



**Universidade de São Paulo**

**Biblioteca Digital da Produção Intelectual - BDPI**

---

Departamento de Física e Ciência Interdisciplinar - IFSC/FCI

Artigos e Materiais de Revistas Científicas - IFSC/FCI

---

2008-01

# Virtual screening and its integration with modern drug design technologies

---

Current Medicinal Chemistry, Sharjah, v. 15, n. 1, p. 37-46, Jan. 2008

<http://www.producao.usp.br/handle/BDPI/49337>

*Downloaded from: Biblioteca Digital da Produção Intelectual - BDPI, Universidade de São Paulo*

# Virtual Screening and Its Integration with Modern Drug Design Technologies

Rafael V.C. Guido, Glaucius Oliva and Adriano D. Andricopulo\*

Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil

**Abstract:** Drug discovery is a highly complex and costly process, which demands integrated efforts in several relevant aspects involving innovation, knowledge, information, technologies, expertise, R&D investments and management skills. The shift from traditional to genomics- and proteomics-based drug research has fundamentally transformed key R&D strategies in the pharmaceutical industry addressed to the design of new chemical entities as drug candidates against a variety of biological targets. Therefore, drug discovery has moved toward more rational strategies based on our increasing understanding of the fundamental principles of protein-ligand interactions. The combination of available knowledge of several 3D protein structures with hundreds of thousands of small-molecules have attracted the attention of scientists from all over the world for the application of structure- and ligand-based drug design approaches. In this context, virtual screening technologies have largely enhanced the impact of computational methods applied to chemistry and biology and the goal of applying such methods is to reduce large compound databases and to select a limited number of promising candidates for drug design. This review provides a perspective of the utility of virtual screening in drug design and its integration with other important drug discovery technologies such as high-throughput screening (HTS) and QSAR, highlighting the present challenges, limitations, and future perspectives in medicinal chemistry.

**Keywords:** Drug design, virtual screening, QSAR, HTS, binding affinity.

## INTRODUCTION

The identification of promising hits and the generation of high-quality leads are crucial steps in the early stages of any drug discovery project. Recent advances in medicinal chemistry at the interface of chemistry and biology have created an important foundation in the search for new drug candidates possessing a combination of optimized pharmacodynamic and pharmacokinetic properties. Despite the impact of the recent technological and scientific advances, drug discovery has become more expensive and time consuming over the same period of time [1,2].

The widespread use of combinatorial chemistry and high-throughput screening (HTS) for the discovery of lead compounds has created a large demand for small organic molecules that act on specific drug targets. These technologies focus on the generation of a huge number of molecules integrated with the biological screening of a very large number of samples. However, due to the ever increasing pressure to reduce drug development time and costs, there is a clear paradigm shift from the random screening of collections of compounds to a more rational process, which would directly affect the success rate of new chemical entity (NCE) generation and, therefore, improve pharmaceutical research and development (R&D) productivity. The definition and assessment of both chemical and biological space have revitalized the screening process model and emphasized the importance of exploring the intrinsic complementary nature of classical and modern methods in drug research. In this context, computational tools play an increasingly critical role in medicinal chemistry research programs [3,4].

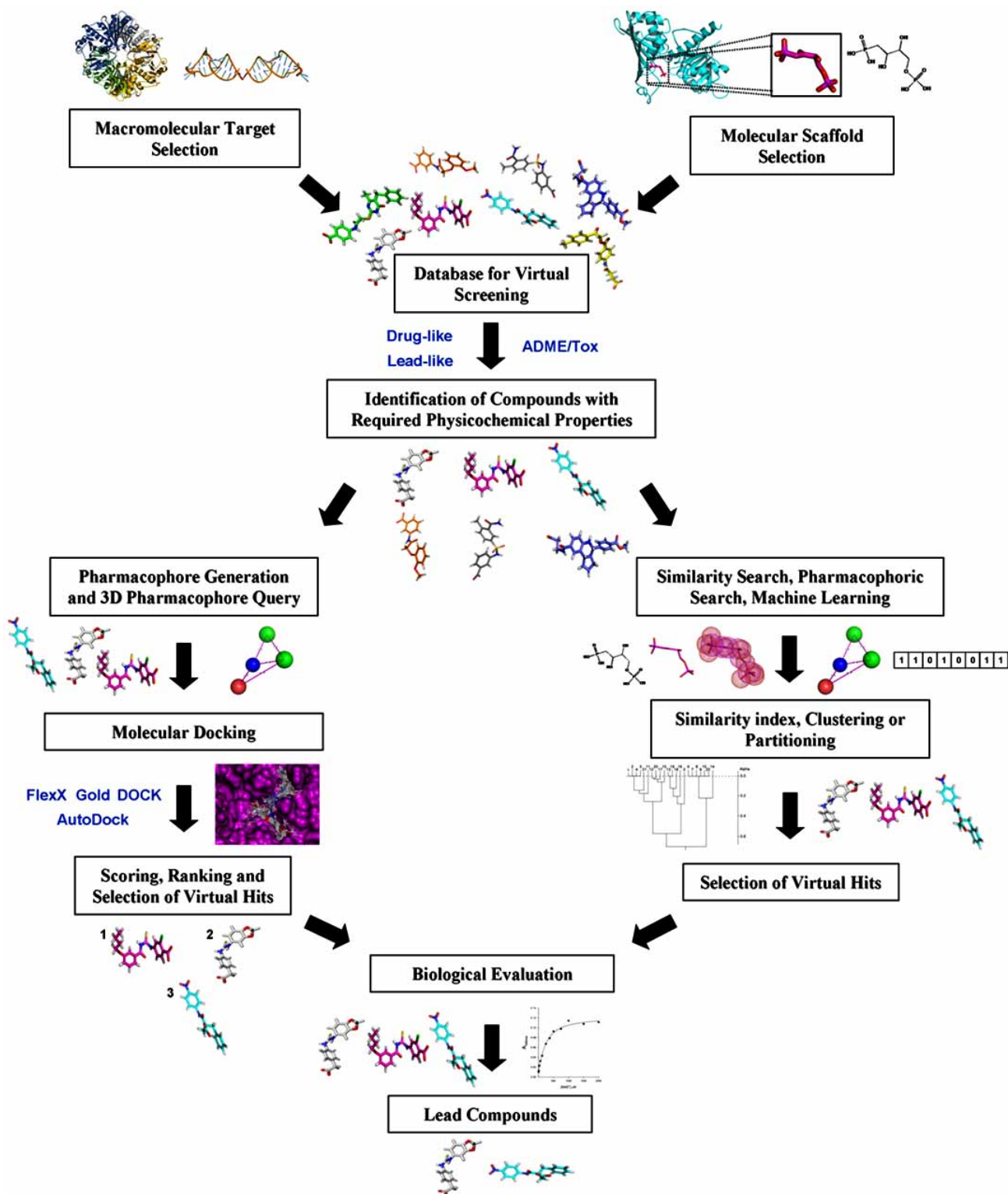
Drug discovery is currently driven by innovation and knowledge employing a combination of experimental and computational methods. One of the most important challenges for the pharmaceutical industry is the identification of innovative NCEs from an incredibly large reservoir of real and virtual possible compounds. Several steps of the drug discovery process (e.g., hit identification, lead optimization, pharmacokinetic profile) can be improved in a rational way with the application of computational methods. Over the past decade, the high-performance computers, algorithms, methods and expertise have evolved and transformed structure-based drug design (SBDD) methods in tools of large impact in

modern drug discovery [5,6]. Although several efforts have been made to improve our understanding of the three-dimensional interactions involved in ligand-receptor binding and molecular recognition, it remains a major challenge for the computational technologies to accurately predict the binding affinity of new drug candidates [7].

In the process of ligand binding, the Gibbs free energy is governed by a combination of complex intermolecular interactions which determines drug-receptor affinity. Therefore, in order to be able to understand the several steps involved and to predict ligand-binding affinity, it is very useful a partition of the free energy of binding into individual, physically interpretable terms. This is not, of course, an easy issue to tackle, particularly concerning the relative calibration of the individual contributions against each other. On one hand, the enthalpic contributions to binding constant are reasonably well characterized in molecular descriptors that encode structural and shape information or empirical functions, such as molecular mechanics approximations and statistically derived atom-type and distance-dependent pair potentials that explore the capabilities of molecules to perform intermolecular interactions with a putative receptor [5,8,9]. On the other hand, entropic contributions are much more difficult to describe, particularly regarding solvent-to-protein transfer, where both the size of the hydrophobic surface area and the release of water molecules from the active site should be considered. Furthermore, ligand conformational flexibility has long been recognized as an important issue, since immobilization of rotational bonds at the binding site involves important entropy changes.

In order to estimate free energies associated with interactions between small-molecules and drug targets (e.g., enzymes, receptors), *ab initio* calculations and free energy perturbation are the preferred computational methods [10]. However, the complexity of the calculations associated with the time-consuming procedures make these methods rarely applicable in drug design [5,10]. In contrast, protein-ligand docking is an area of intense interest to both academia and industry [11]. Computer programs dedicated to docking small-molecules into protein binding sites have been receiving considerable attention in recent years because of their wide applications in medicinal chemistry. Docking methods such as DOCK [12], GOLD [13], FlexX [14], GLIDE [15,16], AUTODOCK [17], and Surflex-Dock [18] are widely used for the high-throughput sampling of ligands into protein binding pockets, with concomitant determination of the most likely binding modes and the estimation of the relative binding affinity [19].

\*Address correspondence to this author at the Laboratório de Química Medicinal e Computacional, Centro de Biotecnologia Molecular Estrutural, Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense 400, 13560-970, São Carlos-SP, Brazil; Tel: + 55 16 3373-8095; Fax: +55 16 3373-9881; E-mail: andricopulo@ifsc.usp.br



**Fig. (1).** General procedure used in SBVS and LBVS.

## VIRTUAL SCREENING

In general, the search for new biologically active molecules from large compound databases by means of computer-assisted techniques is a process known as virtual screening (VS). VS methods have rapidly become an essential component of the modern drug discovery process [20]. High-performance hardware and specialized software, combined with advanced knowledge of 3D pro-

tein structure and small-molecule binding modes, have made this technology a useful complement, and in some cases, a reasonable alternative to HTS. There are, fundamentally, two approaches to VS studies: i. structure-based virtual screening (SBVS), which requires knowledge of the 3D structures of target proteins to prioritize compounds by their complementarity to the binding site; and, ii. ligand-based virtual screening (LBVS), where no information on the protein is needed, instead, compounds known to bind to the protein are

used as queries to search databases for new molecules possessing biological activity [21-25]. Fig. (1) summarizes some important steps in both VS approaches.

### STRUCTURE-BASED VIRTUAL SCREENING (SBVS)

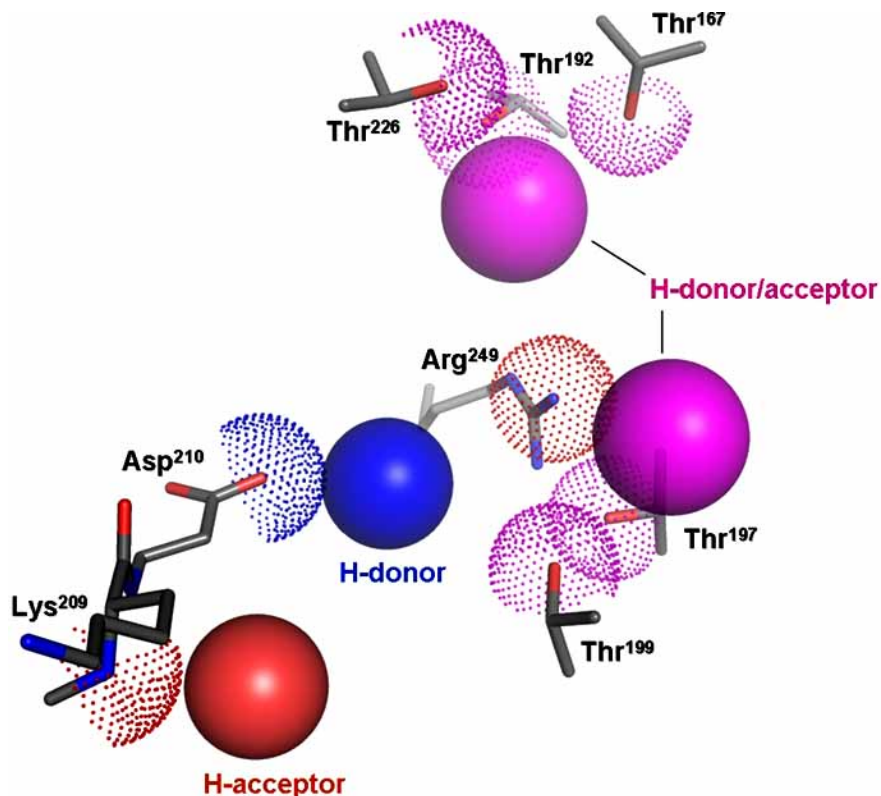
In SBVS approaches, knowledge about the 3D structures of the target proteins is essential to perform *in silico* high-throughput receptor-ligand docking [26]. These macromolecular structures are usually determined by X-ray crystallography, NMR, and homology modeling [27-29]. Owing to its knowledge-based feature, VS strongly depends on the amount and quality of information available about the system under investigation. Regardless of what kind of protein will be employed as molecular target, important issues such as druggability of the target receptor, selection of the most relevant geometry, receptor flexibility, suitable assignment of protonation states and consideration of water molecules in binding site must be properly consider [24]. Another important step in SBVS is the appropriate design of databases of small-molecule candidates for the screening process. Usually, pharmaceutical companies possess their own private compound collections, which guarantee that the detected hits will be exclusive and will cover molecules for which the synthesis is well established. In addition, hundreds of thousands of commercially available compounds for VS can also be found in non-commercial databases or *in-house* collections of natural products and synthetic compounds, and so forth [24,30,31].

Screening libraries generally contain a large number of molecules with broad chemical diversity [32]. Members of such libraries are usually large in terms of size, with an average molecular weight falling close to typical drug-size molecules. However, several examples from drug discovery programs have demonstrated that small core fragments known to bind with substantial affinity are more suitable starting points for further lead optimization [33,34]. There-

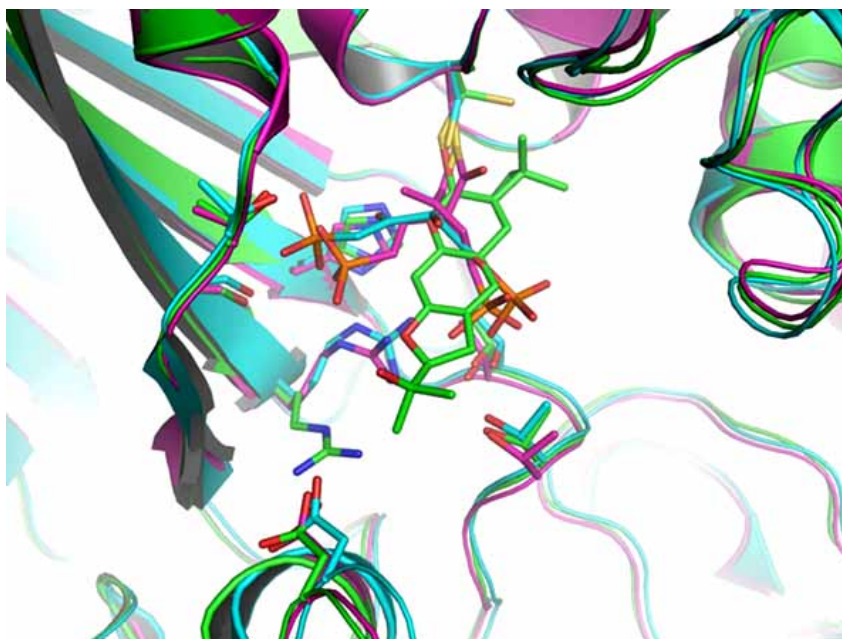
fore, in order to generate a suitable screening subset, several molecular filters are available for selecting those molecules with required features. As a general criteria, compounds are primarily studied with respect to their drug-like (molecules which generally obey Lipinski's rule of five [35]) or lead-like (molecules which have lower molecular complexity when compared to drugs [36-38]) properties. Similarly, other important physicochemical features may be considered, as well as the pharmacokinetic profile (often referred to as ADME properties: Absorption, Distribution, Metabolism and Excretion).

Pharmacophores derived from receptor mapping play an important role in the design of focused collections of compounds. Due to an ever increasing number of structures solved and stored in the Protein Data Bank (PDB) [39], and also due to the development of methods that accurately probe and map ligand binding pockets in protein structures (GRID [40]; LUDI [41]; SuperStar [42]; Drug-Score [9,43]), the use of pharmacophores in SBDD studies has been remarkably increased [6,44,45]. For the generation of structure-based models, 3D pharmacophore hypotheses (Fig. 2) can be derived by considering the information gathered by the superposition of X-ray crystallographic structures (Fig. 3) and also from the analysis of the requirements imposed by the binding site through the identification of favorable regions for intermolecular interactions [46] (Fig. 4). The 3D pharmacophore queries are used to screen databases of compounds with the assistance of appropriate software, such as UNITY (Chemical Information Software, version 4.1, Tripos), Catalyst [47], or FlexX-Pharm [48], and only those molecules that carry the pharmacophoric features in the 3D space are retrieved and selected for the next VS steps.

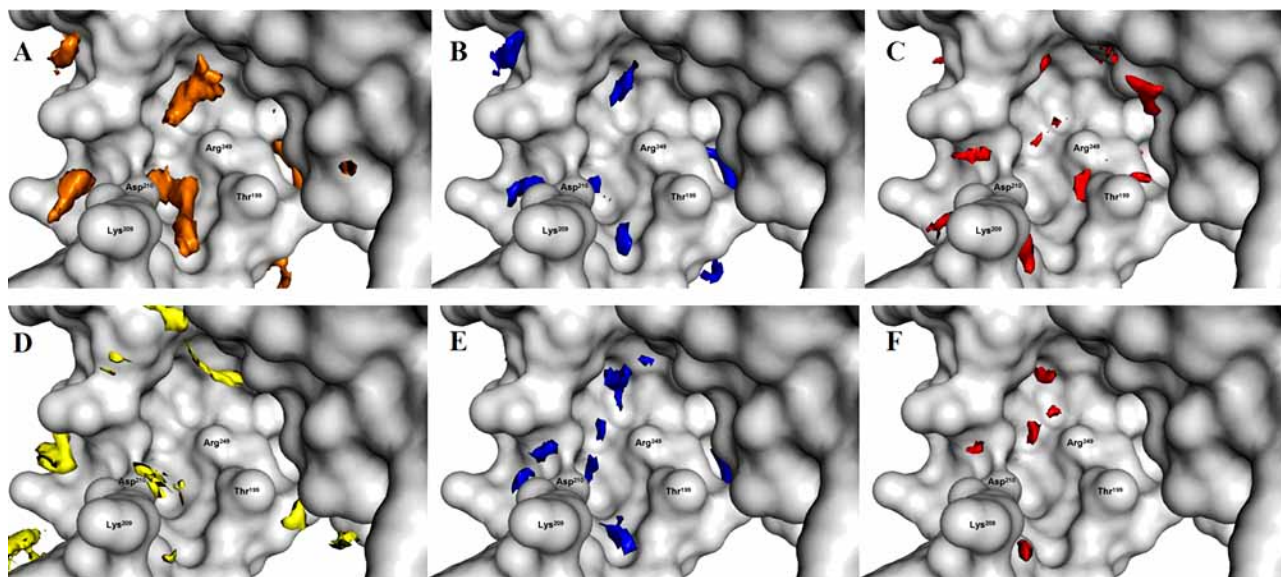
One of the major challenges facing SBVS programs is the selection of an appropriate docking tool. The ligands (orientation and conformation) shall achieve the highest possible degree of



**Fig. (2).** Representation of a structure-based 3D pharmacophore model. Acceptor-site features of the protein and the complementary corresponding partner donor features are shown in blue; donor-site features of the protein and the complementary corresponding partner acceptor features are shown in red; and, the donor/acceptor features and the corresponding donor/acceptor site features of the threonine residues are shown in magenta.



**Fig. (3).** Structural superposition of three crystal structures of *T. cruzi* GAPDH in complex with an analogue of 1,3-bisphospho-d-glyceric acid (PDB code: 1QXS, in cyan), with an irreversible inhibitor 2-(2-phosphono-ethyl)-acrylic acid 4-nitro-phenyl ester (PDB code: 1ML3, in magenta), and with chalepin, a coumarin derivative inhibitor (PDB code: 1K3T, in green).

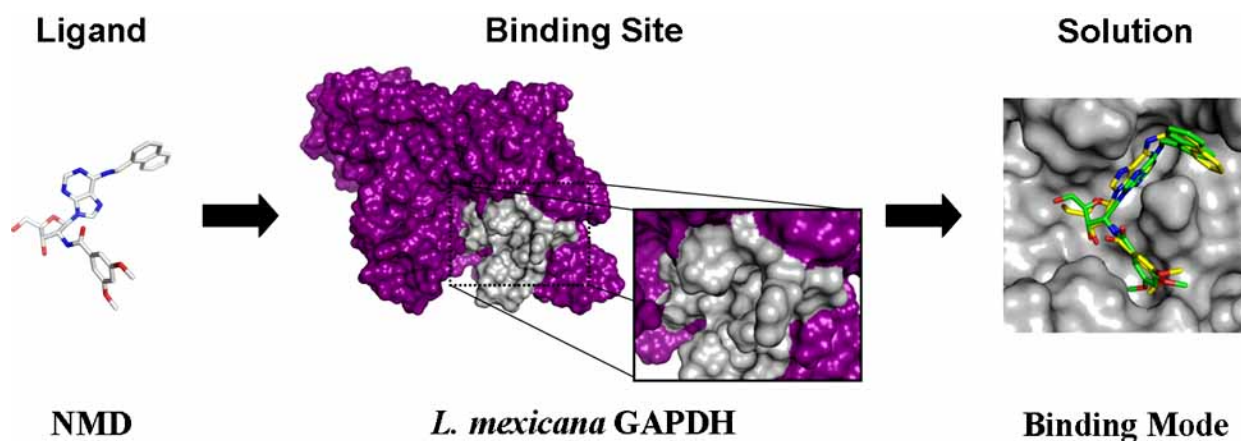


**Fig. (4).** Mapping of putative binding “hot spots” in the active site of *T. cruzi* GAPDH. “Hot spots” indicate areas within the protein binding site where certain functional groups should preferably bind. Panels A and D highlight properties of bulky and hydrophobic groups in a putative ligand: sp<sup>3</sup> carbon (orange) and DRY (yellow) as probes in DrugScore and GRID, respectively. Panels B and E highlight properties of a putative hydrogen-bond donor group in a ligand: amide nitrogen (blue) as probe in DrugScore and GRID. Panels C and F highlight properties of a putative hydrogen-bond acceptor group in a ligand: sp<sup>2</sup> oxygen and carbonyl oxygen (both in red) as probes in DrugScore and GRID, respectively.

molecular complementarity with respect to all binding sites of the receptor active site. Hence, the selection of the docking procedure is a crucial step for the success of the VS process. Currently available docking tools follow slightly different concepts, which make individual programs more suitable for a specific task. It is estimated that there are approximately 30 docking programs available [31]. The docking process involves the prediction of ligand orientation and conformation into the active site, followed by a measure of its fitness into the binding site (Fig. 5).

In general, there are two main complex issues that must be addressed during docking simulations: i. accurate structural modeling; and, ii. correct prediction of activity. Predictions of ligand binding

modes for small molecules in macromolecule binding sites are perhaps the most straightforward step, where, in fact, a considerable degree of success has been achieved. The docking programs employ different search methods (algorithms) to treat ligand flexibility. The search methods are divided into three basic categories: i. Systematic methods, such as incremental construction, conformational search, Hammerhead algorithm and databases, which explore all degrees of freedom in a molecule in order to place ligands (molecular fragments) into active sites of proteins; ii. Random or Stochastic methods, such as Monte Carlo and genetic algorithms (GA). The former generates several ligand configurations into the protein binding site and subsequently score the configurations in a multi-



**Fig. (5).** General docking procedure. Binding mode of the high-affinity selective inhibitor  $N^6$ -(1-naphthalenemethyl)-2'-deoxy-2'-(3,5-dimethoxybenzamido) adenosine (NMD) to the GAPDH from *L. mexicana* (PDB code: 1I32) pocket. NMD minimum energy conformer (in white); NMD crystallographic conformer (in yellow); NMD docking solution (in green).

step procedure, while the latter deals with principles of biological competition and population dynamics to place the ligand in active site; iii. Simulation methods, such as molecular dynamics (MD) and energy minimization, which describe the molecular docking process in more detail. The MD method is based on Newton's equation of motion for an atomic system where atomic forces and masses are used to determine atomic positions over series of very small time steps. Energy minimization procedures are usually employed coupled with other search methods, since only a local energy minimum is reached. For example, minimization steps are carried out after each fragment addition followed by a final minimization before scoring [49].

Docking algorithms are complemented by scoring functions that are designed to predict the biological activity through the evaluation of interactions between compounds and protein active sites. Once the configurations of a system are sampled, the docking programs score them to identify the most likely candidates for the true structure [31,49]. Essentially, three types or classes of scoring functions are currently applied: i. Force field-based, which usually employs molecular mechanics force fields to quantify the sum of energies related to the receptor–ligand interaction energies and the internal ligand energies; ii. Empirical, that is based on the concept that binding energies can be approximated by a sum of individual uncorrelated terms. This type of scoring function is fit to reproduce experimental data, such as binding energies and/or conformations, as a sum of several parameterized functions; iii. Knowledge-based scoring functions, which uses atomic interaction-pair potentials to derive statistical potentials of mean force from large sets of protein–ligand complexes.

Scoring functions implemented in docking programs make different assumptions and simplifications in the evaluation of complexes and do not fully account for a number of physical phenomena that determine molecular recognition. For example, ligand-binding events are driven by a combination of enthalpic and entropic effects, where either entropy or enthalpy can dominate specific interactions. This often presents a conceptual problem for contemporary scoring functions, because most of them are much more focused on capturing energetic than entropic effects [50]. Several scoring functions have been design to reproduce binding energies of protein-ligand systems. Usually, databases of 50-300 protein-ligand complexes are employed for validating the docking programs. Over the past years, a variety of scoring functions have been developed and assessed [11,51,52]. One of the most important parameters during the assessment procedures is the discriminatory power of the applied scoring function for ranking and enriching of

potentially active binders at the top list of the docking solutions. It is, however, a hard issue to address most likely because many physicochemical parameters and the general phenomena involved at the molecular level of the binding process are not yet fully understood. Therefore, these relevant features have not been incorporated into scoring functions in a proper way. It can be illustrated with the knowledge-based scoring function DrugScore [9], which was developed based on crystal structure information from the Cambridge Structural Database (CSD) [53] and protein-ligand complexes from the PDB [39]. This scoring function was employed to validate a test set of 56 crystallography complexes yielding an  $r^2$  value of 0.56 for the correlation between experimental and predicted binding affinities [43]. This indicates that the generation of models possessing improved predictive ability (higher  $r^2$  values) would require the incorporation of a better description of the complex molecular events related to binding affinity.

Scoring function programs have different parameters which rely on distinct atom type schemes and atomic partial charges calculation methods, and have been trained on diverse ligand-protein data sets. As a result, each program returns a particularly different estimate of relative binding affinity, and comparisons are nontrivial. To overcome this limitation, several approaches that assign low ranks to most of inactive compounds while assigning high ranks to most of active compounds have been proposed as alternative methods. One of the most common and employed strategy is to combine estimates from a variety of scoring functions into a single consensus score [20]. Several reports show successful applications of this approach to improve hit rates significantly [54,55]. The impact of consensus scoring strategies in the enrichment of true positives (leads) in SBVS can be explained by the fact that the mean of repeated samplings tends to be closer to the true value than any single sampling, thus, since useful scoring functions perform well, different methods will vote for some of the same actives [55,56]. This process contributes to a better understanding of the chemistry involved in ligand binding and also improves the enrichment of true positives. Consensus scoring is a relatively recent field of research in drug design, with a history of about 10 years. Rapidly, it has become an important tool in the field of *in silico* technologies for drug discovery. However, significant improvements are required in the efficiency of the consensus scoring methods, mainly regarding the full description of the molecular events involved in the prediction of binding affinity. Even though challenging issues such as the consideration of water molecules and protein flexibility are somewhat implemented by some docking tools, they still require substantial development to become more useful in drug design. Re-

ardless of what kind of docking tool is considered, docking and scoring are critical issues in virtual drug screening methods.

### SUCCESSFUL APPLICATIONS OF SBVS

A number of recent successful applications of SBVS can be found in the literature for an impressive variety of drug targets from different therapeutic areas [8,20,24,30,57]. For example, VS studies were performed on a potential antimalarial drug target of the reductase family, an enoyl-acyl carrier protein reductase (ENR), which plays an important role in *Plasmodium falciparum* membrane construction and energy production and does not have a human homolog [58]. In summary, employing a Monte Carlo global energy optimization for flexible ligand docking, a database of 336,600 compounds from the ChemBridge Express Library (San Diego, CA) was fully screened. After applying some ADME filters, 169 compounds with suitable pharmacokinetic properties were retrieved and experimentally tested for their ability to inhibit the ENR activity. In this investigation, 16 inhibitors were identified with three of them having  $IC_{50}$  values in the micromolar range (5, 10 and 25  $\mu M$ , Fig. 6A), which fall close to the inhibitory activity of triclosan, a potent *Pf*-ENR inhibitor ( $IC_{50} = 4 \mu M$ ). In another example, a target protein of the STAT (Signal Transducer and Activator of Transcription) family was studied [59]. Irregular activity of one of the family members (Stat3) contributes to carcinogenesis and tumor progression by up-regulating gene expression and promoting dysregulated growth. Described as a critical step in STAT activation, the dimerization between two Stat3 monomers presents an attractive target to inhibit Stat3 DNA-binding and transcriptional activity. Investigations were conducted with the GLIDE (Grid-based Ligand Docking from Energetics) software [15], which employs an incremental construction algorithm to explore ligand flexibility, for the screening of 150,829 compounds from the National Cancer Institute (NCI) chemical libraries. These studies identified the high-scoring compound NSC 74859 (Fig. 6B), which selectively inhibits Stat3 DNA-binding activity *in vitro* with an  $IC_{50}$  value of 86  $\mu M$  ( $IC_{50}$  values toward Stat1 and Stat5 were  $>300 \mu M$  and 170  $\mu M$ , respectively; no interaction with the Src protein family was observed, no significant effect on Erk1/2 or Shc, and low toxicity to cells with no aberrant Stat3). Furthermore, the compound induces growth inhibition and apoptosis of malignant cells as well as induces human breast tumor regression in xenograft models.

### LIGAND-BASED VIRTUAL SCREENING (LBVS)

Ligand-based methods play an important role in medicinal chemistry and modern hit generation and lead identification. The discovery of a large number of novel biologically active compounds has been possible with the use of LBVS methods, for which no 3D information on the target protein is required [22,25]. In this approach, the screening process is based on the use of active compounds as templates. Ligand-based screening techniques essentially focus on comparative molecular similarity analysis of compounds with known and unknown activity, regardless of the methods or algorithms used. LBVS methods can be divided into three major

classes: i. Similarity searching, including 2D fingerprints (Tanimoto, Cosine, Hamming, Russel-Rao, and Forbes coefficients) [60], rely on the Similar Property Principle. This principle states that molecules that are structurally related are likely to have similar properties, thus, databases of structurally similar molecules which have not been tested for biological activity are likely to have compounds that exhibit the activity of interest; ii. Pharmacophore methods based on the pharmacophore concept, which is commonly defined as a three-dimensional arrangement of molecular features or fragments forming a necessary, but not necessarily sufficient, condition required for binding [61]. These methods involve the identification of the pharmacophoric pattern common to a set of known active compounds, and the use of this pattern in 3D substructure search; and, iii. Machine learning methods, in which a classification rule is developed from a training set containing known active and inactive molecules. For a more comprehensive review of similarity searching, pharmacophore generation and search, machine learning and artificial intelligence approaches, the readers are referred to recent excellent reviews [22,25,62-65].

### SUCCESSFUL APPLICATIONS OF LBVS

There are numerous possible ways of applying LBVS strategies and several examples are available in the literature showing the usefulness of these methods to drug design [22,66]. For example, ligand-based methods were applied for the drug target enzyme 5-lipoxygenase (5-LO), that catalyzes the first steps in the conversion of arachidonic acid into leukotrienes, which has been associated with atherosclerosis, cancer, and osteoporosis [67]. In the search for new bioactive compounds, 43 known 5-LO inhibitors were assembled and used in similarity searching in the speedCATS software (Fig. 7A), which employs chemically advanced template search (CATS) descriptors [68]. Eighteen compounds were selected based on the smallest distance to their query molecules, and their 5-LO inhibitory activities were evaluated in a cell-based assay. Finally, two inhibitors from different structural classes exhibited submicromolar inhibitory activity in an intact PMNL cell assay (Fig. 7A). Another example involves the search for new ligands of the ATP-sensitive potassium channels ( $K_{ATP}$  channels).  $K_{ATP}$  channels couple changes in blood glucose concentrations to insulin secretion in pancreatic  $\beta$ -cells and are considered promising drug targets for the treatment of disorders resulting from excessive insulin release [69]. Accordingly, potassium channel openers (KCOs) comprise chemically heterogeneous classes of compounds and have demonstrated significant potential for the treatment of diabetes. In order to search for new KCOs chemotypes, LBVS studies were carried out employing a pharmacophore model comprising compounds from several structurally diverse classes [69], and a subset of the ZINC database with approximately 65,000 compounds. Pharmacokinetic properties were modeled by molecular interaction fields (MIF) based on Vol-Surf descriptors, which retained about 1,900 compounds with drug-like characteristics. Subsequently, pharmacodynamic properties were computed by GRIND (GRid-INdependent Descriptors) [70], FLAP (Fingerprints for Ligands and Proteins) [71] and TOPP (Triplets Of Pharmacophoric Points) [72] methods, which are based on

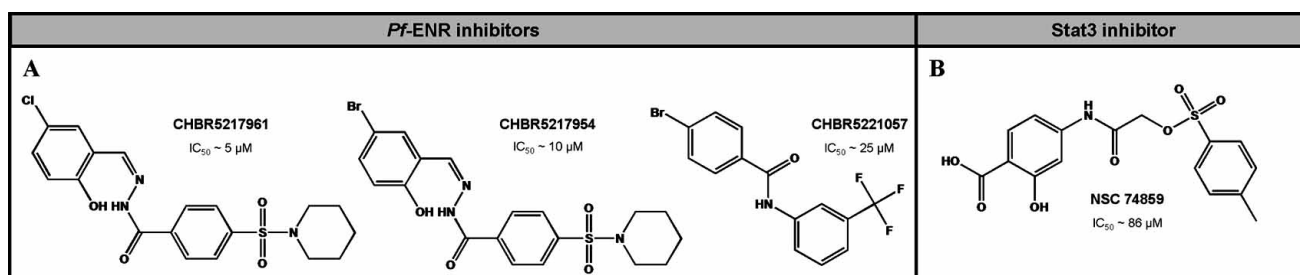


Fig. (6). Inhibitors discovered by SBVS.

MIF and molecular fingerprints. Six potent ligands of the  $K_{ATP}$  channels were used as a template for deriving the pharmacophore hypothesis (Fig. 7B). The top-ranked compounds according to the pharmacophore-based approaches were checked for chemical stability and toxicity by visual inspection, and a final set of 32 molecules were retrieved. The effects on membrane potentials and glucose-stimulated insulin release were assessed by HEK 293 and INS-1 cell assays, respectively, revealing three compounds that were able to inhibit insulin release with micromolar potency (Fig. 7B).

## INTEGRATING VS AND HTS

Modern drug discovery involves the integration of a wide variety of technologies and expertise in multidisciplinary research teams. The influence of the synergy effects between HTS and VS on the selection of hits and lead compounds is a good example of the possible integration of advanced technologies in drug discovery programs. However, it is worth noting that the two approaches are quite distinct in nature and procedures. On one hand, HTS is largely phenomenological and technology-driven, being much more influenced by advances in automation and miniaturization [24,73]. This experimental procedure is costly, complex and labor intensive, with the estimated cost for a basic HTS process, without assay development, being approximately US\$75,000. Alone, biological screening and preclinical pharmacological testing account for about 15% of the total R&D expenditures of the pharmaceutical industry [73]. On the other hand, VS, a knowledge-driven approach, depends on the available information concerning target structure and small-molecule chemical space. As a starting point, either the 3D structure of the macromolecular target or a known active ligand is a prerequisite for VS studies. This computational method offers economy, speed and flexibility to drive drug discovery projects.

These technologies (VS and HTS) are complementary in the sense that they have mutual goals, that is, finding new hits and leads in drug research programs [24,74]. However, an important difference between them is related to the knowledge about the

mechanism of action of ligands upon biological receptors. In the HTS process, no or little information is provided about the possible target binding site or mechanism of action of the selected ligands. In this situation, further structural and kinetic studies would be required to elucidate them. Conversely, VS methods are capable of selecting molecules based on a specific 3D target binding site, therefore, useful information about the binding mode and mechanism of action could also be explored [75]. HTS commonly involves running a primary screen assay (single replicate, single compound concentration) on a large collection of compounds, followed by subsequent rounds of single-shot and dose-response screening. These rapid large-scale assays are monitored by spectroscopic methods such as fluorescence, absorbance and luminescence, being frequently performed in microtiter plates with 96–1536 wells. Since the biochemical assays are designed to screen hundreds of thousands of compounds per run, each plate usually contains internal control wells to ensure quality and comparability of results between plates [76]. Despite the fact that several strategies have been employed to monitor and control the quality and accuracy of the *in vitro* assays, key problems in HTS are the occurrence of false-positive hits (molecules that appear to inhibit the target but turn out to be uninteresting compounds), and the prevalence of nonspecific or promiscuous inhibitors [77]. VS methods are valuable tools designed for the search of large databases of compounds and selection of a reduced number of candidates for biological evaluation. The integration of VS techniques either previously or in connection with HTS methods has three fundamental objectives: i. to extend the scope of the screening to external databases; ii. to identify a larger number and structurally diverse hits; and, iii. to reduce the assay-to-lead attrition rate observed from HTS [78]. VS concepts complement HTS by including compound-filtering techniques based on functions ranging from simple rule-based to more complex neural network architectures. This approach aims to enrich libraries with molecules that have desirable or drug-like properties, eliminating those compounds that have unwanted characteristics for new leads. Alternatively, VS also provides tools for HTS database analysis, being capable of extracting useful information for database mining.

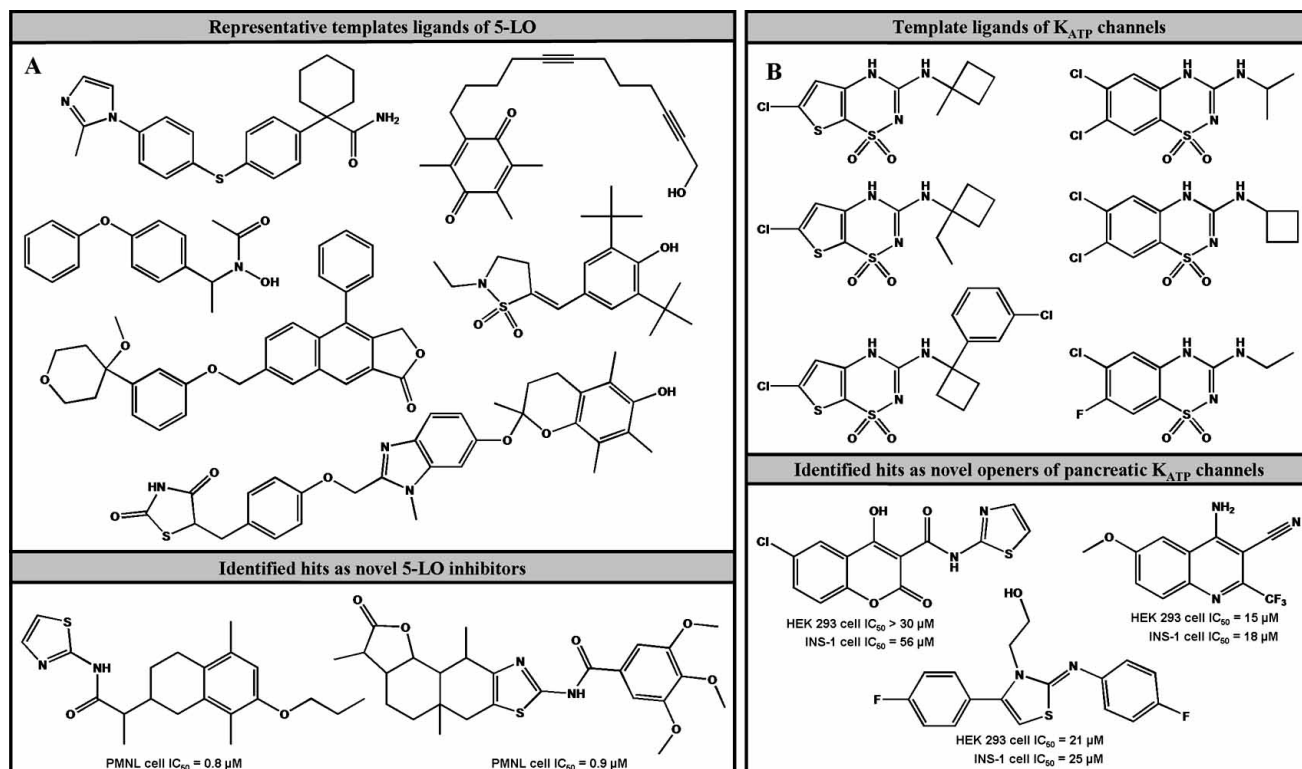


Fig. (7). Inhibitors discovered by LBVS strategies in medicinal chemistry research programs.





results [22]. It is important to emphasize that, in contrast to HTS, VS methods are more easily accessible to academic laboratories around the world, without the need for expensive materials, equipments, and complex infrastructure that an empirical screen demands.

## INTEGRATING VS AND QSAR

Another important example of integration of modern technologies in drug discovery is the combination of VS and QSAR (*Quantitative Structure-Activity Relationships*). QSAR techniques generate descriptors based on molecular structures and uses computational algorithms to relate the key descriptors to the dependent property value of interest [80-85]. Medicinal chemistry studies aimed at elucidating fundamental aspects of the relationships between structural or property descriptors and biological activity are important in the understanding of the activity of interest and may enable the prediction of the biological property for new compounds [86-90]. QSAR has a long history in the drug discovery field, and reached a tremendous impact in the optimization of promising leads that act on specific targets. The availability of advanced molecular modeling techniques and several 2D and 3D QSAR methods has attracted the attention of many scientists around the world for the integration of computational drug design tools. Aiming at overcoming the inherent limitations, the integration of VS and QSAR strategies provides useful opportunities to partially fulfill each method limitation, as well as allows the capture and incorporation of valuable information for the design of new small-molecule drug candidates. A variety of studies describing the integration of these drug design techniques has been reported in the literature [91-96]. For example, investigations were conducted for the discovery of inhibitors of 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) [91], through the integration of 3D QSAR CoMFA (Comparative Molecular Field Analysis) models [97] and FlexE [98] pre-filters based on energy score. As a result, eight novel non-statin-like scaffolds with promising biological activity (Fig. 8A) were identified. Another important example is the development of 3D QSAR CoMSIA (Comparative Molecular Similarity Indices Analysis) models [99] for a series of structurally diverse modulators of the nonsteroidal progesterone receptor (PR) [92]. The QSAR models generated were useful for VS of novel selective nonsteroidal modulators as an alternative receptor-specific scoring function for predicting binding affinities. Instead of searching for models capable of predicting the actual biological activity value, QSAR studies were performed to create models that enable assessment of hypothetical docking scores [93]. In this approach, QSAR models would be capable of identifying most probable nonbinders in docking databases, and also be used together with other pre-docking filters. Similarly, MLR (Multiple Linear Regression) models were developed for a series of chemokine receptor (CCR5) modulators, in order to create a useful alternative to filter out dissimilar compounds and to identify novel potent compounds [94].

Models designed for the prediction of new ligands even when no 3D receptor model is available have been described using a combination of QSAR and VS [95]. Owing to the wide application of QSAR methods, a huge number of molecular descriptors is available. Based on that, techniques that avoid the conformational and alignment ambiguities inherent to 3D QSAR methods have also been employed to produce predictive models useful for database mining and VS [94-96]. In this context, 3D and 2D QSAR models (CoMFA [97], VolSurf [100] and Hologram QSAR [101]) were assessed for the prediction of new ligands towards the nuclear receptor aryl hydrocarbon receptor (AhR) [95]. This study showed the robustness of "structural alignment free" methods as an interesting alternative for VS. HQSAR and VolSurf are fast and highly predictive techniques capable of rapidly generating suitable models for VS of large databases of compounds. In addition, QSAR models derived by *k* nearest neighbor (*k*NN) and molecular topological

descriptors were employed for the discovery of novel anticonvulsant agents [96]. In this study, ten QSAR models were used for the analysis of both Maybridge and National Cancer Institute (NCI) databases. Twenty-two compounds with high predicted activity values were selected and evaluated *in vivo*, leading to the discovery of seven new bioactive compounds (Fig. 8B).

## CONCLUSION

Computational methods play a crucial role in modern medicinal chemistry, presenting a unique potential for transforming the early phases of drug research, particularly in terms of time and cost savings. In this scenario, SBVS and LBVS in combination with other techniques, such as HTS and QSAR, are useful synergistic approaches in drug discovery. VS techniques have been widely employed in the design of focused libraries, compound-filtering, database-mining, and for the rapid analysis of large databases subjected to HTS procedures. Similarly, QSAR methods have been employed as useful tools for guiding the selection of compounds for high-throughput analysis. VS strategies are influencing traditional strategies to analyze large databases of chemical compounds in order to identify possible lead candidates with good pharmacodynamic and pharmacokinetic profiles. In summary, it is always important to understand the limits and scope of any computational tool, and to distinguish between those opportunities which are appropriate to apply the VS technology to pursue a new drug candidate.

## ACKNOWLEDGEMENTS

We gratefully acknowledge financial support from the State of São Paulo Research Foundation (FAPESP) and from the National Council for Scientific and Technological Development (CNPq), Brazil.

## REFERENCES

- [1] Hajduk, P.J.; Greer, J. *Nat. Rev. Drug Discov.* **2007**, *6*, 211-219.
- [2] Hopkins, A.L.; Groom, C.R. *Nat. Rev. Drug Discov.* **2002**, *1*, 727-730.
- [3] Overington, J.P.; Al-Lazikani, B.; Hopkins, A.L. *Nat. Rev. Drug Discov.* **2006**, *5*, 993-996.
- [4] Andricopulo, A.D.; Montanari, C.A. *Mini-Rev. Med. Chem.* **2005**, *5*, 585-593.
- [5] Tame, J.R. *J. Comput. Aided. Mol. Des.* **1999**, *13*, 99-108.
- [6] Grünberg, S.; Stubbs, M.T.; Klebe, G. *J. Med. Chem.* **2002**, *45*, 3588-3602.
- [7] Sippl, W. *J. Comput. Aided Mol. Des.* **2002**, *16*, 825-830.
- [8] Klebe, G. *Perspect Drug Discov. Des.* **1998**, *12-14*, 87-104.
- [9] Gohlke, H.; Hendlich, M.; Klebe, G. *J. Mol. Biol.* **2000**, *295*, 337-356.
- [10] Reddy, M.R.; Singh, U.C.; Erion, M.D. *J. Comput. Chem.* **2007**, *28*, 491-494.
- [11] Hartshorn, M.J.; Verdonk, M.L.; Chessari, G.; Brewerton, S.C.; Mooij, W.T.; Mortenson, P.N.; Murray, C.W. *J. Med. Chem.* **2007**, *50*, 726-741.
- [12] Shoichet, B.K.; Kuntz, I.D. *J. Mol. Biol.* **1991**, *22*, 327-346.
- [13] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727-748.
- [14] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470-489.
- [15] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. *J. Med. Chem.* **2004**, *47*, 1739-1749.
- [16] Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pol-lard, W.T.; Banks, J.L. *J. Med. Chem.* **2004**, *47*, 1750-1759.
- [17] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. *J. Comput. Chem.* **1998**, *19*, 1639-1662.
- [18] Jain, A. N. *J. Med. Chem.* **2003**, *46*, 499-511.
- [19] Clark, R.D.; Strizhev, A.; Leonard, J.M.; Blake, J.F.; Matthew, J.B. *J. Mol. Graph. Model.* **2002**, *20*, 281-295.
- [20] Feher, M. *Drug Discov. Today* **2006**, *11*, 421-428.
- [21] Perola, E.; Walters, W.P.; Charifson, P.S. *Proteins* **2004**, *56*, 235-249.
- [22] Lengauer, T.; Lemmen, C.; Rarey, M.; Zimmermann, M. *Drug Discov. Today* **2004**, *9*, 27-34.
- [23] Cummings, M.D.; DesJarlais, R.L.; Gibbs, A.C.; Mohan, V.; Jaeger, E.P. *J. Med. Chem.* **2005**, *48*, 962-976.
- [24] Klebe, G. *Drug Discov. Today* **2006**, *11*, 580-594.
- [25] Eckert, H.; Bajorath, J. *Drug Discov. Today* **2007**, *12*, 225-233.
- [26] Waszkowycz, B. *Curr. Opin. Drug Discov. Dev.* **2002**, *5*, 407-413.
- [27] Davis, A.M.; Teague, S.J.; Kleywegt, G.J. *Angew. Chem. Int. Ed.* **2003**, *42*, 2718-2736.

- [28] Kasimova, M.R.; Kristensen, S.M.; Howe, P.W.A.; Christensen, T.; Matthiesen, F.; Petersen, J.; Sorensen, H.H.; Led, J.J. *J. Mol. Biol.* **2002**, *318*, 679-695.
- [29] Oshiro, C.; Bradley, E.K.; Eksterowicz, J.; Evensen, E.; Lamb, M.L.; Lantot, J.K.; Putta, S.; Stanton, R.; Grootenhuys, P.D.J. *J. Med. Chem.* **2004**, *47*, 764-767.
- [30] Irwin, J.J.; Shoichet, B.K. *J. Chem. Inf. Model.* **2005**, *45*, 177-182.
- [31] Leach, A.R.; Shoichet, B.K.; Peishoff, C.E. *J. Med. Chem.* **2006**, *49*, 5851-5855.
- [32] Skarzynski, T.; Thorpe, J. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 102-107.
- [33] Erlanson, D.A.; McDowell, R.S.; O'Brien, T. *J. Med. Chem.* **2004**, *47*, 3463-3482.
- [34] Rees, D.C.; Congreve, M.; Murray, C.W.; Carr, R. *Nat. Rev. Drug Discov.* **2004**, *3*, 660-672.
- [35] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3-26.
- [36] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. *Drug Discov. Today* **2003**, *8*, 876-877.
- [37] Hann, M.M.; Oprea, T.I. *Curr. Opin. Chem. Biol.* **2004**, *8*, 255-263.
- [38] Oprea, T.I.; Allu, T.K.; Fara, D.C.; Rad, R.F.; Ostopovici, L.; Bologna, C.G. *J. Comput. Aided Mol. Des.* **2007**, *21*, 113-119.
- [39] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. *Nucleic Acids Res.* **2000**, *28*, 235-242.
- [40] Goodford, P.J. *J. Med. Chem.* **1985**, *28*, 849-857.
- [41] Böhm, H.J. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593-606.
- [42] Verdonk, M.L.; Cole, J.C.; Taylor, R. *J. Mol. Biol.* **1999**, *289*, 1093-1108.
- [43] Gohlke, H.; Hendlich, M.; Klebe, G. *Perspect. Drug Discov. Des.* **2000**, *20*, 115-144.
- [44] Brenk, R.; Naerum, L.; Gradler, U.; Gerber, H.D.; Garcia, G.A.; Reuter, K.; Stubbs, M.T.; Klebe, G. *J. Med. Chem.* **2003**, *46*, 1133-1143.
- [45] Evers, A.; Klebe, G. *J. Med. Chem.* **2004**, *47*, 5381-5392.
- [46] Sottriffer, C.; Klebe, G. *Farmaco* **2002**, *57*, 243-251.
- [47] Kurogi, Y.; Güner, O.F. *Curr. Med. Chem.* **2001**, *8*, 1035-1055.
- [48] Hindle, S.A.; Rarey, M.; Buning, C.; Lengauer, T. *J. Comput. Aided Mol. Des.* **2002**, *16*, 129-149.
- [49] Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.
- [50] Salaniwal, S.; Manas, E.S.; Alvarez, J.C.; Unwalla, R.J. *Proteins* **2007**, *66*, 422-435.
- [51] Stahl, M.; Rarey, M. *J. Med. Chem.* **2001**, *44*, 1035-1042.
- [52] Wang, R.; Lu, Y.; Wang, S. *J. Med. Chem.* **2003**, *46*, 2287-2303.
- [53] Allen, F.H. *Acta Crystallogr.* **2002**, *B58*, 380-388.
- [54] Jenkins, J.L.; Kao, R.Y.; Shapiro, R. *Proteins* **2003**, *50*, 81-93.
- [55] Baber, J.C.; Shirley, W.A.; Gao, Y.; Feher, M. *J. Chem. Inf. Model.* **2006**, *46*, 277-288.
- [56] Wang, R.; Wang, S. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422-1426.
- [57] Kubinyi, H. *Success Stories of Computer-Aided Design, Computer Applications in Pharmaceutical Research and Development.*; John Wiley & Sons: New York, **2006**, pp. 377-424.
- [58] Nicola, G.; Smith, C.A.; Lucumi, E.; Kuo, M.R.; Karagyozov, L.; Fidock, D.A.; Sacchettini, J.C.; Abagyan, R. *Biochem. Biophys. Res. Commun.* **2007**, *358*, 686-691.
- [59] Siddiquee, K.; Zhang, S.; Guida, W.C.; Blaskovich, M.A.; Greedy, B.; Lawrence, H.R.; Yip, M.L.; Jove, R.; McLaughlin, M.M.; Lawrence, N.J.; Sebt, S.M.; Turkson, J. *Proc. Natl. Acad. Sci. USA.* **2007**, *104*, 7391-7396.
- [60] Johnson, M.A.; Maggiora, G.M. *Concepts and Applications of Molecular Similarity.*; John Wiley & Sons: New York, **1990**, pp. 239-288.
- [61] Ghose, A.K.; Wendoloski, J.J. *Perspect. Drug Discov. Des.* **1998**, *9-11*, 253-271.
- [62] Patel, Y.; Gillet, V.J.; Bravi, G.; Leach, A.R. *J. Comput. Aided Mol. Des.* **2002**, *16*, 653-681.
- [63] Willett, P. *Drug Discov. Today.* **2006**, *11*, 1046-1053.
- [64] Duch, W.; Swaminathan, K.; Meller, J. *Curr. Pharm. Des.* **2007**, *13*, 1497-1508.
- [65] Khedkar, S.A.; Malde, A.K.; Coutinho, E.C.; Srivastava, S. *Med. Chem.* **2007**, *3*, 187-197.
- [66] Hert, J.; Willett, P.; Wilton, D.J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. *J. Chem. Inf. Model.* **2006**, *46*, 462-470.
- [67] Franke, L.; Schwarz, O.; Müller-Kuhrt, L.; Hoernig, C.; Fischer, L.; George, S.; Tanrikulu, Y.; Schneider, P.; Werz, O.; Steinhilber, D.; Schneider, G. *J. Med. Chem.* **2007**, *50*, 2640-2646.
- [68] Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, G. *J. Comput. Aided Mol. Des.* **2003**, *17*, 687-698.
- [69] Carosati, E.; Mannhold, R.; Wahl, P.; Hansen, J.B.; Fremming, T.; Zamora, I.; Cianchetta, G.; Baroni, M. *J. Med. Chem.* **2007**, *50*, 2117-2126.
- [70] Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. *J. Med. Chem.* **2000**, *43*, 3233-3243.
- [71] Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J.S. *J. Chem. Inf. Model.* **2007**, *47*, 279-294.
- [72] Sciabola, S.; Morao, I.; de Groot, M.J. *J. Chem. Inf. Model.* **2007**, *47*, 76-84.
- [73] Bajorath, J. *Nat. Rev. Drug Discov.* **2002**, *11*, 882-894.
- [74] Mestres, J. *Biochem. Soc. Transact.* **2002**, *30*, 797-799.
- [75] Klebe, G.; Alvarez, J.; Shoichet, B. *Virtual Screening in Drug Discovery.*; Taylor & Francis: New York, **2005**; pp. 3-4.
- [76] Wu, G.; Doberstein, S.K. *Drug Discov. Today.* **2006**, *11*, 718-724.
- [77] Harper, G.; Pickett, S.D. *Drug Discov. Today.* **2006**, *11*, 694-699.
- [78] Stahura, F.L.; Bajorath, J. *Comb. Chem. High Throughput Screen.* **2004**, *7*, 259-269.
- [79] Parker, C.N.; Shamu, C.E.; Kraybill, B.; Austin, C.P.; Bajorath, J. *Drug Discov. Today* **2006**, *11*, 863-865.
- [80] Honorio, K.M.; Garratt, R.C.; Andricopulo, A.D. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3119-3125.
- [81] Andricopulo, A.D.; Yunes, R.A. *Chem. Pharm. Bull.* **2001**, *49*, 10-17.
- [82] Castilho, M.S.; Postigo, M.P.; de Paula C.B.V.; Montanari, C.A.; Oliva, G.; Andricopulo, A.D. *Bioorg. Med. Chem.* **2006**, *14*, 516-527.
- [83] Leitão, A.; Andricopulo, A.D.; Oliva, G.; Pupo, M.T.; de Marchi, A.A.; Vieira, P.C.; da Silva, M.F.G.F.; Ferreira, V.F.; de Souza, M.C.B.V.; Sa, M.M.; Moraes, V.R.S.; Montanari, C.A. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2199-2204.
- [83] Krogh, R.; Yunes, R.A.; Andricopulo, A.D. *Farmaco* **2000**, *55*, 730-735.
- [84] Andricopulo, A.D.; Yunes, R.A.; Cechinel, V.; Nunes, R.J.; Frazer, J.W.; Cordes, E.H. *Pharmazie* **1999**, *54*, 698-704.
- [85] Castilho, M.S.; Guido, R.V.C.; Andricopulo, A.D. *Letts. Drug Des. Discov.* **2007**, *4*, 106-113.
- [86] Andricopulo, A.D.; Akoachere, M.B.; Krogh, R.; Nickel, C.; McLeish, M.J.; Kenyon, G.L.; Arcott, L.D.; Williams, C.H.; Davioud-Charvet, E.; Becker, K. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 2283-2292.
- [87] Andricopulo, A.D.; Willain, A.; Correa, R.; Santos, A.R.S.; Nunes, R.J.; Yunes, R.A.; Cechinel, V. *Pharmazie* **1998**, *53*, 493-494.
- [88] Cechinel, V.; Correa, R.; Vaz, Z.; Calixto, J.B.; Nunes, R.J.; Pinheiro, T.R.; Andricopulo, A.D.; Yunes, R.A. *Farmaco* **1998**, *53*, 55-57.
- [89] Andricopulo, A.D.; Yunes, R.A.; Nunes, R.J.; Savi, A.O.S.; Correa, R.; Cruz, A.B.; Cechinel, V. *Quím. Nova* **1998**, *21*, 573-577.
- [90] Lima, E.O.; Queiroz, E.F.; Andricopulo, A.D.; Nunes, R.J.; Yunes, R.A.; Correa, R.; Cechinel, V. *Bol. Soc. Chil. Quím.* **1999**, *44*, 185-189.
- [91] Zhang, Q.Y.; Wan, J.; Xu, X.; Yang, G.F.; Ren, Y.L.; Liu, J.J.; Wang, H.; Guo, Y. *J. Comb. Chem.* **2007**, *9*, 131-138.
- [92] Söderholm, A.A.; Lehtovuori, P.T.; Nyronen, T.H. *J. Med. Chem.* **2006**, *49*, 4261-4268.
- [93] Cherkasov, A.; Ban, F.; Li, Y.; Fallahi, M.; Hammond, G.L. *J. Med. Chem.* **2006**, *49*, 7466-7478.
- [94] Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P.A.; Markopoulos, J.; Igglessi-Markopoulou, O. *J. Comput. Aided Mol. Des.* **2006**, *20*, 83-95.
- [95] Lo Piparo, E.; Koehler, K.; Chana, A.; Benfenati, E. *J. Med. Chem.* **2006**, *49*, 5702-5709.
- [96] Shen, M.; Beguin, C.; Golbraikh, A.; Stables, J.P.; Kohn, H.; Tropsha, A. *J. Med. Chem.* **2004**, *47*, 2356-2364.
- [97] Cramer III, R.D.; Patterson, D.E.; Bunce, J.D. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
- [98] Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.* **2001**, *308*, 377-395.
- [99] Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130-4146.
- [100] Cruciani, G.; Pastor, M.; Guba, W. *Eur. J. Pharm. Sci.* **2000**, *11*, S29-S39.
- [101] Tong, W.; Lewis, D.R.; Perkins, R.; Chen, Y.; Welsh, W.J.; Goddette, D.W.; Heritage, T.W.; Sheehan, D.M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669-677.