

---

# Computer-Based Methods of Inhibitor Prediction

---

Silvana Giuliatti

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/52334>

---

## 1. Introduction

This chapter presents *in silico* approaches used in protein structure prediction and drug discovery research.

The structural and functional diversity of animal toxins are interesting tools for therapeutic drug design. This diversity is also of great interest in the search for natural or synthetic inhibitors against these animal toxins.

Computational techniques are highly important in drug design. They are used in the search for candidate ligands binding to a receptor.

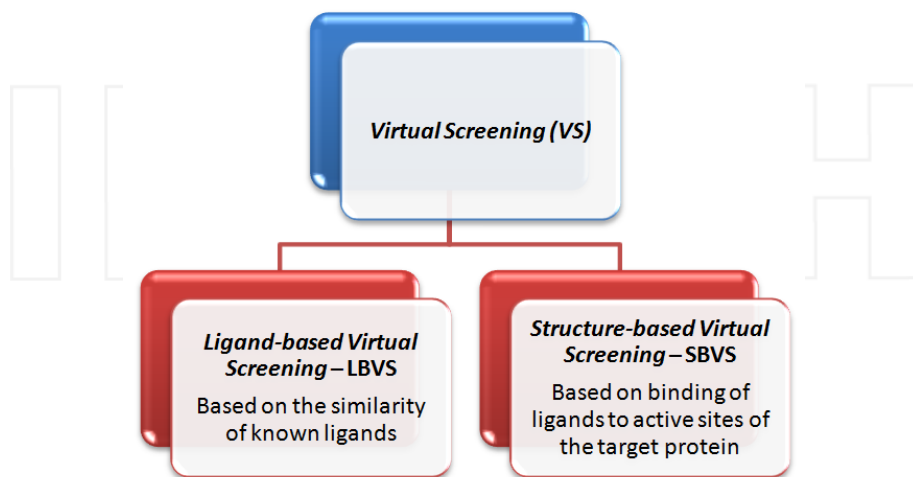
Drug design based on structure has become a highly developed technology and is used in large pharmaceutical companies. Firstly, the structure of the protein of interest must be known. Therefore, molecular modelling plays an important role in the discovery of new drugs.

If the structure of the receptor is known, then the application is essentially a problem of structure-based drug design. These methods have specific goals, such as attempting to identify the location of the active site of the ligand and the geometry of the ligand in the active site. Another goal is to select a number of related binders in terms of affinity or evaluation of the binding free energy.

The strategy of virtual screening has been used to contribute to the increase in hit rate in the selection of new drug candidates.

Virtual screening (VS) is a modern methodology that has been used in the identification of new bioactive substances. It is an *in silico* method that aims to identify small molecules contained in large databases of compounds with high potential for interaction with target proteins for subsequent biochemical analyses.

The strategy of VS can be divided into *ligand-based virtual screening* (LBVS), where a large number of molecules can be evaluated based on the similarity of known ligands, and *structure-based virtual screening* (SBVS), where a number of molecules can be evaluated for specifically binding to the active sites of target proteins (Figure 1).



**Figure 1.** Virtual screening can be divided into ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS).

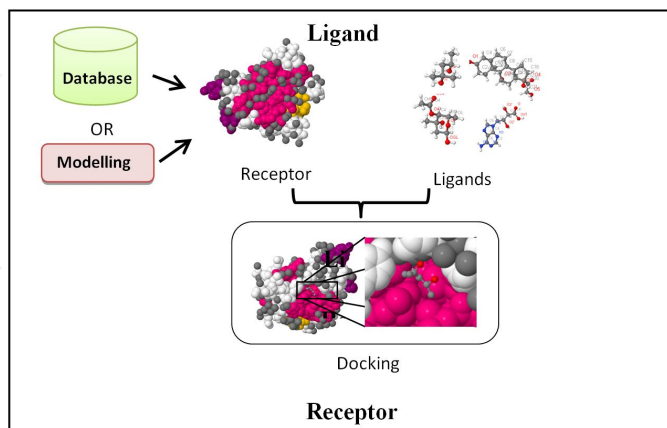
Molecular docking is used to determine the best orientation and conformation of a ligand in its receptor site. The aim is to generate a range of conformations of the protein-ligand complex and sort them according to their scores, which are based on their stabilities. In order to do this, the protein structure and a database of ligands (potential candidates) are used as inputs to the docking software. Thus, large collections of virtual compounds are subjected to docking into a protein-binding site and sorted according to their affinities for the macromolecular target, as suggested by the score function.

The focus of this chapter is to present the strategy of SBVS and the basic concepts of the methodologies involved. Examples of these approaches that have been applied to the identification of animal venom inhibitors have been presented at the end of the chapter.

## 2. Structure-Based Virtual Screening (SBVS)

SBVS involves the evaluation of databases based on the simulation of interactions between the ligands (small molecules) and receptors (target protein). The various steps in the process of SBVS are briefly shown in Figure 2. After obtaining the structure of the receptor and li-

gand, the next step in the process is molecular docking, which involves the coupling of the ligands with the receptor. At this stage, various conformations and orientations are generated and classified according to the score function. The target protein can be obtained from a database or by modelling.



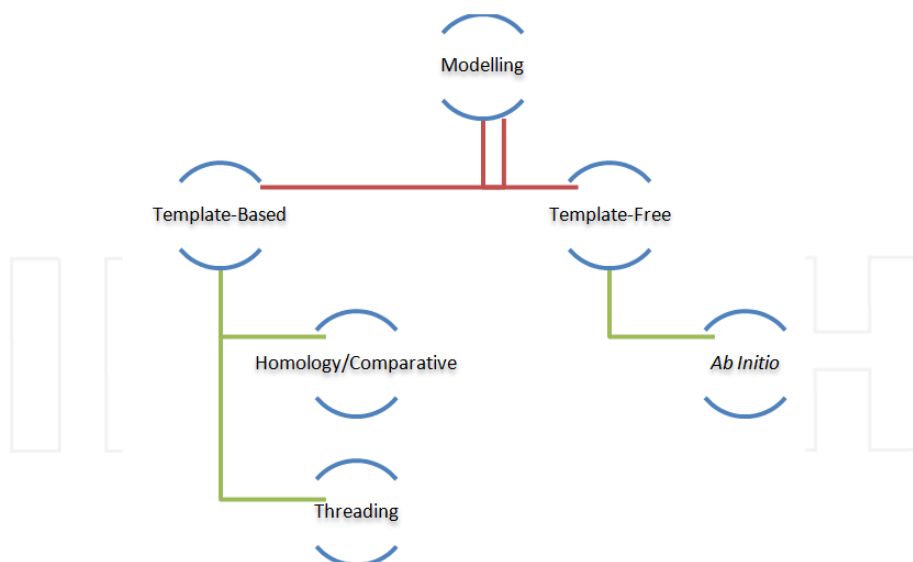
**Figure 2.** Stages of SBVS. The receptor (the target protein) can be obtained from a database or by modelling. Molecular docking completes the structure-based virtual screening.

## 2.1. Obtaining the Structure of the Protein Target

Knowledge of the target protein structure is essential for structure-based drug design. The determination of the 3-dimensional structure of the protein may be achieved experimentally by diffraction of X-rays or by magnetic resonance. If the structure of the target protein has already been solved, it can easily be found deposited in public databases such as PDB [37] which contains more than 80,000 experimentally solved structures.

However, sometimes the structure of the target is not known, and this poses a problem in the drug design process. This situation can be resolved by making use of computational methods for predicting protein structure.

Such methods are divided into 2 groups: those based on templates and those that are template-free. The first group includes comparative or homology modelling and threading. The second group includes methods that do not depend on templates to build the model, such as *ab initio* modelling (Figure 3).



**Figure 3.** Modelling methods can be classified into template-based methods (homology/comparative modelling) and template-free methods (*ab initio*).

### 2.1.1. Template-Based Modelling

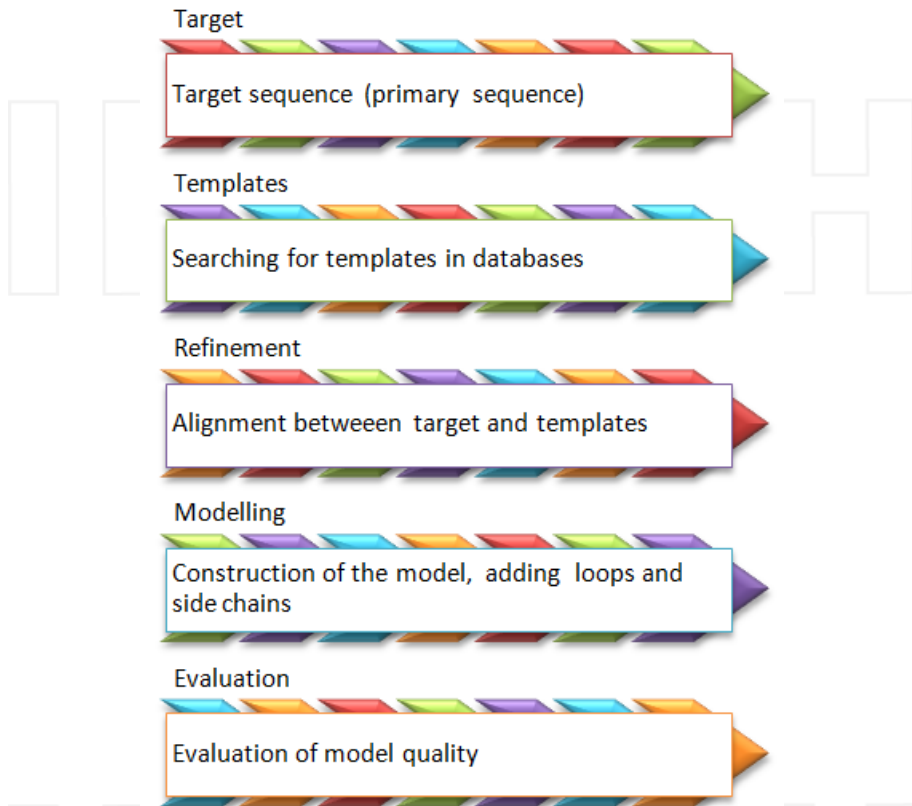
Homology modelling is based on the use of proteins that share an ancestral relationship with the target protein, that is, that they are evolutionarily related and tend to have similar structures. Thus, this method basically involves knowledge of the primary chain of the target protein and a search among databases for homologous proteins that have solved structures. These proteins are used as templates.

Threading modelling is based on the principle that proteins may have similar structures without sharing the same ancestral relationship because the structure tends to be more conserved than the primary sequence. In this case, these methods evaluate the primary chain of the target protein in relation to proteins that have solved structures.

#### 2.1.1.1. Comparative/Homology Modelling

Comparative or homology modelling constructs a model structure of the target protein using its primary chain and the information obtained from homologous proteins that have solved structures. Therefore, this method depends on the availability of proteins that have structures similar to those of the target and can be used as templates. The whole process requires not only the construction of the model, but also the refinement and evaluation of the obtained model. The process can be divided into stages as follows: selection of the templates, which involves the identification of homologous sequences in a database of proteins

that will be used as templates in the modelling process; sequence alignment between the target and the templates; refinement of the alignment; construction of the model, adding loops and side chains; and evaluation of the model (Figure 4).



**Figure 4.** Steps in the comparative modelling process.

The construction of the model depends on the availability of templates. For this purpose, alignment of target and template sequences is widely used and is very efficient. Sequence alignments are typically generated by searching for the result that presents the largest region of identity and similarity. Generally, an identity percentage of at least 25% is considered significant.

There are several tools available for sequence alignment. They differ in the methods used, which can be exhaustive or heuristic, as well as the number of sequences involved in the alignment (multiple or pairwise comparisons). Among these tools, BLAST/PSIBLAST [1; 2]

is a tool that performs local alignments based on the profiles between the target sequence and each sequence belonging to a known database.

The results of the alignment can be evaluated using the E-value. The E-value shows an inverse relationship with the identity/similarity between the sequences. Because it is a heuristic method, the results reported by BLAST are generally suboptimal.

If more than 1 template with similar scores is achieved, the best one can be selected as the template with the higher resolution.

Other methods such as HHpred [34] and Pyre [18] use Markov profiles (Hidden Markov models [HMMs]) combined with structural features.

When more than one template is selected, and taking into account that the results are usually suboptimal, there is a need for an alignment between the target protein and the selected templates. In this case, multiple alignments are indicated. There are several tools that perform multiple alignments, such as ClustalW [21]

After obtaining the alignments between the target and templates, the process of obtaining the model of the target protein begins. There are several software tools available, which differ with respect to the method applied. Prominent among these are MODELLER [9, 33] and SWISS-MODEL [3] The software that has shown the best performance is MODELLER. The program models the backbone using a homology-derived restraint method, which is based on the multiple alignment between the target and templates to differentiate between highly conserved and less conserved residues. The model is optimised by energy minimisation and molecular dynamics methods (Figure 5).



**Figure 5.** The template 3D structures are aligned with the target sequence to be modelled. Spatial features are transferred from the templates to the target and a number of spatial restraints on its structure are obtained. The 3D model is obtained by satisfying all the restraints as thoroughly as possible [33]

The regions of the target that are not aligned with the protein template generally represent loop regions. There are usually some regions caused by insertions and deletions producing gaps in the alignment. Closing these gaps requires modelling of the loops. The loops and the side chains are shaped during the refinement of the model. For this, methods that do not rely on templates can be applied. These include the use of physics parameters and knowledge-based data.

The loops are usually modelled using a database of fragments or by *ab initio* modelling. The use of a database involves finding parts of protein structures known to fit onto 2 regions (stems) of the target protein, which are the regions that precede and follow the loop to be modelled. The conformation of the best matching fragment is used to model the loop.

*Ab initio* methods generate many random loops and look for one that presents a low-energy state and includes conformational angles contained within the allowed regions of the Ramachandran plot [31]. The software CODA [7] can be used for loop modelling.

The side chains can be modelled by programs that make use of libraries of rotamers, such as the software SCRWL4 [20]. The use of rotamer libraries reduces computational time because it reduces the number of favourable torsion angles being examined.

After obtaining the model, its quality must be evaluated. This should be done to make sure that the model has structural features consistent with the physical and chemical rules. Several errors in modelling can occur due to poor choice of template, bad alignment between the target and template, and incorrect determination of loops and side chains.

In the evaluation stage of the model, the structural characteristics as well as the stereochemistry accuracy of the model must be examined.

There are tools available for analysing stereochemical properties, such as PROCHECK [23]. PROCHECK checks the general physicochemical parameters such as phi-psi angles (Ramachandran plot) and chirality. The parameters of the model are compared with those already compiled.

To validate the model for chemical correctness, it is possible to use the software WHAT IF [39]. WHAT IF is a server that checks planarity and bond angles, among other parameters. It also displays the Ramachandran plot.

Verify3D [4, 26] can be used for the analysis of the pseudo-energy profile of the model. It has a database containing environmental profiles based on secondary structures, and the solvent exposure of solved structures at high resolution. It should be noted that the results may be different when different programs are used for verification.

To distinguish correct from incorrect regions, the ERRAT program [6] can be used; this is based on analysis of the characteristics of atomic interactions compared to the highly refined structures.

PROtein Volume Evaluation (PROVE; [30]) calculates the volume of the atoms in the macromolecules using an algorithm that treats the atoms as spheres, analysing the model in relation to the highly resolved and refined structures stored in the PDB.

These software tools are available on servers such as ModFold [27], ProQ (see Section 6 - Table 2), and SAVes (see Section 6 - Table 2).

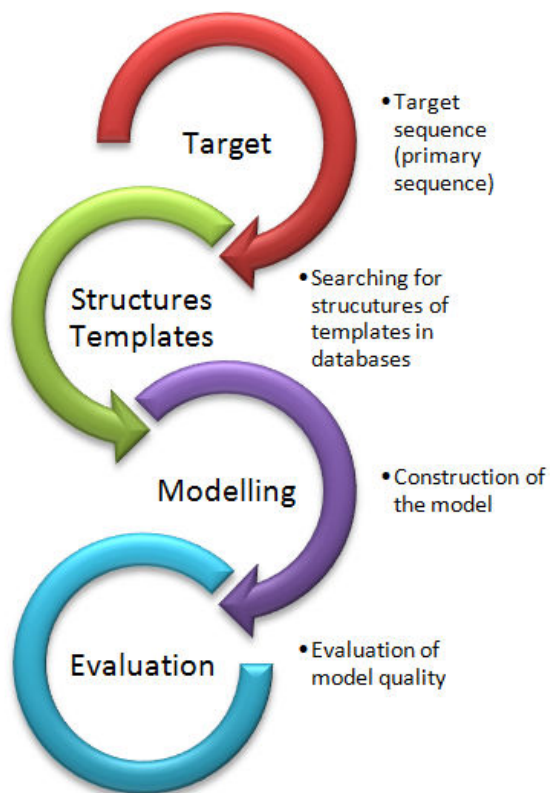
#### 2.1.1.2. Threading

Threading modelling is generally used when the template and target sequences share less than 30% identity. Thus, structures that do not share an evolutionary relationship with the

target protein can be used as templates. However, the target protein has to adopt a fold similar to that of the protein that has had its structure solved. The method can be classified as a pairwise energy-based method.

Using the sequence of the target protein as input, a search is conducted on a database of structures in order to find the best structural match using the criterion of energy calculation. The process is accomplished through a search for solved structures that are most appropriate for the target protein. The comparison highlights secondary structures because they are evolutionarily conserved.

A model is constructed by placing aligned residues between the structure of the template and the target residues. In the next step, the energy of this model is calculated. This is done on various structures in the database. In the end, the models obtained are ranked based on the energy. The model presenting the lowest energy constitutes the most compatible folding model (Figure 6).



**Figure 6.** Steps in the threading modelling process.



Many programs such as THREADER [15, 28] and RAPTOR ([41, 42]) can be used to carry out this process.

### 2.1.2. Template-Free Modelling

One of the biggest problems in comparative modelling is the lack of templates. Template-free methods generate models based on the physicochemical properties and thermodynamic chain of the primary protein target. The processes are iterative. The conformation of the structure is altered until a configuration of lower potential energy is found.

Some methods use force fields based on knowledge as a scoring function. These methods are not strictly free of templates since they employ structures of small fragments of proteins such as, for example, ASTRO-FOLD [19, 35]. Others use energy functions based on first principles of energy and movement of atoms. Generally, these methods involve the calculation of energies of the structures, which has a high computational cost. They are therefore limited to small molecules (approximately 100 residues), as in the case of the software ROSETTA [32].

Firstly, ROSETTA breaks the sequence of the target protein into several short fragments and predicts the secondary structures of the fragments using HMMs. These fragments are then arranged (assembled) into a tertiary setting. Random combinations of these fragments generate a large number of models, which have their energies calculated. The conformation that presents the lowest global energy value is chosen as the best model (Figure 7).

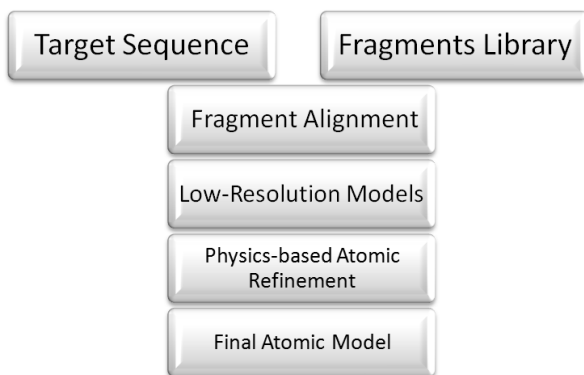


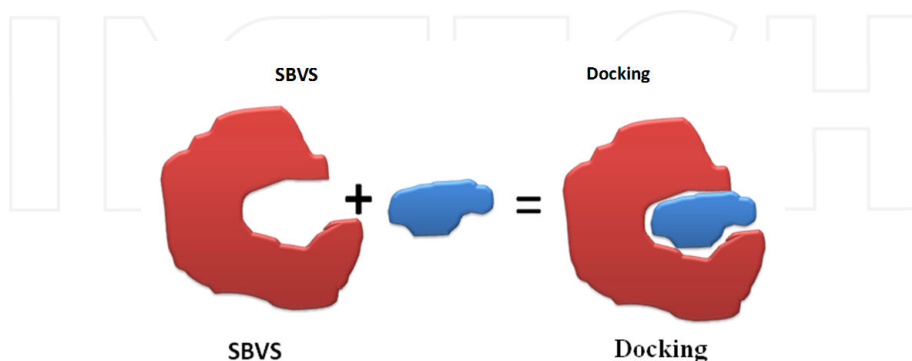
Figure 7. Steps in the ROSETTA process.

## 3. Molecular Docking

One application of molecular docking is virtual screening, in which a library of compounds is compared to one or more targets, thereby providing an analysis of compounds ranked by potential.

Virtual screening computational techniques are applied to the selection of compounds that can be active in a target protein.

In molecular docking, a ligand is usually placed in the binding site of a predetermined structure of a receptor (Figure 8). In other words, this is a method based on structure. The receptor is typically a protein and the ligand is a small molecule or a peptide. The optimal position and orientation of the ligand are determined using a search algorithm and a scoring function that ranks the solutions.



**Figure 8.** Diagram illustrating the docking of a ligand to a receptor to produce a complex.

The first step of the process of molecular docking is to determine the binding sites of the protein. This can be done by software programs such as Q-Sitefinder [24].

The metaPocket method [13] predicts binding sites using 4 methods: LIGSITEcs [12], PASS [5], Q-Sitefinder, and SURFnet [23] – which in combination increase the success rate of prediction. The methods LIGSITEcs, PASS, and SURFnet use only the geometrical characteristics of the protein structure, detecting regions that have the potential to be binding sites. Such methods do not require prior knowledge of the ligands.

In Q-Sitefinder, the surface of the protein is covered with a layer of methyl probes for the calculation of Van der Waals interactions between the protein and the probe. Probes with favourable interaction energies are retained, and are classified into groups based on the number of probes per group. The largest and most energetically favourable group is ranked first and considered the best potential binding site.

Another step is to define the position of the ligand in the pocket. This can be predicted by molecular docking algorithms.

Several methods have developed different scoring functions and different search methodologies.

The search algorithms have to be able to present different configurations and orientations of the ligand in a short time. Search algorithms, such as those used in molecular dynamics,

Monte Carlo simulations, and genetic algorithms, among others, are all suitable for molecular docking.

Scoring functions must be able to discriminate between different ligand-receptor interactions. These can be grouped into field-force, empirical, and knowledge-based methods.

The algorithms can be classified into rigid body docking and flexible docking algorithms. In rigid-body docking, both the ligand and receptor are rigid. These methods are faster, but do not allow ligand and receptor to adapt to the binding. In flexible methods, the computational cost is higher compared to rigid methods. However, in these cases, the flexibility of the ligand and/or receptor is considered.

Another important factor to be considered in ligand-receptor interactions is the presence of water. Some methods allow water molecules to be positioned. In cases where this is not possible, the position of water molecules can be predicted using a software program such as GRID [17].

GRID calculates the interactions between chemical groups and small molecules with known 3-dimensional structures. The energies are calculated using Lennard-Jones interactions, electrostatic and hydrogen bonding between the compounds, and 3-dimensional structures, using a position-dependent dielectric function.

Examples of tools available for docking proteins include AUTODOCK4.2 [29], GOLD [16], and GLIDE [10].

GOLD uses a genetic algorithm that seeks solutions through docking that propagates multiple copies of flexible models of the ligand in the active site of the receptor and recombining segments of copies at random until a converged set of structures is generated.

The process of searching the databases can be time consuming; a way to reduce the search space is filtering databases by performing a search with the fastest algorithms, selecting the best candidates ranked. Subsequently, within this selection, a search algorithm slowly generates a new ranking of the ligands. Another way to reduce the number of ligands being studied in the database is to perform a search for ligands that offer the greatest possibility of being used in drug design. In this case, it is possible to filter the database by using the AD-MET (absorption, distribution, metabolism, excretion, and toxicity) filter.

Lipinski's rule of 5 [25] can be used. The rule of 5 is a set of properties that characterise compounds that exhibit good oral bioavailability. It states that, in general, an orally active drug has no more than 1 violation of the rules (Table 1):

Lipinski's Rule
Not more than 5 hydrogen bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms)
Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms)
A molecular mass less than 500 daltons
An octanol-water partition coefficient log P not greater than 5

Table 1. Lipinski's Rule of Five

Analysis of the metabolic fate and chemical toxicity of the compounds can be accomplished using the software programs DEREK and METEOR [11]. DEREK predicts whether a given chemical is toxic to humans, mammals, and bacteria. METEOR uses the knowledge of metabolism rules to predict the metabolic fate of chemicals, assisting in the choice of more efficient molecules.

#### 4. Ligand-Based Virtual Screening (LBVS)

Other methods can also be used for screening databases of compounds, such as those based on ligands (LBSV). In this case, a similarity search can be made between known bioactive compounds and molecules contained in databases. LBVS techniques include methods based on the pharmacophore and quantitative structure-activity relationship (QSAR) modelling.

In pharmacophore-based virtual screening, a hypothetical pharmacophore is taken as a template. The goal of screening is to identify molecules that show chemical similarities to the template [40].

QSAR is based on the similarity between structures. It is a quantitative relationship between a biological activity and the molecular descriptors that are used to predict the activity. QSAR searches for similarities between known ligands and each structure in a database, investigating how the biological activity of the ligands can be correlated to their structural features [8].

#### 5. Examples of Virtual Screening / Molecular Docking in Animal Venom

[38] performed a virtual screening against  $\alpha$ -Cobratoxin. The neurotoxin  $\alpha$ -Cobratoxin (Cbtx), isolated from the venom of the Thai cobra *Naja kaouthia*, causes paralysis by preventing acetylcholine (ACh) binding to nicotinic acetylcholine receptors (nAChRs). A search for  $\alpha$ -Cobratoxin structures was carried out in the PDB, and the virtual screening of 1990 compounds was performed using the program AutoDock. On [<sup>3</sup>H]epibatidine and on [<sup>125</sup>I]  $\alpha$ -bungarotoxin, NSC121865 (compound 23) was most potent in binding with Ac ( $K_d = 16.26$  nM;  $K_d = 36.63$  nM). The results showed that, in clinical applications, NSC121865 would be a very useful potential lead in the development of a new treatment for snakebite victims. This inhibitor can be used for the development of a more potent and specific anti-cobratoxin.

[14] investigated the effects of protease inhibitors, including phenylmethylsulfonyl fluoride (PMSF), benzamidine (BMD), and their derivatives on the activity of recombinant glosshedobin, a snake venom thrombin-like enzyme (SVTLE), from the snake *Gloydius shedaoensis*. The structural model of glosshedobin was built by homology modelling using modelling package MODELLER. The stereochemical quality of the homology model was assessed using the PROCHECK program and the software AutoDock was used to dock inhibitors onto the structural model of glosshedobin. The docking results indicated that the strongest inhibitor, PMSF, bound covalently to the catalytic Ser195.

[36] evaluated the inhibitory effect of 1-(3-dimethylaminopropyl)-1-(4-fluorophenyl)-3-oxo-1,3-dihydroisobenzofuran-5-carbonitrile (DFD) on viper venom-induced haemorrhagic and PLA2 activities. Molecular docking studies of DFD and snake venom metalloproteases (SVMPs) were performed to understand the mechanism of inhibition by DFD, since SVMPs constitute one of the protein groups responsible for venom-induced haemorrhage. The docking results showed that DFD binds to a hydrophobic pocket in SVMPs with the  $K_i$  of  $19.26 \times 10^{-9}$  (kcal/mol) without chelating  $Zn^{2+}$  in the active site.

## 6. Conclusions

*In silico* approaches used in protein structure prediction and in drug discovery research have been presented in this chapter.

Computational methods used in the search for inhibitors play an essential role in the process of discovering new drugs.

The application of protein modelling methods has contributed significantly in cases where the structure of the target protein has not been solved, allowing the SBVS process be completed.

Good results obtained by virtual screening depend on the quality of structures, databases to be scanned, the search algorithms, and scoring functions. Therefore, there must be a good interaction and exchange of information between *in silico* and experimental methods. Careful application of these strategies is necessary for successful drug design.

Table 2 presents a list of software tools and server web sites.

Summary Tools	
PDB	<a href="http://www.rcsb.org/pdb/home/home.do">http://www.rcsb.org/pdb/home/home.do</a>
BLAST	<a href="http://blast.ncbi.nlm.nih.gov/">http://blast.ncbi.nlm.nih.gov/</a>
HHpred	<a href="http://toolkit.tuebingen.mpg.de/hhpred">http://toolkit.tuebingen.mpg.de/hhpred</a>
ClustalW	<a href="http://www.ebi.ac.uk/Tools/msa/clustalw2/">http://www.ebi.ac.uk/Tools/msa/clustalw2/</a>
SWISS-MODEL	<a href="http://swissmodel.expasy.org/">http://swissmodel.expasy.org/</a>
MODELLER	<a href="http://salilab.org/modeller/">http://salilab.org/modeller/</a>
SCRWL4	<a href="http://dunbrack.fccc.edu/scwrl4/">http://dunbrack.fccc.edu/scwrl4/</a>
PROCHECK	<a href="http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/">http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/</a>
WHAT IF	<a href="http://swift.cmbi.ru.nl/whatif/">http://swift.cmbi.ru.nl/whatif/</a>
Verify3D	<a href="http://nihserver.mbi.ucla.edu/Verify_3D/">http://nihserver.mbi.ucla.edu/Verify_3D/</a>
ERRAT	<a href="http://nihserver.mbi.ucla.edu/ERRATv2/">http://nihserver.mbi.ucla.edu/ERRATv2/</a>
PROVE	<a href="http://www.doe-mbi.ucla.edu/Software/PROVE.html">http://www.doe-mbi.ucla.edu/Software/PROVE.html</a>

<b>Summary Tools</b>	
<b>modFold</b>	<a href="https://www.reading.ac.uk/bioinf/ModFOLD/">https://www.reading.ac.uk/bioinf/ModFOLD/</a>
<b>ProQ</b>	<a href="http://www.sbc.su.se/~bjornw/ProQ/ProQ.html">http://www.sbc.su.se/~bjornw/ProQ/ProQ.html</a>
<b>ROSETTA</b>	<a href="http://www.rosettacommons.org/home">http://www.rosettacommons.org/home</a>
<b>Q-sitefinder</b>	<a href="http://www.modelling.leeds.ac.uk/qsitfinder/">http://www.modelling.leeds.ac.uk/qsitfinder/</a>
<b>SAVES</b>	<a href="http://nihserver.mbi.ucla.edu/SAVES/">http://nihserver.mbi.ucla.edu/SAVES/</a>
<b>THREADER</b>	<a href="http://bioinf.cs.ucl.ac.uk/software_downloads/threader/">http://bioinf.cs.ucl.ac.uk/software_downloads/threader/</a>
<b>metaPocket</b>	<a href="http://projects.biotech.tu-dresden.de/metapocket/">http://projects.biotech.tu-dresden.de/metapocket/</a>
<b>PASS</b>	<a href="http://www.ccl.net/cca/software/UNIX/pass/overview.shtml">http://www.ccl.net/cca/software/UNIX/pass/overview.shtml</a>
<b>SURFNET</b>	<a href="http://www.ebi.ac.uk/thornton-srv/software/SURFNET/">http://www.ebi.ac.uk/thornton-srv/software/SURFNET/</a>
<b>AUTODOCK</b>	<a href="http://autodock.scripps.edu/">http://autodock.scripps.edu/</a>
<b>GOLD</b>	<a href="http://www.ccdc.cam.ac.uk/products/life_sciences/gold/">http://www.ccdc.cam.ac.uk/products/life_sciences/gold/</a>
<b>GLIDE</b>	<a href="http://www.schrodinger.com/products/14/5/">http://www.schrodinger.com/products/14/5/</a>
<b>Derek/Meteor</b>	<a href="https://www.lhasalimited.org/">https://www.lhasalimited.org/</a>
<b>Raptorx</b>	<a href="http://raptorx.uchicago.edu/">http://raptorx.uchicago.edu/</a>
<b>RAPTOR</b>	<a href="http://www.bioinformaticsolutions.com/raptor/downloadpricing/freetrial.html">http://www.bioinformaticsolutions.com/raptor/downloadpricing/freetrial.html</a>
<b>Phyre</b>	<a href="http://www.sbg.bio.ic.ac.uk/~phyre/">http://www.sbg.bio.ic.ac.uk/~phyre/</a>
<b>MUSTER</b>	<a href="http://zhanglab.ccmb.med.umich.edu/MUSTER/">http://zhanglab.ccmb.med.umich.edu/MUSTER/</a>
<b>I-TASSER</b>	<a href="http://zhanglab.ccmb.med.umich.edu/I-TASSER/">http://zhanglab.ccmb.med.umich.edu/I-TASSER/</a>

**Table 2.** Software tools and server web sites.

## Acknowledgements

The author would like to thank CAPES-PROEX and CNPq for financial support.

## Author details

Silvana Giuliatti\*

Address all correspondence to: [silvana@fmrp.usp](mailto:silvana@fmrp.usp)

Faculty of Medicine of Ribeirão Preto - University of São Paulo, Brazil

## References

- [1] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. (1997). Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research*, 25(17), September, 3389-3402, 1362-4962.
- [2] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215(3), October, 403-410, 0022-2836.
- [3] Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL Workspace: a Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics*, 22(2), January 2005, 195-201, 1460-2059.
- [4] Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991). A Method to Identify Protein Sequences that Fold into a Known Three-Dimensional Structure. *Science*, 253(5016), July, 164-170, 0036-8075.
- [5] Brady, G., & Stouten, P. (2000). Fast Prediction and Visualization of Protein Binding Pockets with PASS. *Journal of Computer-Aided Molecular Design*, 14(4), May, 383-401, 1573-4951.
- [6] Colovos, C., & Yeates, T. O. (1993). Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Science*, 12(9), September, 1511-1519, 0036-8075.
- [7] Deane, C. M., & Blundell, T. L. (2001). CODA: A Combined Algorithm for Predicting the Structurally Variable Regions of Protein Models. *Protein Science*, 10(3), March, 599-612, 0146-9896 X.
- [8] Ebalunode, J. O., Zheng, W., & Tropsha, A. (2011). Application of QSAR and Shape Pharmacophore Modeling Approaches for Targeted Chemical Library Design. *Methods in Molecular Biology*, 685, 111-133, 1064-3745.
- [9] Eswar, N., Marti-Renom, M. A., Webb, B., Madhusudhan, M. S., Eramian, D., Shen, M., Pieper, U., & Sali, A. (2007). Comparative Protein Structure Modelling With MODELLER. *Current Protocols in Bioinformatics*, 50, (November), unit 2.9.1-2.9.31, 1934-340X.
- [10] Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shaw, D. E., Shelley, M., Perry, J. K., Francis, P., & Shenkin, P. S. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medical Chemistry*, 47(7), March, 1739-1749, 1520-4804.
- [11] Greene, N., Judson, P., Langowski, J., & Marchant, C. A. (1999). Knowledge-based expert Systems for Toxicity and Metabolism Prediction: DEREK, StAR and METEOR. *SAR QSAR Environmental Research*, 10(2-3), 299-313, 0013-9351.

- [12] Huang, B., & Schroeder, M. (2006). LIGSITE<sub>cs</sub>: Prediction of Ligand Binding Sites using the Connolly Surface and Degree of Conservation. *BMC Structural Biology*, 6, September, 19, 1472-6807.
- [13] Huang, B. (2009). MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. *OMICS: A Journal of Integrative Biology*, 13(4), August, 325-330, 1557-8100.
- [14] Jiang, X., Chena, L., Xua, J., & Yanga, Q. (2010). Molecular Mechanism Analysis of Gloydius Shedaoensis Venom Gloschedobin. *International Journal of Biological Macromolecules*, 48(1), January, 129-133, 0141-8130.
- [15] Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). A New approach to Protein Fold Recognition. *Nature*, July, 358, 86-96, 0028-0836.
- [16] Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and Validation of a Genetic Algorithm for Flexible Docking. *Journal of Molecular Biology*, 267(6381), July, 727-748, 0022-2836.
- [17] Kastenholz, M. A., Pastor, M., Cruciani, G., Haaksmas, E. E. J., & Fox, T. (2000). GRID/CPCA: A New Computational Tool to Design Selective Ligands. *Journal of Medical Chemistry*, 43(16), August, 3033-3044, 1520-4804.
- [18] Kelley, L. A., & Stemberg, J. E. (2009). Protein Structure Prediction on the Web: a Case Study using the Phyre Server. *Nature Protocols*, 4(3), February, 363-371, 1754-2189.
- [19] Klepeis, J. L., & Floudas, C. A. (2003). ASTRO-FOLD: A Combinatorial and Global Optimization Framework for Ab Initio prediction of Three-Dimensional Structures of Proteins from the Amino Acid Sequence. *Biophysical Journal*, 85(4), October, 2119-2146, 0006-3495.
- [20] Krivov, G. G., Shapovalov, M. V., & Dunbrack, R. L. (2009). Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins*, 77(4), December, 778-795, 1097-0134.
- [21] Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X Version 2.0. *Bioinformatics*, 23(21), November, 2947-2948, 1460-2059.
- [22] Laskowski, R. (1995). SURFNET: a Program for Visualizing Molecular Surfaces, Cavities and Intermolecular Interactions. *Journal of Molecular Graphics*, 13(5), October, 323-330, 0263-7855.
- [23] Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography*, 26(2), April, 283-291, 1600-5767.



- [24] Laurie, A., & Jackson, R. (2005). Q-SiteFinder: an Energy-based Method for the Prediction of Protein-Ligand Binding Sites. *Bioinformatics*, 21(9), May, 1908-1916, 1046-2059.
- [25] Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Advanced Drug Delivery Reviews*, 46(1-3), March, 3-26, 0016-9409 X.
- [26] Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of Protein Models with Three-Dimensional Profiles. *Nature*, 356(6364), March, 83-85, 0028-0836.
- [27] Mc Guffin, L. J. (2008). The ModFOLD Server for the Quality Assessment of Protein Structural Models. *Bioinformatics*, 24, 586-587, 1460-2059.
- [28] Milleer, R. T., Jones, D. T., & Thornton, J. M. (1996). Protein Fold Recognition by Sequence Threading: Tools and Assessment Techniques. *The FASEB Journal*, 10(1), January, 171-178, 1530-6860.
- [29] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2004). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry*, 30(16), December, 2009, 2785-2791, 0109-6987 X.
- [30] Pontius, J., Richelle, J., & Wodak, S. J. (1996). Deviations from Standard Atomic Volumes as a Quality Measure of Protein Crystal Structures. *Journal of Molecular Biology*, 264(1), November, 121-126, 0022-2836.
- [31] Ramachandran, G. N., Ramakrishnan, C., & Sasisekharan, V. (1963). Stereochemistry of Polypeptide Chain Configurations. *Journal of Molecular Biology*, 7, July, 95-99, 0022-2836.
- [32] Rohl, C. A., Strauss, C. E., Misura, K. M. S., & Baker, D. (2004). Protein Structure Prediction using Rosetta. *Methods Enzymol*, 383, 66-93, 0076-6879.
- [33] Sali, A. E., & Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, 234, 779-815, 0022-2836.
- [34] Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred Interactive Server for Protein Homology Detection and Structure Prediction. *Nucleic Acids Research*, 33(3), December, W244-W248, 1362-4962.
- [35] Subramani, A., Wei, Y., & Floudas, C. A. (2012). ASTRO-FOLD 2.0: An Enhanced Framework for Protein Structure Prediction. *American Institute of Chemical Engineers Journal*, 58(5), May, 1619-1637, 1547-5905.
- [36] Sunitha, K., Hemshekhar, M., Gaonkar, S. L., Santhosh, M. S., Kumar, M. S., Basappa Priya, B. S., Kemparaju, K., Rangappa, K. S., Swamy, S. N., & Girish, K. S. (2011). Neutralization of Hanemorhagic Activity of Viper Venoms by 1-(3-Dimethylamino-

- propyl)-1-(4-Fluorophenyl)-3-Oxo-1, 3-Dihydroisobenzofuran-5-Carbonitrile. *Basic & Clinical Pharmacology & Toxicology*, 109(4), October, 292-299, 1742-7843.
- [37] Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O., & Abola, E. E. (1998). Protein data bank (PDB): a Database of 3D Structural Information of Biological Macromolecules. *Acta Crystallographica*, D54, 1078-1084, 1600-5759.
- [38] Utsintong, M., Talley, T. T., Taylor, P. W., Olson, A. J., & Vajragupta, O. (2009). Virtual Screening Against  $\alpha$ -Cobratoxin. *Journal of Biomolecular Screening*, 14(9), October, 1109-1118, 1087-0571.
- [39] Vriend, G. (1990). WHAT IF: A Molecular Modelling and Drug Design Program. *Journal of Molecular Graphics*, 8(1), March, 52-56, 0263-7855.
- [40] Yang, U. S. (2010). Pharmacophore Modeling and Applications in Drug Discovery: Challenges and Recent Advances. *Drug Discovery Today*, 15(11-12), June, 446-450, 1359-6446.
- [41] Peng, J., & Xu, J. (2010). Low-homology protein threading. *Bioinformatics*, 26, i294-i300, 10-1093.
- [42] Peng, J., & Xu, J. (2011). RaptorX: Exploiting Structure Information for protein alignment by statistical inference. *Proteins: Structure, Function, and Bioinformatics*, 79(S10), 167-171, 10-1002.

INTECH