In silico screening strategies for novel inhibitors of parasitic diseases

http://www.producao.usp.br/handle/BDPI/49713

Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo
**In silico** screening strategies for novel inhibitors of parasitic diseases

Rafaela S Ferreira, Rafael VC Guido, Adriano D Andricopulo & Glaucius Oliva

Laboratório de Química Medicinal e Computacional, Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos-SP, Brazil

**Introduction:** Parasitic diseases are a major global problem causing long-term disability and death, with severe medical and psychological consequences around the world. Despite the prevalence of parasitic disease, the treatment options for many of these illnesses are still inadequate and there is a dire need for new antiparasitic drugs. In silico screening techniques, which are powerful strategies for hit generation, are widely being applied in the design of new ligands for parasitic diseases.

**Areas covered:** This article analyses the application of ligand- and structure-based virtual screening strategies against a variety of parasitic diseases and discusses the benefits of the integration between computational and experimental approaches toward the discovery of new antiparasitic agents. The analysis is illustrated by recent examples, with emphasis on the strategies reported within the past 2 years.

**Expert opinion:** Virtual screening techniques are powerful tools commonly used in drug discovery against parasitic diseases, which have provided new opportunities for the identification of several novel compound classes with antiparasitic activity.

**Keywords:** drug discovery, neglected diseases, virtual screening


1. Introduction

Parasitic diseases are a major global cause of illness, morbidity, long-term disability and death, with severe medical and psychological consequences for millions of men, women and children [1]. Despite the high prevalence of parasitic diseases worldwide, in most cases their treatment is inadequate, generating an urgent demand for new antiparasitic drugs. However, in addition to the traditional challenges involved in the complex process of drug discovery and development, there is the hurdle of the lack of investments in this field [2]. This situation is especially problematic in de novo drug discovery, regarded as a high risk and costly process [3]. Therefore, strategies that allow high quality hit identification rate as well as reduction in drug discovery costs are extremely useful in this field.

The biology of parasitic organisms has been continuously studied in detail, providing a solid base for the selection of relevant molecular targets for drug discovery. Usually, hit and lead discovery begin with the application of either experimental or in silico high-throughput screening (HTS) strategies [4]. In the particular case of parasitic diseases, broadly investigated in academia, virtual screening (VS) strategies play a major role in comparison to the traditional HTS. This is not only a consequence of the lower costs and less infrastructure required, but also of the advantages of this modern approach, which allows the identification of hits from a set of privileged compounds.
In silico screening strategies for novel inhibitors of parasitic diseases

### Article highlights.

- Ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS) have been used in the discovery of new classes of compounds that significantly differ from the original set of known actives and ii) the complexity to deal with the presence of activity cliffs (i.e., substantial differences in biological activity in very similar compounds) within structure-activity relationship (SAR) guided series [17]. However, recent approaches using LBVS highlight its power and versatility for drug discovery being applied to the identification of new classes of compounds that significantly differ from the ligands used to derive the models or integrated with scaffold hopping as a relevant tool to increase structural diversity and exploit unpatented chemical space [12,13,16].

In the field of parasitic diseases, recent successful examples underscore the diversity of ligand-based strategies used to build models capable of explaining and predicting the biological activity within a particular series of antiparasitic compounds. Inhibitors of the enzymes cruzain and glyceraldehydes-3-phosphate dehydrogenase (GAPDH) from Trypanosoma cruzi, molecular targets for Chagas disease, have been investigated by 2D and 3D quantitative structure-activity relationship (QSAR) methods [7,18-20]. Additionally, studies based on 2D and 3D QSAR strategies have been used to investigate the biological activity underlying a series of benzimidazoles against Trichomonas vaginalis and Giardia intestinalis [21]. Regardless of the methods used in the VS, the strategy consists in rationally selecting a set of high ranked molecules for experimental evaluation. The integration of VS and QSAR strategies provides useful opportunities to capture valuable information for compound selection (Figure 1). For instance, in the design of new anticoccidial agents, QSAR models were generated based on a set of 38 known drugs and 144 inactive compounds, randomly divided into training and test sets. The best QSAR model was subsequently used in a VS campaign to prioritize the compounds for experimental testing. Ultimately, this combined approach guided the identification of compound 1 with substantial in vivo activity [22]. Similarly, molecular fingerprints and QSAR models methods have been used in the discovery of new classes of antitrypanosomal (2) [6], antitrichomonal (3) [23] and antimalarial (4) compounds (Figure 1) [5].

### 2. Ligand-based virtual screening

Ligand-based virtual screening (LBVS) strategies are powerful methods that enable the discovery of bioactive compounds based only on small-molecule information. LBVS essentially focus on comparative molecular similarity analysis of compounds with known and unknown activity. To this end, a panel of known active and inactive compounds is used to build robust models. Subsequently, the models are used to select and sort the library molecules according to their likelihood of binding to the target of interest. The VS process can be based on several techniques, including molecular similarity methods [12], pharmacophore models [14] or machine learning methods [15]. These methods have been recently reviewed elsewhere [15-17].

Challenges in the field include: i) the generation of models able to identify ligands that significantly differ from the original set of known actives and ii) the complexity to deal with the presence of activity cliffs (i.e., substantial differences in biological activity in very similar compounds) within structure-activity relationship (SAR) guided series [17]. However, recent approaches using LBVS highlight its power and versatility for drug discovery being applied to the identification of new classes of compounds that significantly differ from the ligands used to derive the models or integrated with scaffold hopping as a relevant tool to increase structural diversity and exploit unpatented chemical space [12,13,16].

In the field of parasitic diseases, recent successful examples underscore the diversity of ligand-based strategies used to build models capable of explaining and predicting the biological activity within a particular series of antiparasitic compounds. Inhibitors of the enzymes cruzain and glyceraldehydes-3-phosphate dehydrogenase (GAPDH) from Trypanosoma cruzi, molecular targets for Chagas disease, have been investigated by 2D and 3D quantitative structure-activity relationship (QSAR) methods [7,18-20]. Additionally, studies based on 2D and 3D molecular similarity have been used to investigate the biological activity underlying a series of benzimidazoles against Trichomonas vaginalis and Giardia intestinalis [21]. Regardless of the methods used in the VS, the strategy consists in rationally selecting a set of high ranked molecules for experimental evaluation. The integration of VS and QSAR strategies provides useful opportunities to capture valuable information for compound selection (Figure 1). For instance, in the design of new anticoccidial agents, QSAR models were generated based on a set of 38 known drugs and 144 inactive compounds, randomly divided into training and test sets. The best QSAR model was subsequently used in a VS campaign to prioritize the compounds for experimental testing. Ultimately, this combined approach guided the identification of compound 1 with substantial in vivo activity [22]. Similarly, molecular fingerprints and QSAR models methods have been used in the discovery of new classes of antitrypanosomal (2) [6], antitrichomonal (3) [23] and antimalarial (4) compounds (Figure 1) [5].

### 3. Structure-based virtual screening

Advances in molecular biology have been essential for the identification and validation of biological targets of pharmaceutical interest, while genomic and proteomic approaches have fundamentally contributed to the analysis of 3D macromolecular organization and function. Concomitantly, improvements in physical techniques applied for structural determination and molecular analyses, such as X-ray crystallography, NMR and calorimetry, provided a deep understanding of both spatial and energetic components underlying ligand–receptor binding interactions [3]. Structure-based drug design methods incorporate information from the target receptor; hence, these knowledge-driven approaches require a good deal of information about the target topology under investigation (e.g., X-ray crystal structure, NMR or robust homology
modeling) [4]. In this context, public databases such as the Protein Data Bank, Protein Data Bank in Europe and the Structural Biology Knowledge Base are important data sources to retrieve and analyze 3D structures of target proteins.

The wealth of structural information of attractive molecular targets and the evolution of computational methods have prompted the development of a highly specialized screening tool to identify promising compounds based on their complementarity to a specific binding site. This method, known as structure-based virtual screening (SBVS), has made a remarkable impact on the discovery of new drug candidates [3,4,10].

Owing to its intrinsic features, SBVS strongly depends on the amount and quality of data available about the biological system under investigation. Some general considerations regarding the suitability of the molecular target for SBVS studies include, but are not limited to: i) assessment of target validation and druggability, incorporating features such as pocket size, geometry, surface complexity and roughness, and their complementarity in shape and polarity with respect to a putative drug-like ligand; ii) analysis and selection of the most relevant geometry of the target receptor, including macromolecular flexibility for ligand binding; iii) assignment of the correct protonation and tautomeric states, which is required for the definition of crucial molecular properties for ligand binding and affinity (e.g., local dielectric conditions within the binding pocket can modulate pKa values of functional groups which can easily turn a hydrogen acceptor group into a donor or a charge-assisted hydrogen bond into a neutral one) and iv) definition of structurally conserved water molecules in the receptor binding site, which should be taken into consideration for model development [4,10].

The molecular recognition phenomenon relies on the properties and features of a binding pocket, which are determined by the amino acids present in the binding cavity. The spatial arrangement of the amino acids within the binding site specifies structural and physicochemical constraints that must be met by any putative ligand. Consequently, a detailed analysis of the stereo-electronic properties of the target binding pocket provides useful insights into relevant ligand–receptor interactions. In this context, the use of the structure-based pharmacophores has been successfully used in VS of drug-like databases [24]. The pharmacophore models are used to select compounds with specific chemical and spatial features that represent the essential interactions (e.g., hydrogen bonding, charge transfer, electrostatic, hydrophobic interactions) between small-molecule ligands and the receptor binding pocket.

The use of structure-based pharmacophore models is especially attractive for drug discovery for parasitic diseases. Often, the host and the parasite share common biochemical pathways and the development of compounds which selectively bind to the parasite enzymes is required. In this context, pharmacophore models are commonly used to explore the structural differences between close homologs for the design of selective inhibitors. The following examples provide a perspective of the utility of the pharmacophore-based VS and its integration.
with other important experimental and computational strategies in medicinal chemistry.

An example of the integration of pharmacophore models and SBVS can be observed in the identification and optimization of a novel class of inhibitors of Schistosoma mansoni purine nucleoside phosphorylase (SmPNP), a key enzyme involved in the purine salvage pathway of S. mansoni [25,26]. In this study, a 3D pharmacophore model was used to select compounds with new structural scaffolds as competitive inhibitors of SmPNP. By applying this filter, the initial set containing > 300,000 molecules was reduced by 98% [8]. The resulting focused-library was sequentially docked into the SmPNP binding site. Ultimately, this combined strategy allowed the identification of three thioxothiazolidinones derivatives (5 – 7) as new reversible and competitive inhibitors of SmPNP with IC$_{50}$ values in the low micromolar range (Figure 2). These inhibitors represent new potential lead compounds for further development for the therapy of schistosomiasis.

The impact of pharmacophore models on SBVS can also be seen in the discovery of pyrazole-urea derivatives as potential candidates for the treatment of malaria. In the apicomplexan parasites, the actin–myosin A (MyoA) motor complex is a key component in the gliding motility that is required for the cell invasion process. A critical interaction occurs between the myosin tail interacting protein (MTIP) and myosin A (MyoA) facilitating the gliding action of the parasite [27]. The inhibition of this protein–protein interaction impairs the gliding motility, thereby preventing the parasite to invade the host cells. In order to identify compounds that target the MTIP–MyoA interaction interface, a detailed structural analysis of the MTIP–MyoA complex was performed. The analysis supported the construction of a four-point structure-based pharmacophore model, which was subsequently used to screen a library of nearly 300,000 compounds. Of these, 40 compounds were selected and docked into the protein–protein interface. Finally, 15 compounds were acquired and tested against Plasmodium falciparum cultures. The most potent inhibitor (compound 8, EC$_{50}$ = 145 nM) was selected as lead compound for SAR studies (Figure 2). Several derivatives with EC$_{50}$ values in the nanomolar range were identified as inhibitors of the parasite growth. The pharmacophore model used in this work was designed to identify potential inhibitors that block the interaction between MTIP and MyoA, consequently inhibiting the gliding motility of the parasite. To explore the mechanism of inhibition in more detail, the most potent inhibitors were tested for their ability to reduce the parasite’s motility. The series of pyrazole-urea derivatives (e.g., compound 9, EC$_{50}$ = 385 nM) has shown impairment in gliding motility, providing indirect evidence that the compounds inhibit the MTIP–MyoA interaction (Figure 2) [27].

The use of robust structural and biochemical information for specific molecular systems has significantly contributed to the discovery of antiparasitic agents bearing innovative scaffolds [3,10]. This assumption is corroborated by the identification of the structural determinants for ligand binding of the enzyme pteridine reductase 1 from Trypanosoma brucei (TbPTR1), an attractive biological target for the therapy of human African trypanosomiasis (HAT) [28]. X-ray crystallographic studies revealed extensive van der Waals and hydrogen bonding interactions in the tertiary complexes incorporating potent inhibitors (co-crystallized ligands), the cofactor NADP$^+$ and key amino acids of TbPTR1. Additionally, a close inspection of the TbPTR1 binding pocket indicated important structural features for a fragment-based approach. In light of this, a robust structure-based pharmacophore model was generated and then used to screen a database of 250,000 compounds (Figure 2). It is worth noting that the initial database was reduced by ~90% after the application of specialized fragment-like property filters (e.g., < 20 heavy atoms, 1 – 2 ring systems, ≥ 1 hydrogen-bond donor group, < 4 rotatable bonds, and ClogP/ClogD < 3.5) [28]. The designed fragment library was docked into the TbPTR1 binding site and the predicted orientation evaluated according to the pharmacophore model. The criteria for compound selection were based on the quality of the hydrogen-bond network and the shape complementarity between the ligands and the binding site. This procedure resulted in the identification of a series of 45 compounds that have their inhibitory activity evaluated against TbPTR1. The aminobenzimidazole derivative 10 (K$_i$ = 10 µM) was selected as an innovative scaffold for further SAR and crystallographic studies (Figure 2). On the basis of the docked binding mode, several analogs were synthesized and tested in order to evaluate the SARs underlying this series. The investigation led to the development of compound 11 (K$_i$ = 0.4 µM), which exhibited binding affinity 25-fold higher than the parent compound 10. Subsequently, the experimental binding mode of compounds 10 and 11 was determined using X-ray crystallography, which allowed the identification of a large hydrophobic pocket adjacent to the 7-position of the aminobenzimidazole core. The molecular modification of 11 with moieties suitable to fill the hydrophobic pocket led to the discovery of compound 12 (K$_i$ = 0.007 µM), a 7-phenyl-aminobenzimidazole derivative with binding affinity improved by >1400-fold in comparison to the initial VS hit. Finally, the inhibitory activity of 12 was assessed against T. brucei cultures, confirming that the compound was also active in vitro (EC$_{50}$ = 10 µM).

### 4. Integration of screening techniques

Several approaches using a combination of screening techniques have recently been applied in the search for new antiparasitic drugs, ranging from strategies that merge computational techniques to cases in which virtual and experimental screenings are used sequentially or in parallel to improve the hit rate. The power of these integrated approaches is demonstrated by their enhanced performance,
Figure 2. SBVS strategies for the development of new leads for schistosomiasis (green), malaria (purple) and human African trypanosomiasis (cyan). Pharmacophore models: H-donor groups (blue spheres); H-acceptor groups (red spheres); H-donor/acceptor groups (magenta spheres) and hydrophobic groups (cyan spheres).

SBVS: Structure-based virtual screening.
as demonstrated by the discovery of a large number of biologically active compounds [9,13,29]. Fundamentally, these methods differ in their speed and the necessary resources for their application. Therefore, combined strategies make it possible to use faster and cost-effective VS techniques as rational filters for prioritizing compounds for further detailed computational studies or experimental investigations.

A rich diversity of methods can be used in the integration of LBVS and SBVS. For instance, the chemical space of the hits identified by structure-based methods can be further explored by LBVS (e.g., similarity searches). Alternatively, LBVS may be used as a fast and reliable tool for the prioritization of compounds to be screened by SBVS. An example can be seen in the discovery of new trypanothione reductase inhibitors, a therapeutic relevant target to treat Chagas disease, HAT and leishmaniasis [12]. A library of > 8 million compounds was screened by LBVS using the Molinspiration VS ‘miscreen’, reducing the initial database to nearly 1300 compounds, which were then filtered based on absorption, distribution, metabolism and excretion (ADME)-Tox properties, further decreasing the number of compounds to ~600. The programs AUTODOCK and X-PLORE were used to dock the molecules within the enzyme binding site, leading to the selection of a final set of 19 compounds for biochemical evaluation. Of these, 10 compounds were active against T. cruzi trypanothione reductase, 6 of which with IC\text{50} < 50 \mu M (e.g., 13 – 18, Figure 3) [12]. The high hit rates obtained and the novelty of the compounds discovered underscored the power of the application of LBVS as an efficient filter to rationally prioritize compounds for SBVS.

VS techniques can be complemented by more sophisticated, lower throughput computational methodologies, such as molecular dynamics. The application of molecular dynamics simulations to a privileged set of compounds selected by VS approaches allows a more detailed evaluation of the molecular events and interactions between the ligands and target protein [11,30]. For example, to search for inhibitors of the T. cruzi GAPDH, ligand-based methods were used to select a panel of 35 natural products to be tested \textit{in vitro} against the enzyme. Of these, seven molecules exhibited inhibitory activity in the micromolar range. The three most potent compounds were then docked into the binding site of the enzyme, and the binding mode of the most potent compound was evaluated by molecular dynamics. The simulation indicated reasonable agreement with the SAR data, thereby providing detailed information underlying the structural requirements for ligand recognition and binding [11]. In another recent example, steered molecular dynamics has been successfully applied to identify true inhibitors from a set of active and inactive polyhydroxylated flavones as inhibitors of the \textit{P. falciparum} \alpha-hydroxyacyl-ACP dehydratase enzyme. In this case, the only differences among the structurally related compounds lay on the number and positioning of the hydroxyl substituents around the flavone core. Furthermore, a compound not included in the initial set was correctly predicted as an inhibitor of the target enzyme [30]. These examples illustrate the applicability and efficiency of molecular dynamics as a complementary strategy for LBVS and SBVS studies. However, it is worth mentioning that the computational costs and the complexity of the simulation methods may hinder its application as an actual screening strategy.

Improved performance for a significant number of targets was observed by the application of LBVS and SBVS parallel screening strategies as an \textit{in silico} tool for drug design. In this context, both approaches can be applied to the same library of compounds and the final prioritization is based on a consensus scoring and comparison of the methods (e.g., the sum of the normalized scores or the combination of the ranked compounds lists) [13]. Similarly, the benefits of combining \textit{in silico} and experimental HTS campaigns have recently been evaluated [4,29]. Such an approach is exemplified by the screening of the same library of nearly 198,000 compounds in parallel by both SBVS and HTS against the enzyme cruzain, the major cysteine protease of \textit{T. cruzi}. Given the high number of hits obtained in the HTS (~1000 compounds not easily identified as artifacts), the compounds were prioritized for confirmatory assays either based on their VS ranking (e.g., among the top 1%) or on the most frequent chemotypes among the hits. This combined approach led to the discovery of five classes of reversible competitive inhibitors, with K\text{\textit{i}} < 10 \mu M (e.g., compounds 19 – 23). Out of these five classes, only one was identified by both approaches (compound 21), whereas two classes were found exclusively by HTS (compounds 22 and 23) and the other two were poorly represented among HTS hits, which would have been ignored for secondary assays if the VS results were not taken into account (compounds 19 – 20) [9]. Accordingly, this study emphasizes the complementarity between these techniques, as each of the screening methods was able to identify ligands that would be otherwise missed if a single screening strategy was used (Figure 3).

\textbf{5. Expert opinion}

The application of \textit{in silico} screening is a powerful strategy that allows the identification of promising hits without the complex infrastructure required to perform experimental screenings (HTS). In the past few years, \textit{in silico} screening techniques have been successfully applied in the discovery of new classes of inhibitors for several key therapeutic targets for parasitic diseases. Although LBVS and SBVS present intrinsic technical limitations, there are many examples demonstrating that these methods are cost-effective and efficient in generating leads for further medicinal chemistry development. Interestingly, the classes of compounds identified usually differ considerably from the drugs already available. Therefore, these approaches provide a source for innovation and encouraging results have been obtained, providing starting points for lead optimization efforts. The combination
Figure 3. Integrated strategies for HTS. (A) Sequential application of LBVS and SBVS against trypanothione reductase (TryR). (B) Selected examples of TryR inhibitors discovered by LBVS and SBVS; (C) Integration of SBVS and HTS in the discovery of new classes of cruzain inhibitors.

of LBVS and SBVS, as well as the integration of SBVS and HTS, either sequential or in parallel can be beneficial as indicated by the increasing number of hits identified. Alternatively, the integrated approaches significantly contribute to the prioritization of compounds to be further investigated by more sophisticated computational approaches or experimental testing. It is expected that SBVS methods will continue to gain strength with the steadily increasing number of 3D structures experimentally determined. At the same time, new in silico strategies or the combination of the existing ones are constantly being attempted and assessed. There is a clear tendency of taking into account pharmacokinetic features, as early as possible, in addition to the small-molecule interactions related to affinity and biological activity, with the objective of improving potency and efficacy, incorporating ADME properties in the selection of molecules with drug (and lead)-like properties. The next challenges will consist in improving the quality and performance of both drug design methods, as well as their integration with other modern drug discovery strategies. We believe that this trend is likely to increase the efficiency of delivering drug candidates for parasitic diseases in the incoming years. To this end, partnerships with industries can be essential or, at least, highly recommended for boosting the success of the projects. In spite of the many scientific and technological challenges in this field, what is clear is that the integration of LBVS and SBVS methods will continue to enable and expand the application of these approaches in the development of new chemotherapy agents having promise of utility in clinical medicine.

### Declaration of interest

The authors state no conflict of interest and have received no payment in preparation of this manuscript.

### Bibliography


Affiliation
Rafaela S Ferreira, Rafael VC Guido, Adriano D Andricopulo & Glaucius Oliva

†Author for correspondence
Laboratório de Química Medicinal e Computacional, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-Carlense 400, 13560-970 São Carlos-SP, Brazil
Tel: +55 16 3373 9844; Fax: +55 16 3373 9881; E-mail: oliva@ifsc.usp.br