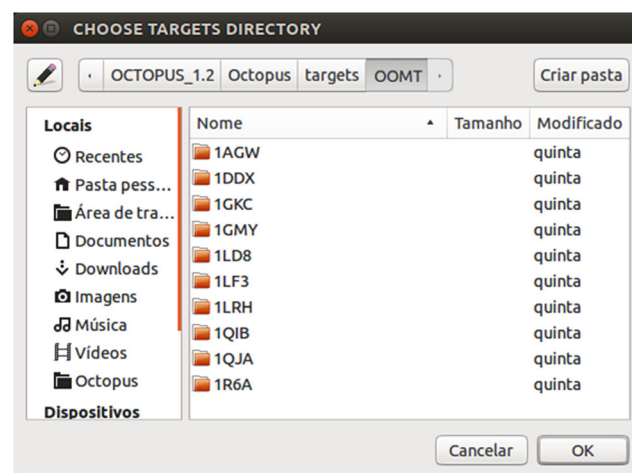


Fig. 4 The automated workflow of run_MOPAC

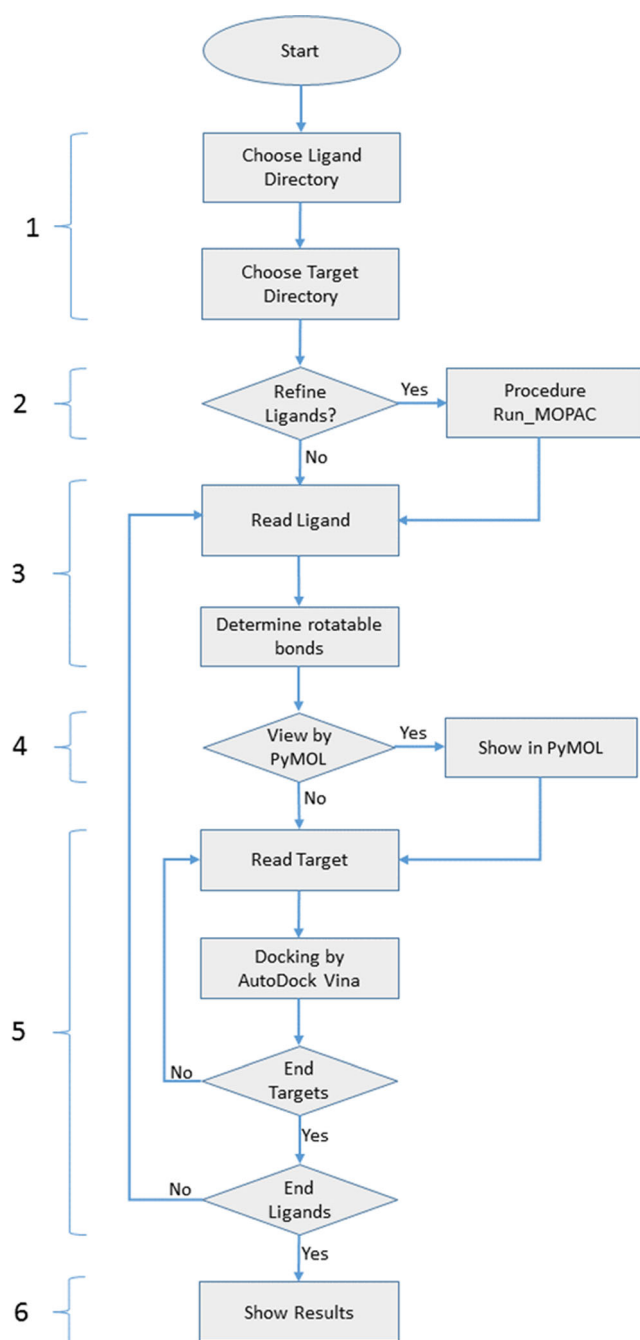


Fig. 5 Automated workflow of Octopus

Octopus are available. *Install software* is used to install other applications available in Octopus.

Octopus can perform IVS in automatic or manual mode. In automatic mode, the entire experiment is performed without user intervention after choosing the ligand and molecular target directories. PyMOL is not executed in this case. In manual mode, user intervention is required after every step shown in Figure 5. In addition, the entire process can be repeated while storing the previous results. To test out this Octopus process, we used it to perform two case studies examining the

metalloprotease activities of (phenylamino)urea derivatives and the antimalarial activity of a pyrazole derivative [58] (see the next section).

The docking approach is limited by the flexibility of the receptor, which is generally considered to be rigid, and the fixed bond angles and lengths generally assumed for the ligands. Consequently, improper results can be obtained for molecular targets when using the induced-fit mechanism. This issue can be resolved by using an ensemble of protein structures or flexible docking [22]. These tools are complementary to docking methods as they reduce computational costs. Even though Octopus uses rigid receptors from the OOMT databank, all molecular targets are evaluated based on RMSD and AUC values to gauge the accuracy that can be achieved. In addition, explicit water molecules (which participate in two hydrogen bonds) were retained in the docking simulation, whereas water molecules in the molecular targets were removed [24]. Docking using receptors with flexible side chains will be considered in subsequent versions of the program.

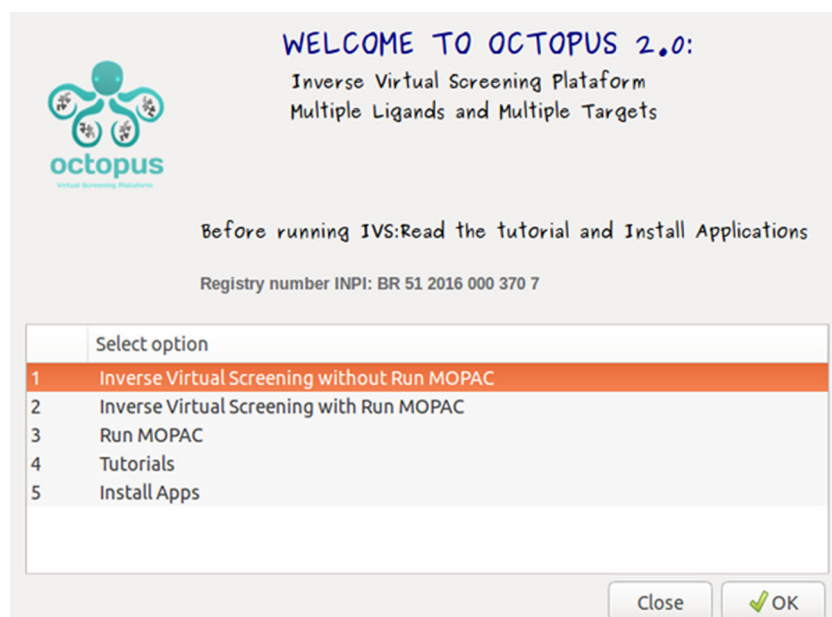
Results

This section discusses two successful applications of Octopus. In the first case study, the IVS process was applied to determine the metalloprotease activities of (phenylamino)urea derivatives. In the second (which has been reported previously), the process was applied to check whether a particular pyrazole derivative possesses antimalarial activity.

Successful case study 1: metalloprotease activities of (phenylamino)urea derivatives

A set of 22 (phenylamino)urea derivatives (“LSO&ME” compounds) were submitted to Octopus. Docking results from the IVS approach suggested that, among the 40 molecular targets studied, the metalloproteinases were feasible targets. The matrix metalloproteinases (MMPs) are zinc-dependent enzymes that have collagen (present in the extracellular matrix) as one of their substrates. They participate in the tissue remodeling process. Moreover, they are involved in tumor metastasis because they are overexpressed in some types of tumors. The IVS methodology showed that binding energies with the metalloproteinase with PDB code 1GKC ranged from -8.0 kcal/mol to -9.5 kcal/mol [59]. The corresponding crystallographic binding energy was -6.6 kcal/mol. 1GKC recognized LSO&ME007, with interactions including hydrogen bonds, van der Waals interactions, and intramolecular π -stacking. This molecular target is a metalloprotease involved in cancer pathology; it was evaluated previously based on the RMSD and the ROC curve [24], yielding values of 0.55 Å and 0.60 , respectively. RMSD values of <2.0 Å imply good pose fidelity

Fig. 6 User-friendly interface of Octopus: the main menu



[21], while AUC values of >0.5 enable the methodology to distinguish between true- and false-positive compounds. In other words, docking studies of this system should be evaluated by performing a corresponding experimental study.

Figure 7 summarizes the intermolecular interactions between 1GKC and two ligands in the form of 2D diagrams. L-Valinamide (Fig. 7a) and LE&007 (Fig. 7b) present similar molecular interactions in terms of van der Waals and hydrogen bonds, although LE&007 shows additional intermolecular interactions, such as π - π stacking and T-shaped stacking. In addition, the interaction (at a distance of 2.39 Å) between

the zinc atom of 1GKC and the lone pairs of the carbonyl moiety of LE&007 is highlighted in the figure. These additional molecular interactions with LE&007 help to explain the binding energies of the (phenylamino)urea derivatives with the metalloproteinases. The compounds of interest were studied in a biological assay, and LE&007 was found to inhibit 80% of the enzymatic activity of the metalloproteinase 1GKC.

Following the IVS experiments, the effects of the (phenylamino)urea derivatives (LSO&ME compounds) on the proteolytic activities of MMP gelatinases were measured by gelatin zymography performed according to a previous

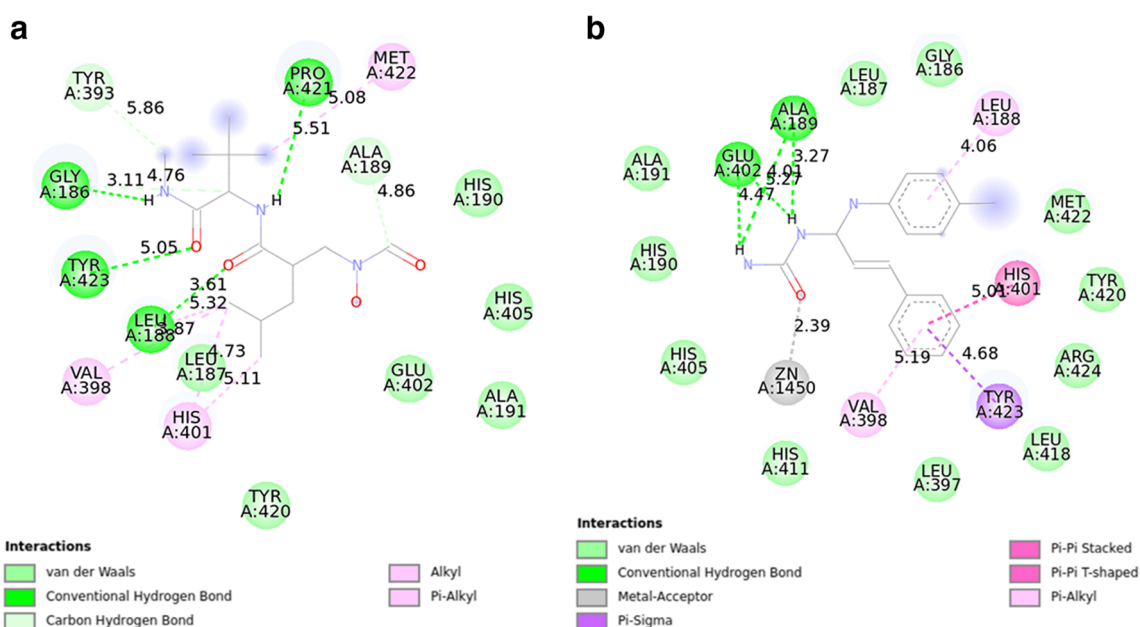


Fig. 7a–b 2D diagrams of the binding between 1GKC and two ligands, as visualized using Discovery Studio Visualizer 4.5 [60]: **a** the crystallographic ligand L-valinamide; **b** LE&007

report [61]. The samples were dissolved in dimethyl sulfoxide (DMSO) at 6 mg/mL, and 10 μ L were applied to a well of gel containing the substrate-rich MMPs: saliva (20 U of protein) in sample buffer (SDS 2.5 wt% and saccharose 1 wt%). This corresponded to the same quantity of saliva was used as the standard, and this represented 100% of the active enzymes. Electrophoresis (PROTEAN II, Bio-Rad, Hercules, CA, USA) was conducted under reducing conditions (0.025 M Tris, 0.192 M glycine, and 0.1% SDS, pH 8.5) at 70 V and 4 °C for 3.5 h.

After electrophoresis, the gels were washed for 1 h with Triton X-100 (2.5 g%) to remove the SDS, and then submerged (with stirring) in an activation buffer (Tris–HCl 0.05 M, CaCl₂ 0.6 g%, pH 8.0) for 16 h at room temperature. Next, the gels were stained (0.25% Coomassie Blue R-250, methanol 45%, and acetic acid 10%) for 1 h and then bleached (using 30% ethanol/10% acetic acid) for another hour.

The compounds LSO&ME005, LSO&ME004, and LSO&ME007 suppressed the activity of MMP-9 by approximately 80%, and partial inhibition was observed when LSO&ME004 was applied. LSO&ME005 and LSO&ME028 suppressed the activity of MMP-2 by approximately 55%. Inhibition of gelatinase activity was measured by comparing the decrease in the amount of undigested bound substrate in solutions containing MMPs and the LSO&ME compounds with the decrease in the amount of undigested bound substrate observed in solutions of MMPs that did not contain the LSO&ME compounds.

Successful case study 2: antimalarial activity of a pyrazole derivative

Several reports have shown that pyrazole derivatives possess biological activities (e.g., [62]). Hence, our group performed VS of the pyrazole derivative Tx001. Octopus showed that this compound can complex with a model of *Plasmodium falciparum* ATP6 (PfATP6) [63], with a binding energy of -8.6 kcal/mol (as compared to -7.7 kcal/mol for the binding energy of thapsigargin (TG)—a natural compound that is an inhibitor of PfATP6) calculated for docking into the hydrophobic cavity of this model. The complex Tx001–PfATP6 was then evaluated by molecular simulation utilizing an implicit solvent model, and the system was observed to reach equilibrium in 30 ns. The potential energy of the system decreased during the simulation to approximately -5500 kcal/mol. The main ligand–PfATP6 interactions were van der Waals, electrostatic, and hydrogen bonding between the guanidinium moiety of Tx001 and Ile752 of PfATP6. Finally, Tx001 was evaluated for antimalarial activity, and it presented a good inhibitory concentration (IC₅₀) of 8.2 μ M. Its antimalarial activity is therefore stronger than that of chloroquine (IC₅₀ = 0.38 μ M), a widely used antimalarial drug, which

motivated us to optimize this ligand. Second-generation derivatives of Tx001 are currently being evaluated [58].

Conclusions

Drug development is a difficult task for small academic groups. Thus, applying a theoretical approach can increase the “hit” rate, and these hits have the potential to become lead compounds for new therapies. This motivated us to develop Octopus as a tool for the vHTS of multiple compounds against a set of molecular targets. It can also reduce the number of biological assays needed to determine a pharmacological mechanism. It is limited principally by the time to draw the structures of the ligands as well as the choice of desired targets. The entire Octopus protocol can run automatically, although computational chemists are still needed to visually inspect the intermolecular interactions.

In this manuscript, we also showed two successful examples of the application of Octopus to find molecular targets. Octopus identified a new hit compound, LE&007, that can be optimized to generate a new lead compound for antineoplastic drugs, and it was also used to determine the antimalarial activity of the pyrazole derivative Tx001. Neither LE&007 nor Tx001 were lead candidates originally identified for these diseases. Thus, Octopus provides a second chance to find a use for these compounds as lead compounds.

Finally, Octopus provides a user-friendly Linux-based interface for MOPAC2012, PyMOL, and AutoDock Vina. Work to enhance Octopus by adding a new molecular dynamics simulation code is also in progress. Octopus can be obtained from www.drugdiscovery.com.br

Acknowledgements The authors are grateful for the support provided by the Foundation for Research Support of Minas Gerais (FAPEMIG APQ-00557-14 and APQ-02860-16), the Higher Level Personnel Improvement Commission (CAPES), the National Research Council (CNPq UNIVERSAL 449984/2014-1), and Graduated Programs in Pharmaceutical Sciences (PPGCS) and Biotechnology (PPGBiotec) from the Federal University of Sao Joao del Rei (UFSJ) and the Federal Center for Technological Education of Minas Gerais (CEFET-MG). A.G. Taranto is grateful to Mr. Pedro for the “Ignorância Zero” initiative.

References

1. Ferreira RS, Oliva G, Andricopulo AD (2011) Integrating virtual and high-throughput screening: opportunities and challenges in drug research and development. *Quim Nov.* 34:1770–1778. doi:10.1590/S0100-40422011001000010
2. Bannani YL (2011) Drug discovery in the next decade: Innovation needed ASAP. *Drug Discov Today* 16:779–792. doi:10.1016/j.drudis.2011.06.004
3. Keserü GM, Makara GM (2009) The influence of lead discovery strategies on the properties of drug candidates. *Nat Rev Drug Discov* 8:203–212. doi:10.1038/nrd2796

4. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3:935–949. doi:10.1038/nrd1549
5. Katsuno K, Burrows JN, Duncan K et al (2015) Hit and lead criteria in drug discovery for infectious diseases of the developing world. *Nat Rev Drug Discov* 14:751–8. doi:10.1038/nrd4683
6. Andricopulo A, Ferreira L (2014) Medicinal chemistry approaches to neglected diseases drug discovery. *J Mod Med Chem* 2:20–31. doi:10.12970/2308-8044.2014.02.01.4
7. Polgar T, Keseru GM (2011) Integration of virtual and high throughput screening in lead discovery settings. *Comb Chem High Throughput Screen* 14:889–897. doi:10.2174/138620711797537148
8. Ripphausen P, Nisius B, Bajorath J (2011) State-of-the-art in ligand-based virtual screening. *Drug Discov Today* 16:372–376. doi:10.1016/j.drudis.2011.02.011
9. Sousa SF, Cerqueira NMFS, Fernandes PA, Ramos MJ (2010) Virtual screening in drug design and development. *Comb Chem High Throughput Screen* 13:442–453. doi:10.2174/138620710791293001
10. Berman HM, Kleywegt GJ, Nakamura H, Markley JL (2013) The future of the Protein Data Bank. *Biopolymers* 99:218–22. doi:10.1002/bip.22132
11. Irwin JJ, Shoichet BK (2005) ZINC—a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45:177–182. doi:10.1021/ci049714+
12. Ferreira L, dos Santos R, Oliva G, Andricopulo A (2015) Molecular docking and structure-based drug design strategies. *Molecules* 20:13384–13421. doi:10.3390/molecules200713384
13. Morris G, Huey R (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791. doi:10.1002/jcc.21256.AutoDock4
14. Elokely KM, Doerksen RJ (2013) Docking challenge: protein sampling and molecular docking performance. *J Chem Inf Model* 53:1934–1945
15. Jaghoori MM, Bleijlevens B, Olabarriaga SD (2016) 1001 ways to run AutoDock Vina for virtual screening. *J Comput Aided Mol Des* 30:237–249. doi:10.1007/s10822-016-9900-9
16. Hui-fang L, Qing S, Jian Z, Wei F (2010) Evaluation of various inverse docking schemes in multiple targets identification. *J Mol Graph Model* 29:326–330. doi:10.1016/j.jmgm.2010.09.004
17. Carregal AP, Comar M, Alves SN et al (2012) Inverse virtual screening studies of selected natural compounds from Cerrado. *Int J Quantum Chem* 112:3333–3340. doi:10.1002/qua.24205
18. Rognan D (2010) Structure-based approaches to target fishing and ligand profiling. *Mol Inform* 29:176–187. doi:10.1002/minf.200900081
19. Stewart JPP (2012) MOPAC2012. Stewart Computational Chemistry, Colorado Springs
20. DeLano WL (2002) The PyMOL molecular graphics system, version 1.8. Schrödinger, LLC, New York. <http://www.pymol.org>. doi:10.1038/hr.2014.17
21. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. doi:10.1002/jcc.21334
22. Forli S, Piche ME, Sanner M et al (2016) Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 5:905–919. doi:10.1038/nbt.3121.ChIP-nexus
23. Carregal AP, Comar Jr MC, Taranto AG (2013) Our Own Molecular Targets data bank (OOMT). *Biochem Biotechnol Reports* 2:14–16
24. Carregal AP, Maciel FV, Carregal JB, et al. (2016) Docking-based virtual screening of Brazilian natural compounds using OOMT as the pharmacological target database. *J Mol Model* (submitted)
25. Haga JH, Ichikawa K, Date S (2016) Virtual screening techniques and current computational infrastructures. *Curr Pharm Des* 22:3576–84
26. Harrison RL (2010) Introduction to Monte Carlo simulation. *AIP Conf Proc* 1204:17–21. doi:10.1063/1.3295638.Introduction
27. Verdonk ML, Cole JC, Hartshorn MJ et al (2003) Improved protein–ligand docking using GOLD. *Proteins Struct Funct Genet* 52:609–623. doi:10.1002/prot.10465
28. Taylor JS, Burnett RM (2000) DARWIN: a program for docking flexible molecules. *Proteins Struct Funct Genet* 41:173–191. doi:10.1002/1097-0134(20001101)41:2<173::AID-PROT30>3.0.CO;2-3
29. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N et al (2014) rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol* 10:1–7. doi:10.1371/journal.pcbi.1003571
30. Chemical Computing Group Inc. (2004) Molecular Operating Environment (MOE). *Sci Comput Instrum* 22:32
31. Abagyan R, Totrov M, Kuznetsov D (1994) ICM—a new method for protein modeling and design: applications to docking and structure prediction from distorted native conformation. *J Comput Chem* 15:488–506
32. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein–small molecule docking methods. *J Comput Aided Mol Des* 16:151–166. doi:10.1023/A:1020155510718
33. McMartin C, Bohacek RS (1997) QXP: powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des* 11:333–344. doi:10.1023/a:1007907728892
34. Friesner RA, Banks JL, Murphy RB et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749. doi:10.1021/jm0306430
35. Hu B, Lill MA (2014) PharmDock: a pharmacophore-based docking program. *J Cheminform* 6:1–14. doi:10.1186/1758-2946-6-14
36. Accelrys Software Inc. (2013) Discovery Studio Modeling Environment, release 4.1. Accelrys Software Inc., San Diego
37. McGann M (2011) FRED pose prediction and virtual screening accuracy. *J Chem Inf Model* 51:578–596. doi:10.1021/ci100436p
38. Kearsley SK, Underwood DJ, Sheridan RP, Miller MD (1994) Flexibase: a way to enhance the use of molecular docking methods. *J Comput Aided Mol Des* 8:565–582. doi:10.1007/BF00123666
39. McGann M (2012) FRED and HYBRID docking performance on standardized datasets. *J Comput Aided Mol Des* 26:897–906. doi:10.1007/s10822-012-9584-8
40. Schnecke V, Kuhn LA (2000) Virtual screening with solvation and ligand-induced complementarity. *Perspect Drug Discov Des* 20:171–190. doi:10.1023/A:1008737207775
41. Zsoldos Z, Reid D, Simon A et al (2007) eHiTS: a new fast, exhaustive flexible ligand docking system. *J Mol Graph Model* 26:198–212. doi:10.1016/j.jmgm.2006.06.002
42. Spitzer R, Jain AN (2012) Surflex-Dock: docking benchmarks and real-world application. *J Comput Aided Mol Des* 26:687–699. doi:10.1007/s10822-011-9533-y
43. Lang PT, Brozell SR, Mukherjee S et al (2009) DOCK 6: combining techniques to model RNA—small molecule complexes. *RNA* 15:1219–1230. doi:10.1261/ma.1563609.HIV
44. Pang YP, Perola E, Xu R, Prendergast FG (2001) EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J Comput Chem* 22:1750–1771. doi:10.1002/jcc.1129
45. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 261:470–89. doi:10.1006/jmbi.1996.0477

46. Allen WJ, Balias TE, Mukherjee S et al (2015) DOCK 6: impact of new features and current docking performance. *J Comput Chem* 36: 1132–1156. doi:10.1002/jcc.23905
47. Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FlexX incremental construction algorithm for protein–ligand docking. *Proteins Struct Funct Genet* 37:228–241. doi:10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8
48. Welch W, Ruppert J, Jain AN (1996) Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol* 3:449–462. doi:10.1016/S1074-5521(96)90093-9
49. Eisen MB, Wiley DC, Karplus M, Hubbard RE (1994) HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins Struct Funct Genet* 19:199–221. doi:10.1002/prot.340190305
50. Tripos International (2011) SYBYL-X 1.2. Tripos International, St. Louis
51. Antes I (2010) DynaDock: a new molecular dynamics-based algorithm for protein–peptide docking including receptor flexibility. *Proteins Struct Funct Bioinf* 78:1084–1104. doi:10.1002/prot.22629
52. Huang S-Y, Grinter SZ, Zou X (2010) Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys Chem Chem Phys* 12:12899–908. doi:10.1039/c0cp00151a
53. Breda A, Basso LA, Santos DS, de Azevedo Jr WF (2008) Virtual screening of drugs: score functions, docking, and drug design. *Curr Comput Aid Drug Des* 4:265–272. doi:10.2174/157340908786786047
54. Stewart JJP (2016) MOPAC2016. Stewart Computational Chemistry, Colorado Springs. <http://openmopac.net/MOPAC2016.html>
55. Dutra JDL, Filho MAM, Rocha GB et al (2013) Sparkle/PM7 lanthanide parameters for the modeling of complexes and materials. *J Chem Theory Comput* 9:3333–3341. doi:10.1021/ct301012h
56. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* 36:3219–3228. doi:10.1016/0040-4020(80)80168-2
57. Forli S, Huey R, Pique ME et al (2016) Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc* 11:905–919. doi:10.1038/nprot.2016.051
58. Nunes RR, dos Costa MS, dos Santos BR et al (2016) Successful application of a virtual screening and molecular dynamics simulation against antimalarial molecular targets. *Mem Inst Oswaldo Cruz* 111:721–730. doi:10.1590/0074-02760160207
59. Rowsell S, Hawtin P, Minshull CA et al (2002) Crystal structure of human MMP9 in complex with a reverse hydroxamate inhibitor. *J Mol Biol* 319:173–181. doi:10.1016/S0022-2836(02)00262-0
60. Accelrys Software Inc. (2015) Discovery Studio modeling environment, release 4.5. Accelrys Software Inc., San Diego
61. RIMA R, Kuribayashi JS, Borges Júnior PC et al (2010) Inibição de metaloproteinases por extratos aquosos de Aloe vera, Annona muricata e chá preto. *Biosci J* 26:121–127
62. Adhikari A, Kalluraya B, Sujith KV et al (2012) Synthesis, characterization and pharmacological study of 4,5-dihydropyrazolines carrying pyrimidine moiety. *Eur J Med Chem* 55:467–474. doi:10.1016/j.ejmech.2012.07.002
63. Guimarães DSM, Da Fonseca AL, Batista R et al (2015) Structure-based drug design studies of the interactions of *ent*-kaurane diterpenes derived from *Wedelia paludosa* with the *Plasmodium falciparum* sarco/endoplasmic reticulum Ca²⁺-ATPase PfATP6. *Mem Inst Oswaldo Cruz* 110:255–258. doi:10.1590/0074-02760140415