J Mol Model (2017) 23:111
DOI 10.1007/s00894-017-3253-8

ORIGINAL PAPER

Docking-based virtual screening of Brazilian natural compounds using the OOMT as the pharmacological target database

Ana Paula Carregal[1] · Flávia V. Maciel[1] · Juliano B. Carregal[1] · Bianca dos Reis Santos[2] · Alisson Marques da Silva[2] · Alex G. Taranto[1]

Received: 27 April 2016 / Accepted: 23 January 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract The demand for new therapies has encouraged the development of faster and cheaper methods of drug design. Considering the number of potential biological targets for new drugs, the docking-based virtual screening (DBVS) approach has occupied a prominent role among modern strategies for identifying new bioactive substances. Some tools have been developed to validate docking methodologies and identify false positives, such as the receiver operating characteristic (ROC) curve. In this context, a database with 31 molecular targets called the Our Own Molecular Targets Data Bank (OOMT) was validated using the root-mean-square deviation (RMSD) and the area under the ROC curve (AUC) with two different docking methodologies: AutoDock Vina and DOCK 6. Sixteen molecular targets showed AUC values of >0.8, and those targets were selected for molecular docking studies. The drug-likeness properties were then determined for 473 Brazilian natural compounds that were obtained from the ZINC database. Ninety-six compounds showed similar drug-likeness property values to the marked drugs (positive values). These compounds were submitted to DBVS for 16 molecular targets. Our results showed that AutoDock Vina was more appropriate than DOCK 6 for performing DBVS experiments. Furthermore, this work suggests that three compounds—ZINC13513540, ZINC06041137, and ZINC1342926—are inhibitors of the three molecular targets 1AGW, 2ZOQ, and 3EYG, respectively, which are associated with cancer. Finally, since ZINC and the PDB were solely created to store biomolecule structures, their utilization requires the application of filters to improve the first steps of the drug development process.

Keywords Structure-based drug design · Docking · Natural products · Virtual screening

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

Electronic supplementary material The online version of this article (doi:10.1007/s00894-017-3253-8) contains supplementary material, which is available to authorized users.

✉ Ana Paula Carregal
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, Belo Horizonte, MG, Brazil

Introduction

Ensuring the availability of new drugs on the market is a big challenge for the modern pharmaceutical industry. The number of new drugs approved and released every year has remained constant, despite a steady increase in investment in research and development [1]. This situation has encouraged the development of different strategies for identifying new lead compounds ("hits") [2], as the utilization of biological assays and methodologies is limited by their high cost [3]. Moreover, further complications arise when the active compound occurs in an organism because of the inherently reduced availability of the compound and the presence of many important metabolites in low quantities, which mean that a large number of the organisms are needed to extract and isolate the compound of interest [4].

Considering the number of potential biological targets for new drugs, docking-based virtual screening (DBVS) plays a prominent role among current strategies for identifying promising bioactive substances. DBVS is a theoretical approach that enables the identification of lead compounds from the

Published online: 11 March 2017

three-dimensional structure of the receptor of interest using docking programs. These docking programs measure the affinity of a small molecule (the ligand) for a molecular target to determine the interaction energy of the resulting complex. Moreover, starting from the complex between the ligand and the receptor, visualization software can present the intermolecular interaction that is responsible for molecular recognition. As a result, DBVS can identify the most promising lead compounds for biological assays and decrease the costs associated with drug development [2, 5].

The root mean square deviation (RMSD) is a statistical measure that can be used to evaluate the ability of a molecular docking program to reproduce the ligand conformation observed in the corresponding experimental crystallographic structure [6]. The RMSD is calculated by superposing the crystallographic ligand pose on the theoretically determined ligand pose when redocked, and then calculating the deviations between the positions of the heavy atoms in the crystallographic ligand pose and their positions in the theoretical ligand pose. The RMSD can be used to deduce the following: (i) if the chosen docking methodology can accurately reproduce the binding conformation; (ii) if the configuration parameters are suitable for the docking process; and (iii) a reference score or binding energy for the crystallographic ligand (active). On the other hand, although RMSD calculations are simple and are relatively computationally inexpensive, their application to evaluate the quality of a docking program has limitations. The first relates to the number of atoms in the ligands. The RMSD is only practical to use for small molecules (with relatively few atoms), which poses a problem if we need a method that can be applied to molecules of various sizes. Furthermore, high deviations can be obtained for groups of ligands that do not participate in some common types of intermolecular interactions, even though the fundamental characteristics of the overall ligand–receptor docking interaction are maintained by other interactions [7]. However, the main limitation of DBVS is the ability of the docking methodology to discriminate true-positive from false-positive compounds. Some tools have been developed to validate the docking methodology and identify false-positive compounds, such as the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) [8]. The ROC curve is a plot of the true-positive rate against the false-positive rate, and the AUC indicates the probability of retrieving true-positive results before false-positive results [9, 10]. In other words, the closer the AUC value is to 1.0, the more likely it is that the methodology will distinguish true-positive from false-positive compounds. An AUC value of ≤ 0.5 implies a methodology that selects true-positive and false-positive compounds at random.

Inverse virtual screening (IVS) is a screening technique for molecular targets [11]. This technique can be applied to natural products, which are generally substances for which there have been no previous reports of biological activity. Moreover, many of the drugs currently available on the market were obtained from natural sources or were inspired by them [12]. IVS enables the identification of molecular targets as well as the prediction of the pharmacological properties and the potential adverse effects of a natural compound [4]. In this context, computational methodologies can be used to identify structural modifications of natural compounds that could improve ligand–receptor molecular recognition and thus lead to the ligand-optimization process. Lauro and collaborators built a bank of specific molecular targets for cancer and conducted an IVS study with natural compounds to identify molecular targets and natural compounds with favorable interactions [4]. They docked 43 natural compounds into a library of 126 protein targets, all of which are involved in tumor processes, using the AutoDock Vina program [13]. In addition to the IVS test, they developed a statistical approach to eliminate false-positive compounds. As a result, six compounds with affinities for nine receptors involved in cancer were described as hits. In a previous study, our group used IVS to select molecular targets for natural compounds obtained from cerrado, which is a typical biome in Brazil [14]. The study included docking and molecular dynamics (MD) simulation to evaluate the methodology. However, the MD simulation has a high computational cost, which limits its use to a small set of compounds. These previous studies motivated our group to develop the Our Own Molecular Targets Data Bank (OOMT) [15]. The OOMT contains various receptors present in the Protein Data Bank (PDB) [16], including those related to the physiopathology of cancer and malaria. The main objective of the OOMT is facilitate DBVS on specific molecular targets for which biological assays can also be performed.

The ZINC database is a free database with 35 million commercially available compounds. Since it was first released ten years ago, ZINC has increased in size more than tenfold. ZINC is maintained by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry of the University of California, San Francisco (UCSF). ZINC allows searches to be performed by ligand structure, biological activity, and physical properties, as well as by seller [17].

Because of the availability of various databases that provide thousands of compounds for high-throughput screening (HTS) and virtual screening (VS), a need to be able to select the most promising compounds with suitable physicochemical characteristics to become drugs has emerged. Among various filters that can be applied, drug-likeness can be employed to evaluate whether molecules in the ZINC database have similar physicochemical characteristics or similar functional groups to drugs available in the marketplace. This evaluation circumvents the need for screening tests for potentially toxic compounds or for those with physical properties that will impair bioavailability [5, 18].

In the study reported in the present paper, in silico methods were used to identify lead compounds with potential

biological activity. Initially, our goals were to determine which methodology is more suitable for DBVS experiments to evaluate the performance of AutoDock Vina and DOCK 6 using RMSD and AUC curve calculations. Then the OOMT was used to perform virtual-screening studies to infer the biological activities of natural compounds from Brazilian flora that are deposited in the ZINC database.

Methodology

Evaluation of molecular docking methodologies

For each OOMT molecular target (Table 1), five active compounds were selected from the ChEMBL [19, 20]. Then 50 decoys were generated from the most active compound for each molecular target using the DUD-E platform [21]. Next, the active and decoy compounds were submitted to molecular docking in the AutoDock Vina [13, 22] and DOCK 6 [23] programs. However, two different methodologies were used by DOCK 6: the grid score [24] and the generalized Born (AMBER GB-SA) solvation model [23, 25, 26]. The output of the grid score was used as the input for the GB-SA calculation, which performed rescoring. The details of the AutoDock Vina and DOCK 6 parameters are provided in Table S1 of the "Electronic supplementary material" (ESM). The ROC curve and the area under the ROC curve (AUC) were built for each methodology [9] using a trial version of IBM SPSS Statistics for Windows [27]. In addition, the methodologies were evaluated by redocking, which involves

Table 1 AUC and RMSD values calculated for the OOMT proteins

	AUC			RMSD	
OOMT database	AutoDock Vina	Grid score	GB-SA score	AutoDock Vina	DOCK6 grid score
1AGW	0.80	0.90	0.91	1.42	0.53
1DDX	0.59	0.26	0.32	1.58	1.04
1GKC	0.60	0.35	0.60	0.55	0.27
1GMY	0.33	0.34	0.27	1.58	0.84
1LD8	0.48	0.44	0.34	0.25	0.44
1LF3	0.73	0.71	0.73	0.41	0.18
1 W22	0.27	0.19	0.48	7.85	1.21
1W6M	0.00	0.22	0.48	0.68	0.30
1Z57	0.58	0.68	0.59	0.29	0.29
1ZZ1	0.92	0.94	0.78	0.87	1.04
2HYY	0.59	0.63	0.50	0.27	0.07
2QHN	0.49	0.36	0.24	1.22	0.55
2VV9	0.59	0.70	0.60	13.49	1.02
2 W15	0.63	0.85	0.52	1.19	0.98
2ZOQ	0.89	0.29	0.26	0.98	0.95
3BPF	0.71	0.94	0.78	7.88	1.96
3BZ3	0.65	0.63	0.16	5.97	1.18
3C4C	0.92	0.41	0.39	1.20	0.16
3DV3	0.15	0.54	0.24	1.21	0.18
3EDQ	0.57	0.92	0.73	1.45	1.90
3ENE	0.79	0.58	0.44	2.99	1.40
3EYG	0.84	0.84	0.37	1.37	0.63
3FAP	0.46	0.68	0.64	0.21	0.05
3FL5	1.00	0.63	0.43	1.61	0.53
3G0E	0.92	0.46	0.55	0.82	0.33
3HIG	0.87	0.61	0.71	0.74	0.50
3JYA	0.80	0.43	0.48	0.90	0.10
4AGN	0.08	0.23	0.25	1.02	0.18
4EY7	0.92	0.78	0.91	1.92	1.63
4IAR	0.64	0.92	0.60	0.31	0.25
4 J56	0.10	0.02	0.04	0.11	12.38

calculating the RMSD between the crystallographic ligand structure and the best docked ligand pose, using DOCK 6 [23, 28, 29], the academic version of the DS Visualizer 4.1 software [30], and AutoDock Vina.

Docking-based virtual screening (DBVS)

The Brazilian natural-compound database was obtained from the ZINC platform [17, 31] and consisted of 473 compounds. The fragment-based drug-likeness property was calculated for each compound using the OSIRIS DataWarrior 4.2.2 [32] program to select the compounds with the best pharmacokinetic properties [18]. The resultant compounds were submitted to MarvinSketch 15.5.4.0 [33] to correct the protonation states of the molecules at pH 7.4. The geometries of the compounds were optimized using Parametric Method 7 (PM7) [34] in MOPAC2012 [35]. Only the molecular targets that had an AUC ROC value of ≥0.8 were submitted to screening tests [36, 37]. The AutoDock Vina 1.0.2 [13, 22] software and the DOCK 6 [23] program were used for docking simulation. Meanwhile, the molecular targets were prepared from the OOMT (Table. 1). Initially, the H++ program [38–40] was used to adjust the ionizable groups to the appropriate protonation state at pH 7.4 by adding hydrogen atoms where appropriate, except for 3BPF and 1LF3, which were adjusted to pH 4.0 (lysosomal environment of *Plasmodium falciparum*). After that, the crystallographic water molecules that formed a bridge structure between the ligand and the protein were added using DS Visualizer 4.1. For AutoDock Vina, the polar hydrogens and Gasteiger charges were adjusted using the AutoDockTools program [41]; for DOCK 6, the hydrogens and Gasteiger charges of the proteins and ligands were added using the Chimera software [42]. The Gasteiger charges are the partial charges of the atoms in molecules, which were calculated using an electrostatic model that considers the partial equalization of orbital electronegativities (PEOE). Atomic charges are useful for predicting the physical or chemical properties of molecules [43, 44]. AutoDock Vina calculations are based on a pseudo-randomly generated algorithm, so each calculation was performed in quintuplicate and the binding energies were obtained from the average with an exhaustiveness of 20. Finally, the compounds were submitted to the DBVS methodology using the OOMT database (see Table 1) [15].

Results and discussion

Evaluation of molecular docking methodologies

Initially, the AutoDock Vina and DOCK 6 programs were selected to carry out DBVS. AutoDock Vina performed stochastic optimization to search for the binding conformation. Then the Metropolis criterion was used to decide whether to accept this optimization. In addition, AutoDock Vina uses both knowledge-based potentials and empirical scoring functions [13, 22], so AutoDock Vina combines speed and accuracy, which is ideal for DBVS experiments. In contrast, the DOCK 6 program uses a systematic search based on the incremental construction method. In this method, the ligand is divided into rigid and flexible fragments. Because the rigid cores are set, they are docked into the active site and the flexible fragments are incrementally added [23, 45]. Unlike AutoDock Vina, DOCK 6 is a force-field-based method. In our studies, the grid-based score [24] and GB-SA [23] were used. The grid-based score computes the contact and energy scores and identifies if a ligand atom sterically overlaps with a receptor atom. In addition, GB-SA requires the output of the grid-based score, which includes a solvation model in the molecular-docking calculation. This output enables the electrostatic, van der Waals, and hydrophobic contributions to the free energy of the protein–ligand complex to be estimated. The GB-SA methodology is typically used as a second-step DBVS assay because of its computational cost [23].

In other words, two docking programs were used, but the performance of each of three different methodologies was evaluated in our studies. As a result, the area under the ROC curve (AUC) and the RMSD were calculated for 31 molecular targets for each methodology, which yielded 93 ROC curves and 62 RMSD calculations in total. Table 1 shows the AUC and RMSD values for each molecular target in the OOMT for AutoDock Vina and DOCK 6 (score functions). AUC values larger than 0.5 designate that the method was able to distinguish between true-positive and false-positive compounds. Among the 31 curves constructed, the success rates were 71% (22), 55% (17), and 48% (15) for the AutoDock Vina program, grid-based score, and GB-SA, respectively. However, to select molecular targets for which the methodologies could separate a true-positive from a false-positive compound, DBVS assays were applied to molecular targets with AUC values >0.8. The complete lists of docked molecular targets and compounds are available in Tables 2 and 3 in the ESM. Although GB-SA is a robust method because it includes solvation and rescores the grid-based score, it was less efficient at distinguishing true-positive from false-positive compounds compared to other methods. In addition, it was computationally costly. Thus, docking simulations of natural compounds were not performed using the GB-SA methodology. AutoDock Vina showed the best performance in this case.

In general, the docking methodologies were evaluated by calculating the RMSD. The ligand was removed from the binding site and docked again. Overlaying the crystallographic on the theoretically docked structure enabled the RMSD to be calculated. RMSD values of <2.0 Å indicate that the methodology is suitable for docking [7]. In other words, the RMSD indicates the ability to reproduce the binding conformation of

the crystallographic ligand–protein complex. According to Table 1, our results indicate that AutoDock Vina and the grid-based score method of the DOCK 6 program were generally able to reproduce the binding conformations of crystallographic ligand–target complexes for which the molecular target is in the OOMT. Moreover, AutoDock Vina and the grid-based score were able to efficiently redock 83% (26) and 97% (30), respectively, of the ligands in cases where the RMSD value was <2.0. Thus, the grid-based score was more efficient for calculating the RMSD. In addition, Tables 2 and 3

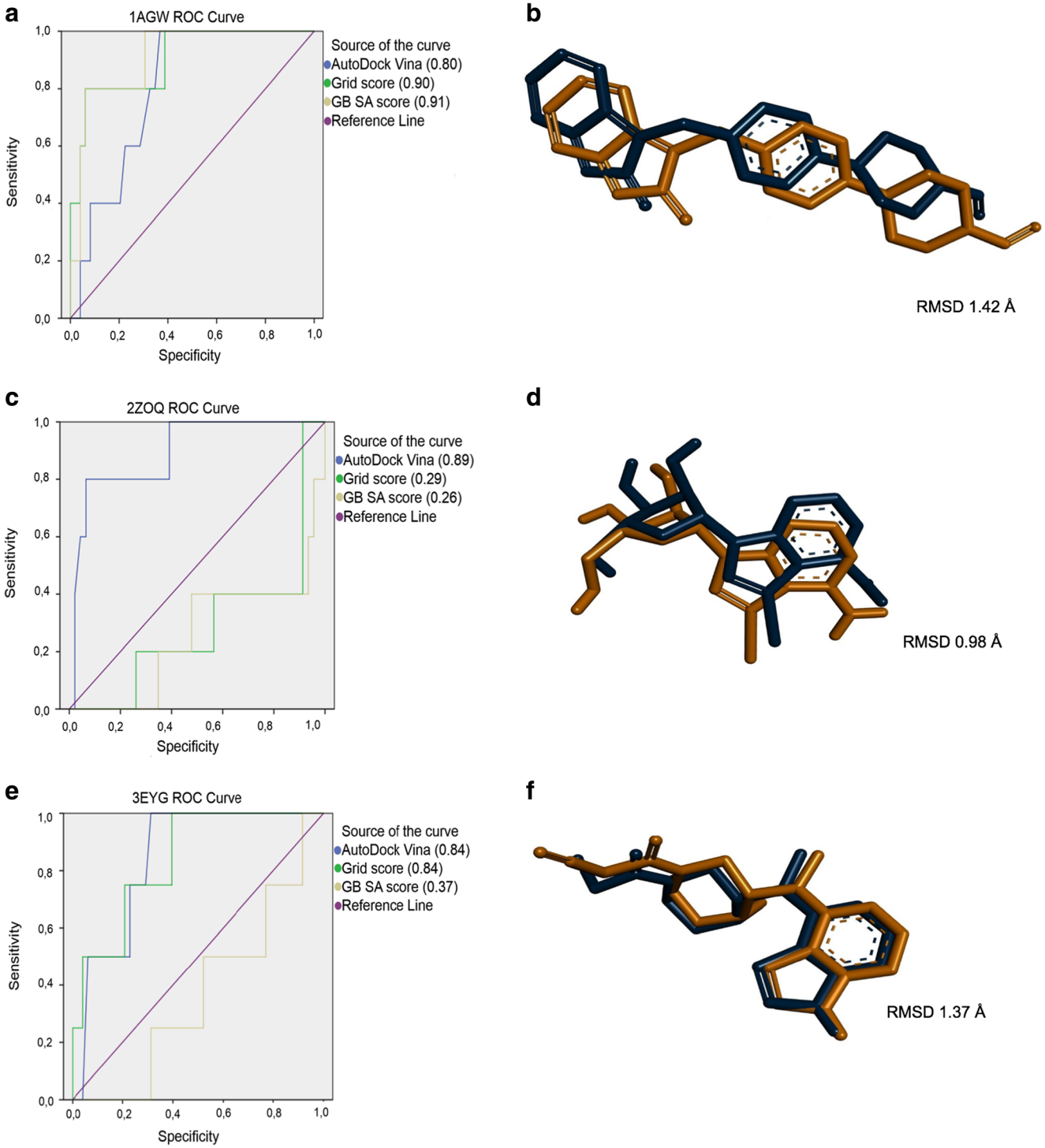

Fig. 1 **a–f** ROC curves and AUC values for selected molecular targets from OOMT (**a** 1AGW, **c** 2ZOQ, and **e** 3EYG). **b** (for 1AGW), **d** (for 2ZOQ), and **f** (for 3EYG) show overlays of the crystallographic ligands (*yellow*) on the redocked structures (*blue*), as calculated using the AutoDock Vina program

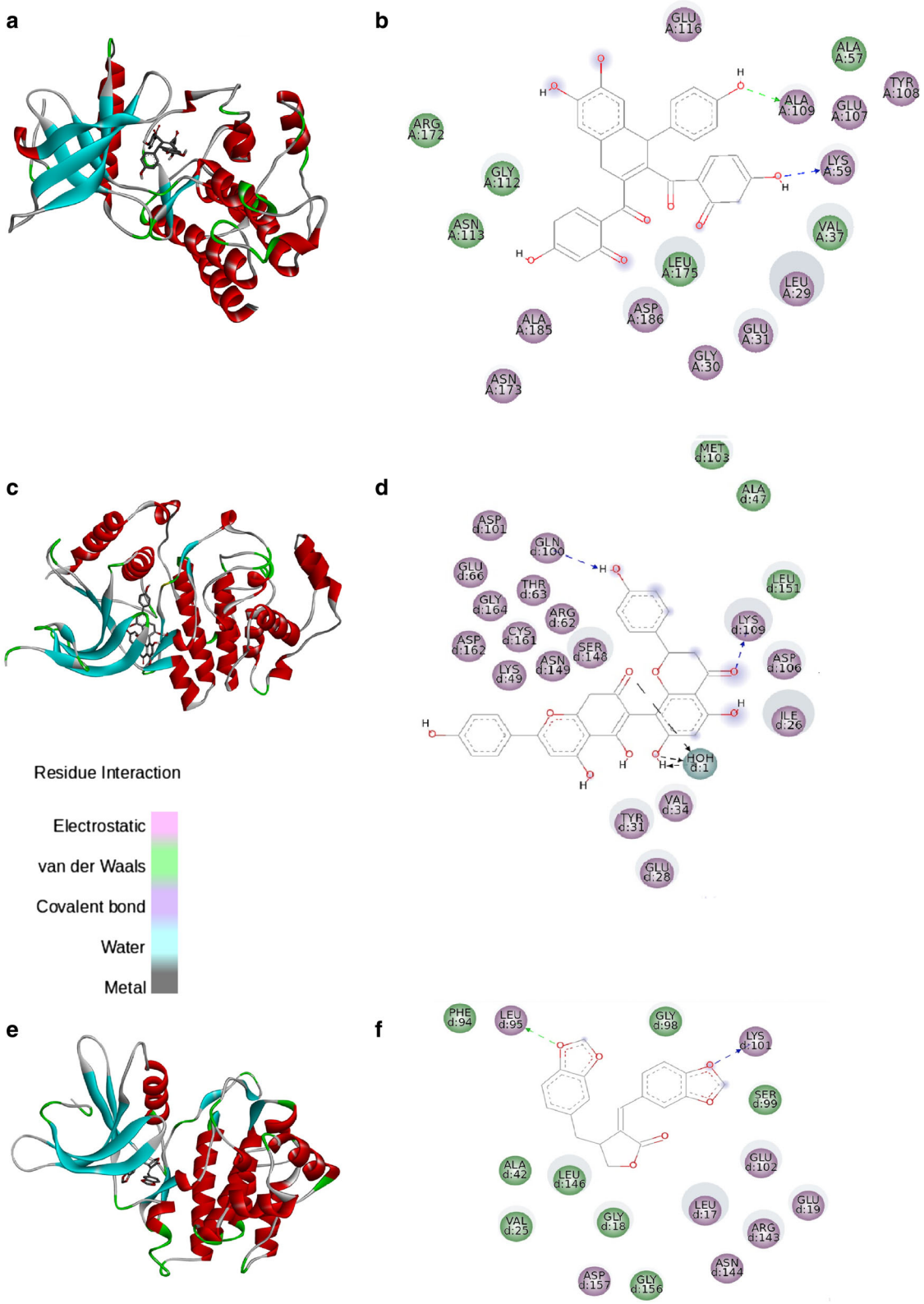

Fig. 2 **a**–**f** DBVS-selected results for the docking of natural compounds into targets from the OOMT. The complexes 1AGW–ZINC 69482570, 2ZOQ–ZINC06041137, and 3EYG–ZINC1342926 are shown in **a**, **c**, and **e**, respectively, whereas the 2D intermolecular interaction diagrams are shown in **b**, **d**, and **f**. Van der Waals and electrostatics interactions are shown in *green* and *pink*, respectively, and the *dashed lines* represent hydrogen bonds

in the ESM show the binding energy values obtained with the AutoDock Vina and DOCK 6 programs, respectively. To help visualize the results of the study, we also present in Fig. 1 the molecular targets with the lowest binding energy obtained upon re-docking 1AGW, 2ZOQ, and 3EYG (whose AUC values were 0.8, 0.9, and 0.8, respectively) using AutoDock Vina. Figure 1 also shows the redocking results and the respective RMSD values for the crystallographic ligands when they bind to the 1AGW, 2ZOQ, and 3EYG receptors.

Because collections of compounds are generally huge databases, it is impractical to apply the DBVS process to all of them in a rational drug design approach. Thus, the use of filters is recommended to improve the success rate [5]. In this context, the fragment-based drug-likeness [32] was used as a filter with the 473 compounds from ZINC. The fragment-based drug-likeness methodology scores ligand fragments from 15,000 commercial compounds. A positive value indicates that the compound has similar functional groups to most of the available drugs on the market. In general, most commercial drugs have a drug-likeness value of close to 2. As a result, 96 compounds showed a positive fragment-based drug-likeness, and those compounds were docked against ten and six receptors using AutoDock Vina and DOCK 6 (grid-based score), respectively. The drug-likeness results for the fragments are provided in Table 2, and the energy values obtained using DBVS are shown in Tables 3 and 4 of the ESM.

Because of the large amount of data, the molecular targets with the lowest redocking binding energies, which were 1AGW, 2ZOQ, and 3EYG [46–48] (with binding energies of −7.7, −7.1, and −7.4 kcal/mol, respectively), and the compounds ZINC13513540, ZINC06041137, and ZINC1342926 (with drug-likeness values of 0.35, 0.81, and 0.75, respectively) were selected to illustrate the DBVS experiment. All of them were obtained from the AutoDock Vina results, which indicated that the 1AGW–ZINC69482570, 2ZOQ–ZINC06041137, and 3EYG–ZINC1342926 complexes were formed with binding energies of −9.6, −9.9, and −9.5 kcal/mol, respectively (Fig. 2a, c, d). These results suggest that these ligands could be more active than the corresponding crystallographic ligands. These molecular targets are associated with cancer, particularly hematopoietic malignancies and tumor angiogenesis.

The intermolecular interactions responsible for molecular recognition are shown in Fig. 2b, d, f as 2D diagrams. As observed, there are electrostatic interactions with Leu29, Gly30, Glu31, Lys59, Glu107, Try108, Ala109, Glu116, Asn173, Ala185, and Asp186 in the complex 1AGW–ZINC69482570. In addition, van der Waals interactions occur with Val37, Ala57, Gly112, Asn113, Arg172, and Leu175, and two hydrogen bonds form with Lys59 and Ala109 (Fig. 1b). Similarly, the overall intermolecular interaction between 2ZOQ and ZINC06041137 was observed to involve electrostatic interactions with Ile26, Glu28, Tyr31, Val34, Lys49, Arg62, Thr63, Glu66, Gln100, Asp101, Asp106, Lys109, Ser148, Asn149, Vys161, Asp162, and Gly164 as well as van der Waals interactions with Ala47, Met103 and Leu151 (Fig. 2d). Additionally, Fig. 2d highlights a hydrogen bond between a water molecule and ZINC06041137. Moreover, a sodium ion was removed from the binding site of 2ZOQ. This ion is an artifact that was used in the crystallization process as a buffer. In general, docking protocols suggest the removal of any ion that does not interact with the ligand [49]. Finally, Fig. 1f shows the interactions of the molecular target 3EYG with ZINC13462926. Here, electrostatic interactions occur with Leu17, Glu19, Leu95, Lys101, Glu102, Arg143, Asn144, and Asp157, and van der Waals interactions occur with Gly18, Val25, Ala42, Phe94, Gly98, Ser99, Leu146, and Gly156.

Finally, our results show that the use of a huge database to search for new lead compounds should be performed with a criterion. The main problem with DBVS is the selection of false-positive compounds, which leads to source loss in biological assay experiments. Thus, our results highlight new hit compounds and suitable molecular targets among 473 ligands and 31 molecular targets, thus prompting a lead-compound optimization process in subsequent steps and increasing the chances of drug development success. Moreover, if the computational cost is considered, AutoDock Vina is faster than the grid-based score implemented in DOCK 6. This advantage motivated our group to develop the Octopus software [50, 51], an automated workflow management tool that is sufficiently scalable to allow high-throughput virtual screening (vHTS) and integrates MOPAC2016, AutoDock tools, and AutoDock Vina. Octopus can prepare the ligands using the PM7 method after adding Gasteiger charges and before they are automatically docked into OOMT molecular targets with a usability interface. In addition, Octopus can check the ligand structure for the presence of an unpaired electron. These characteristics make DBVS faster and more accurate.

Conclusion

In general, natural compounds are difficult to obtain and require a large quantity of source material, so their full biological activities are often underestimated. Furthermore, unlike synthetic compound collections, they have wide structural diversity. Thus, many experimental techniques are required to determine the most suitable biological activity of each natural compound. In this context, new technologies that can search for the most promising candidates in the chemical space of natural compounds should be evaluated to check whether they could be useful in rational drug design.

Generally, the initial (and sometimes only) evaluation of a docking methodology is achieved by calculating the RMSD. Our results were improved by calculating the AUC, which

increased the accuracy of the results. The AUC is a statistical methodology used to discriminate between false- and true-positive results of a test. It is widely used in various subjects, including drug development. When applied to DBVS, the AUC enables the methodology's ability to recover active compounds in preference to inactive compounds from a collection of ligands to be assessed [8]. The AUC can be applied to evaluate the accuracy of a test and to compare methods on this basis. An AUC value of close to 1.0 indicates that a test is good at distinguishing true-positive results from false-positive results. If the AUC value is equal to 0.5, the test returns true-positive and false-positive results at random. In this study, the AUC was calculated for OOMT proteins, and its value was used to validate three docking methodologies using two software programs. Our findings were observed to be consistent with previous docking studies in that structure-dependent evaluation was found to be necessary [11].

DBVS validation methodologies are becoming increasingly popular, and performing VS tests without applying these tools reduces the reliability of the study. After validating the molecular docking method using the AUC and RMSD, the reliability of the DBVS test enabled us to identify compounds that showed smaller binding energies than those for the crystallographic ligand of each molecular target, and the compounds selected could then be examined in biological assays. Our results suggest that the AutoDock Vina program is more appropriate for performing DBVS experiments, and our group also uses Octopus to make high-throughput DBVS more feasible. Finally, experimental assays of the selected hits are in progress.

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 the Graduate Program in Pharmaceutical Sciences from the Federal University of Sao Joao del Rei (PPGCF/UFSJ).

References

1. Khanna I (2012) Drug discovery in pharmaceutical industry: productivity challenges and trends. Drug Discov Today 17:1088–1102. doi:10.1016/j.drudis.2012.05.007
2. Geromichalos GD, Alifieris CE, Geromichalou EG, Trafalis DT (2016) Overview on the current status of virtual high-throughput screening and combinatorial chemistry approaches in multi-target anticancer drug discovery; part I. J Buon 21:764–779
3. Cheng T, Li Q, Zhou Z et al (2012) Structure-based virtual screening for drug discovery: a problem-centric review. AAPS J 14:133–41. doi:10.1208/s12248-012-9322-0
4. Lauro G, Romano A, Riccio R, Bifulco G (2011) Inverse virtual screening of antitumor targets: pilot study on a small database of natural bioactive compounds. J Nat Prod 74:1401–1407. doi:10.1021/np100935s
5. Glaab E (2016) Building a virtual ligand screening pipeline using free software: a survey. Brief Bioinform 17:352–366. doi:10.1093/bib/bbv037
6. Bursulaya BD, Bursulaya BD, Totrov M et al (2004) Comparative study of several algorithms for flexible ligand docking. J Comput Aided Mol Des 17:755–763
7. Ding Y, Fang Y, Moreno J et al (2016) Assessing the similarity of ligand binding conformations with the contact mode score. Comput Biol Chem 64:403–413. doi:10.1016/j.compbiolchem.2016.08.007
8. Triballeau N, Acher F, Brabet I et al (2005) Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4. J Med Chem 48:2534–2547. doi:10.1021/jm049092j
9. Zhao W, Hevener KE, White SW et al (2009) A statistical framework to evaluate virtual screening. BMC Bioinformatics 10:225. doi:10.1186/1471-2105-10-225
10. Lätti S, Niinivehmas S, Pentikäinen OT (2016) Rocker: open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. J Cheminform 8:45. doi:10.1186/s13321-016-0158-y
11. Westermaier Y, Barril X, Scapozza L (2015) Virtual screening: an in silico tool for interlacing the chemical universe with the proteome. Methods 71:44–57. doi:10.1016/j.ymeth.2014.08.001
12. Li JW-H, Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? Science 325:161–165. doi: 10.1126/science.1168243
13. Trott O, Olson AJ (2009) 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.
14. 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
15. Carregal AP, Comar M Jr, Taranto AG (2013) Our Own Molecular Targets Data Bank (OOMT). Biochem Biotechnol Rep 2:14–16
16. Rose PW, Prlić A, Bi C et al (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res 43:D345–D356. doi:10.1093/nar/gku1214
17. Irwin JJ, Sterling T, Mysinger MM et al (2012) ZINC: a free tool to discover chemistry for biology. J Chem Inf Model 52:1757–1768. doi:10.1021/ci3001277
18. Walters WP, Murcko MA (2002) Prediction of "drug-likeness". Adv Drug Deliv Rev 54:255–271. doi:10.1016/S0169-409X(02)00003-0
19. Gaulton A, Bellis LJ, Bento AP et al (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:1100–1107. doi:10.1093/nar/gkr777
20. Bento AP, Gaulton A, Hersey A et al (2014) The ChEMBL bioactivity database: an update. Nucleic Acids Res 42:D1083–D1090. doi:10.1093/nar/gkt1031
21. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49:6789–6801. doi:10.1021/jm0608356
22. 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
23. 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/rna.1563609
24. Meng EC, Shoichet BK, Kuntz ID (1992) Automated docking with grid-based energy evaluation. J Comput Chem 13:505–524. doi:10.1002/jcc.540130412

25. Zou X, Sun Y, Kuntz I (1999) Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. J Am Chem Soc 121:8033–8043. doi:10.1021/ja984102p
26. Liu HY, Kuntz ID, Zou XQ (2004) Pairwise GB/SA scoring function for structure-based drug design. J Phys Chem B 108:5453–5462. doi:10.1021/jp0312518
27. IBM Corporation (2010) IBM SPSS Statistics for Windows. IBM Corporation, Armonk
28. Munkres J (1957) Algorithms for the assignment and transportation problems. J Soc Ind Appl Math 5:32–38. doi:10.1137/0105003
29. Kuntz ID, Blaney JM, Oatley SJ et al (1982) A geometric approach to macromolecule–ligand interactions. J Mol Biol 161:269–288. doi:10.1016/0022-2836(82)90153-X
30. Accelrys Software Inc. (2015) Discovery Studio Modeling Environment, release 4.1. Accelrys Software Inc., San Diego
31. dos Santos IA (2012) Construção de uma base de dados para triagem virtual de potenciais inibidores da UDP-*N*-acetilglicosamina pirofosforilase do *Moniliophthora perniciosa*. Universidade Estadual de Feira de Santana
32. Sander T, Freyss J, von Korff M, Rufener C (2015) DataWarrior: an open-source program for chemistry aware data visualization and analysis. J Chem Inf Model 55:460–473. doi:10.1021/ci500588j
33. ChemAxon (2015) MarvinSketch. ChemAxon, Budapest
34. Stewart JJP (2013) Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J Mol Model 19:1–32. doi:10.1007/s00894-012-1667-x
35. Stewart JJP (2012) MOPAC. Stewart Computational Chemistry, Colorado Springs
36. Balius TE, Mukherjee S, Rizzo RC (2011) Implementation and evaluation of a docking-rescoring method using molecular footprint comparisons. J Comput Chem 32:2273–89. doi:10.1002/jcc.21814
37. Brozell SR, Mukherjee S, Balius TE et al (2012) Evaluation of DOCK 6 as a pose generation and database enrichment tool. J Comput Aided Mol Des 26:749–773. doi:10.1007/s10822-012-9565-y
38. Gordon JC, Myers JB, Folta T et al (2005) H++: a server for estimating pK_as and adding missing hydrogens to macromolecules. Nucleic Acids Res 33:W368–W371. doi:10.1093/nar/gki464
39. Myers J, Grothaus G, Narayanan S, Onufriev A (2006) A simple clustering algorithm can be accurate enough for use in calculations of p*K*s in macromolecules. Proteins Struct Funct Bioinf 63:928–938. doi:10.1002/prot.20922
40. Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: Automating p*K* prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res 40:537–541. doi:10.1093/nar/gks375
41. Morris GM, Huey R, Lindstrom W et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791. doi:10.1002/jcc.21256
42. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera? A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. doi:10.1002/jcc.20084
43. Gasteiger J, Marsili M (1978) A new model for calculating atomic charges in molecules. Tetrahedron Lett 19:3181–3184. doi:10.1016/S0040-4039(01)94977-9
44. 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
45. 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
46. Mohammadi M, McMahon G, Sun L, et al. (1997) Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors. Science 276:955–960
47. Kinoshita T, Yoshida I, Nakae S et al (2008) Crystal structure of human mono-phosphorylated ERK1 at Tyr204. Biochem Biophys Res Commun 377:1123–1127. doi:10.1016/j.bbrc.2008.10.127
48. Williams NK, Bamert RS, Patel O et al (2009) Dissecting specificity in the Janus kinases: the structures of JAK-specific inhibitors complexed to the JAK1 and JAK2 protein tyrosine kinase domains. J Mol Biol 387:219–232. doi:10.1016/j.jmb.2009.01.041
49. 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
50. Nunes RR, dos Santos Costa M, dos Reis Santos B, et al (2016) Successful application of a virtual screening and molecular dynamics simulation against antimalarial molecular targets. Mem Inst Oswaldo Cruz 1–10. doi:10.1590/0074-02760160207
51. Habib E, Maia B, Campos VA, et al (2017) Octopus: a platform for the virtual high-throughput screening of a pool of compounds against a set of molecular targets. J Mol Model 23:26. doi: 10.1007/s00894-016-3184-9