Bioinformatics Tools for Screening of Antiparasitic Drugs

Walter Filgueira de Azevedo Junior[1,*], Raquel Dias[1], Luis Fernando Saraiva Macedo Timmers[1], Ivani Pauli[1], Rafael Andrade Caceres[1,2] and Milena Botelho Pereira Soares[3,4]

[1]*Faculdade de Biociências, Laboratório de Bioquímica Estrutural, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre – RS, Brazil;* [2]*Programa de Pós-Graduação em Medicina e Ciências da Saúde. Programa de Pós-Graduação em Biologia Celular e Molecular. Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre – RS, Brazil;* [3]*Centro de Pesquisas Gonçalo Moniz, Fundação Oswaldo Cruz. Rua Waldemar Falcão, 121. Candeal 40296-710 - Salvador, BA – Brazil and* [4]*Hospital São Rafael. Av. São Rafael, 2152. São Marcos 41253-190 - Salvador, BA, Brazil*

Abstract: Drug development has become the Holy Grail of many structural bioinformatics groups. The explosion of information about protein structures, ligand-binding affinity, parasite genome projects, and biological activity of millions of molecules opened the possibility to correlate this scattered information in order to generate reliable computational models to predict the likelihood of being able to modulate a target with a small-molecule drug. Computational methods have shown their potential in drug discovery and development allied with *in vitro* and *in vivo* methodologies. The present review discusses the main bioinformatics tools available for drug discovery and development.

Key Words: Virtual screening, protein-ligand, docking, empirical scoring function, protein target, drug development, *in silico* screening.

INTRODUCTION

One of the most important challenges in the Post-genomic Era is the understanding of protein networks and their interaction with small molecule drugs. Until the 1980s, most of the information related to drug mechanisms and drug receptors could fit in few thousands of printed pages [1]. However, with the recent explosion in biological and chemical information, this is no longer true. Sequencing of human and parasite genomes made possible to pursue the identification of a potential portion of a genome that could be targeted by a drug, which is of extreme importance in the early stages of drug development. Based on the analysis of the first version of a human genome, the total number of "druggable" target macromolecules has been estimated to be approximately 8000, of which ~5000 could be hit by small drug-like molecules. More recent estimates revealed only ~3000 protein targets with pharmacological potential [2].

On the other hand, the sequencing of parasite genomes opened the possibility to focus drug development initiatives on protein targets. These initiatives usually involve high throughput screening of large libraries of diverse molecular structures to determine if they inhibit a specific protein target, previously identified in the parasite genome sequencing. A typical success rate for the initial screening of possible drug candidates is 3-6 hit series out of 750,000 compounds. If the three-dimensional structure of the protein target is known, either through X-ray crystallography, Nuclear Magnetic Resonance (NMR) or homology modeling, then the screening process and success rate can be significantly enhanced through docking simulations of the molecular structure into the active site of the protein target [2].

The process of drug development involves several steps, from target identification and screening, hit and lead generation and optimization, preclinical and clinical studies to final registration of a drug [3]. Several methodologies can be employed in order to identify drug-like molecules able to interfere with a protein function, including *in silico*, *in vitro*, and *in vivo* tests. *In silico* methodologies, also known as bioinformatics tools, have increased their contribution in the last decade. Fig. (**1**) illustrates the main steps involved in drug development using bioinformatics tools, showing a general scheme for virtual screening, where we start with identification of a protein target, followed by the use of a database with hundreds of thousands of molecules. This dataset is submitted to docking simulations. The ligand-binding affinity is evaluated. A smaller subset of potential hits is selected based on the binding affinity. The main computational steps involved in these processes are discussed in this review.

SELECTION OF PROTEIN TARGET

Experimentally determined three-dimensional structures of protein targets are of extreme importance for drug development. In the last few decades over 55,000 structures have been deposited in the Protein Data Bank (PDB) [4-8]. In the case of antiparasitic drugs there are several well-established protein targets that had their structures solved, either by X-ray crystallography or NMR methods. Enzymes of shikimate pathway [9-15], protein kinases [16], and purine nucleoside phosphorylases [17-28] are few examples of these protein targets. In addition, many similar protein targets, identified

*Address correspondence to this author at the Av. Ipiranga, 6681, CEP 90619-900, Porto Alegre, Rio Grande do Sul, Brazil; Tel: +55 51 33203500; E-mail: walter@azevedolab.net
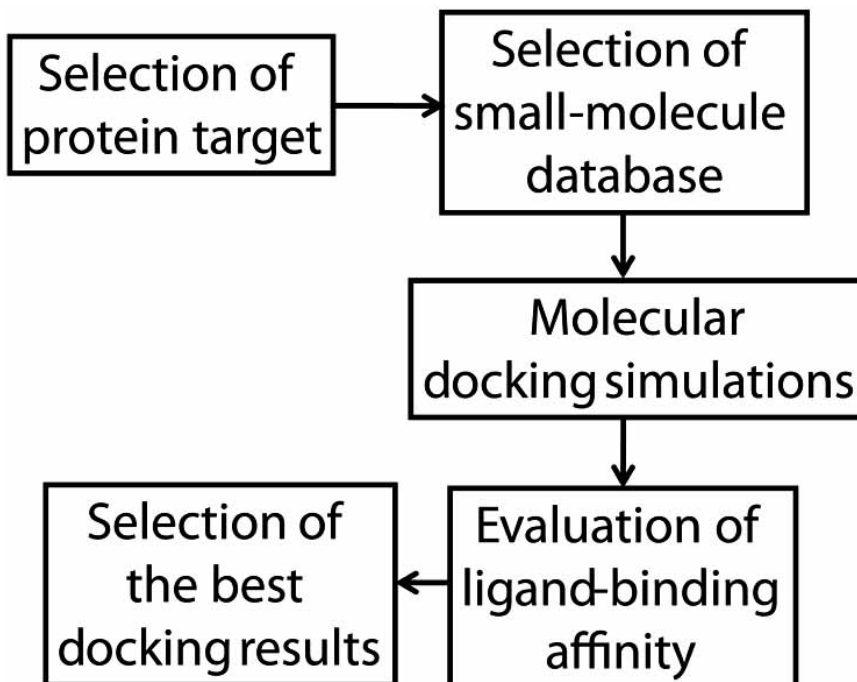
Fig. (1). Main steps involved in drug development using bioinformatics tools.

in parasite genomes may be modeled using experimental determined structures as templates [29-34]. This richness of structural information opens the possibility to engage several virtual screening projects, focused either on specific protein targets or on specific parasites.

SELECTION OF SMALL-MOLECULE DATABASE

The Internet has fast become the first choice for all kind of searches. The increasing quantity of chemistry-related resources now available provides chemists with a direct path to the discovery of information, earlier provided only *via* library services and limited to commercial and costly resources. The diversity of information available online is expanding at a dramatic rate and a shift to publicly available resources offers significant opportunities in terms of the benefit to science and society. We describe here the most used databases available for virtual screening projects.

PubChem is one of the most used small molecule database in virtual screening initiatives. This database is a scientific showcase of the National Institutes of Health (NIH) Roadmap Initiatives (http://nihroadmap.nih.gov). It is a public molecular information repository which was constructed to facilitate information exchange and sharing data deposits among the NIH funded ten Molecular Libraries Screening Centers Network (MLSCN) [35]. The database system is maintained by the National Center for Biotechnological Information (NCBI) and can be accessed on the internet at http://PubChem.ncbi.nlm.nih.gov. The focus of this database is on the chemical, structural and biological properties of small molecules, particularly their application as diagnostic and therapeutic agents.

PubChem is divided into dynamically growing primary databases. First, PubChem Compound database, which has over 10 million entries of pure and characterized chemical compounds. Second, PubChem Substance, with approximately 15.8 million entries of mixtures, extracts, complexes and uncharacterized substances [36]. PubChem continues to grow in stature, content and capability.

Information about molecules deposited in PubChem database and bioactivity data deposited by all MLSCN centers have generated many research opportunities for the scientific community to exploit the available data in PubChem in order to optimize the structure-activity relationships and pharmacology, pharmacokinetics, metabolism and toxicology profiles of target compounds both *in vitro* and *in silico*. Many of the molecules present in this database had antiparasitic activities [35].

The PubChem databases link not only to other Entrez databases such as PubMed and PubMed Central but also to Entrez Structure and Protein, to provide a bridge between the macromolecules of genomics and the small organic molecules of cellular metabolism. They are searchable using, in addition to text queries, structural queries based on chemical Smiles, formulas or 3D chemical structures provided in a variety of formats [37].

PubChem can be searched by alphanumeric text variables, such as names of chemicals, property ranges or by structure, substructure or structural similarity. The system has an important role as a central repository for chemical vendors and content providers enabling evaluation of commercial compound libraries and saving biomedical research-

ers from the work associated with gathering and searching commercial databases.

Another useful database is called eMolecules [38]. It is a free online database that offers suppliers and information for over 10 million unique chemical structures. The database is assembled from data supplied by over 150 suppliers and provides a path to identify a vendor for a particular chemical compound. By providing access to compounds for purchase they are providing a free access online service similar to those of commercial databases, such as Symyx Available Chemical Directory [39], CAS' ChemCats [40] and Cambridgesoft's ChemACX [41] as well as a number of other providers [42]. This database is considered the world's most comprehensive openly accessible search engine for chemical structures.

It is also worth to mention a database called DrugBank, which was designed to serve as a comprehensive, fully searchable *in silico* drug resource that linked sequence, structure and mechanistic data about drug molecules (including antiparasitic drugs) with sequence, structure and mechanistic data about their drug targets [43]. Fundamentally, DrugBank is a dual purpose bioinformatics-chemoinformatics database with a strong focus on quantitative, analytic or molecular-scale information about both drugs and drug targets [1]. As a clinically oriented drug encyclopedia, DrugBank is able to provide detailed, up-to-date, quantitative, analytic or molecular-scale information about drugs, drug targets and the biological or physiological consequences of drug actions. As a chemically oriented drug database, DrugBank is able to provide many built-in tools for viewing, sorting, searching and extracting text, image, sequence or structure data. Since its initial release, DrugBank has been used in a wide range of applications including *in silico* drug discovery [44], drug "rejuvenation" [45], drug docking or screening [46], drug metabolism prediction [47], drug target prediction [48], and general pharmaceutical education.

The database is hosted by the University of Alberta, Canada [42], and in the version 1, that was released in 2006, DrugBank presented >4100 drug entries, corresponding to >12 000 different trade names and synonyms. These drug entries were chosen according to the following rules: the molecule must contain more than one type of atom, be non-redundant, have a known chemical structure and be identified as a drug or drug-like molecule by at least one reputable data source [1]. DrugBank also supports an extensive array of visualizing, querying and search options including a structure similarity search tool and an easy-to-use relational data extraction system [1].

The DrugBank team has employed enormous efforts to facilitate the drug discovery and drug development and hoped that DrugBank will serve as a useful resource to not only members of the pharmaceutical research community but to educators, students, clinicians and the general public [43].

Several virtual screening initiatives have employed ZINC database. This database has a library of 727,842 molecules, the number of molecules is growing, each with 3D structure, using catalogs of compounds from vendors (the size of this library continues to grow). The molecules have been assigned biologically relevant protonation states and are anno-

tated with properties such as molecular weight, calculated LogP, and number of rotatable bonds [49].

The users of this database can access it through several criteria, e.g. the molecular weight, polar desolvation, apolar desolvation, H-acceptors, H-donors, rotatable bonds, xLogP, net charge or the user can draw the own molecular structure, using the Java Molecular Editor (JME) [50], and search for similarities. The Database Browser displays molecules in a table containing ZINC registration code, a 2D sketch, purchasing information, and molecular properties such as calculated LogP and number of rotatable bonds, clicking on a vendor's catalog number links to the vendor's e-commerce Website, if available. The following options are also available: (a) download individual molecules or the set of all molecules matched in SMILES, mol2, SDF, and DOCK flexi-base formats, (b) download a table of molecular properties including purchasing information for analysis in a spreadsheet, and (c) create a subset for docking or download [49].

One interesting feature of ZINC is a collection of pre-defined subsets of molecules. These subsets were defined based on physicochemical criteria in order to facilitate virtual screening initiatives. Another useful option is the possibility to create your own subset. This option allows that a useful subset may be used for other group interested in the same problem.

All previously described databases can be easily accessed, and 3D structures may be downloaded. In the case of virtual screening projects focused on the discovery of antiparasitic drugs all databases are useful, especially ZINC, which can be used to build specific subset of molecules to be used in molecular docking simulations.

MOLECULAR DOCKING

Molecular docking is a computer simulation methodology to predict the conformation of a receptor-ligand complex. Each docking algorithm makes use of one or more specific search algorithms, which are the methods employed to predict the possible conformations of a binary complex. It can also be defined as a simulation process where a ligand position is estimated in a predicted or pre-defined binding site. Molecular docking simulations may be used for reproducing experimental data through docking validations algorithms, where protein-ligand or protein-protein conformations are obtained *in silico* and compared to structures obtained from X-ray crystallography or nuclear magnetic resonance. Furthermore, docking is one of main tools for virtual screening procedures, where a library of several compounds is "docked" against one drug target and returns the best hit. The source of these compounds was described in the previous section.

Speed and accuracy are key features for obtaining a successful result in docking simulations. There are several docking programs, such as DOCK [51], AUTODOCK [52, 53], GOLD [54, 55], FLEXX [56, 57], ZDOCK [58], M-ZDOCK [58], MS-DOCK [60], Surflex [61], MCDOCK [62], and others. The docking program can search for best fit between two or more molecules taking into account several parameters, obtained from receptor and ligand input coordinates as:

geometrical complementarity, regarding atomic VDW radius and charge, receptor or ligand structure flexibility; or considering interatomic interactions, such as hydrogen bonds and hydrophobic contacts. As result, docking applications return the predicted orientations (poses) of a ligand in the target's binding site. Usually the posing process returns several possible conformations. Scoring functions, which are able to evaluate intermolecular binding affinity or binding free energy, are employed in order to optimize and rank results, obtaining the best orientation after the docking procedure.

Docking applications can be classified by their search algorithms, which are defined by a set of rules and parameters applied to predict the conformations. When we consider the flexibility of the ligand and/or the receptor docking algorithms can be classified in two large groups: rigid-body and flexible docking. Rigid-body docking method does not take into account the flexibility of neither ligand nor receptor, limiting the specificity and accuracy of results, considering essentially geometrical complementarities between two molecules. Flexible docking methods can consider several possible conformations of ligand or receptor, as well as for both molecules at the same time, at a higher computational time cost. Docking applications usually make use of one or more of the following methods: fast shape matching (SM) [64], incremental construction (IC) [56, 57], Monte Carlo simulations (MC) [62], distance geometry (DG) [65], evolutionary programming (EP) [66, 67], genetic algorithms (GA) [52, 63], tabu search (TS) [68, 69] and simulated annealing (SA) [70, 71].

One key point in the development of docking algorithms is the accuracy of the docking simulation. The accuracy may vary depending on what target is being tested and what kind of molecules composes the screening library. Highest speed and highest accuracy are ideal, although opposite features for virtual screening through docking simulations. Methods which are more complex, considering many physicochemical and thermodynamic properties, tend to present higher accuracy. However, these methods consume more computational time. Likewise, methods which take into account simpler parameters, as shape matching algorithms, are able to predict docking conformations in fast speed, however at lower accuracy rate.

Generally, docking algorithms predict several orientations (poses) for the ligand inside the binding site. In virtual screening, this process is repeated for a small-molecules database and results are ranked according to the degree of binding affinity between ligand and receptor (*selection of best docking results*) (Fig. (**1**)). Among other factors, the accuracy of docking hits depends on the quality of scoring functions, which are used in result analysis. There are several scoring functions available [72, 73-80]. Recently, many scoring functions have been compared and their accuracy have been tested [81, 82].

Scoring functions are mathematical approximating methods for estimating binding affinity, which is a main tool for lead optimization of virtual screening results, finding the highest-affinity ligand against a target. Its accuracy can be validated in docking refinement tests, if the scoring method is able to exclude docking false positive hits ("decoys"), fit-

ting experimental data [83]. It also has shown to be useful in methods for localization of protein's unidentified binding sites [84-86], prediction of protein-protein binding affinity [87-89] and specific optimizations for proteins from a same structural family also are available [90]. In addition, it has been demonstrated the empirical scoring functions are capable of predicting the best docking results, when compared with a crystallographic structure [90].

The performance of methods for virtual screening has been widely explored, where different docking algorithms have been shown to be successful when combined together in one virtual screening protocol [91], as well as adaptations and tests of known algorithms have been developed in order to optimize virtual screening speed and accuracy [92-94]. Furthermore, many scoring functions have been improved and combined to obtain a correctly ranking of docking results [79, 95, 96]. It indicates that some of the most important sources of improvement for virtual screening are in designing smarter strategies for the setup of virtual screening library or post processing of docking simulations and its results, as well as developing more accurate or specific fast scoring functions.

EVALUATION OF LIGAND-BINDING AFFINITY AND SELECTION OF THE BEST DOCKING RESULTS

The last two steps in the schematic diagram shown in Fig. (**1**) are related to the evaluation of ligand-binding affinity, initially aiming to estimate ligand-binding affinity and finally to select the best docked structure. Computational analysis of protein-ligand systems is one of the open problems in the field of computational chemistry, even with all development reached in the last decade. Specifically, the direct calculation of ligand-binding affinities, using solely information contained in the atomic coordinates of the protein-ligand complexes, is of extreme relevance for the process of scanning virtual libraries of small molecules in order to identify new medications and biological probes. Indeed, the discovery of a new lead compound that binds tightly to a protein is the main objective of early-stage drug discovery and also of chemical genomics projects seeking inhibitors to elucidate gene function [97]. This scenario strongly indicates the need of precise and reliable methodologies to computational evaluate protein-ligand interaction. This computational methodologies would be of little applicability or even useless without experimentally determined affinities to correlate with the predicted affinities. The experimental determined affinities are used as guidelines to calibrate these empirical scoring functions.

Evaluation of ligand-binding affinities can be addressed using empirical scoring functions, based on approximations that use few terms such as intermolecular hydrogen-bonds, van der Waals interaction, deformation effect, hydrophobic interaction, entropy and others. This approach to the study of ligand-binding affinity started with the pioneering work of Böhm [98-101]. Specificity of a ligand for a protein depends on several features, such as intermolecular hydrogen bonds, van der Waals contacts, and shape and charge complementarity between ligand and protein [102-108]. Therefore, assessing intermolecular hydrogen bonds from crystallographic

data is of pivotal importance for determining specificity of ligand against a protein.

The majority of the empirical scoring functions in current use are based on the model where binding affinity can be decomposed in terms that reflect the various contributions to the binding. This hypothesis can be used to build empirical scoring functions to estimate Gibbs free energy of binding ($\Delta G_{binding}$).

In the proposition of new empirical scoring functions it is necessary to use a training set of crystallographic structures of binary complexes of proteins and ligands, where experimental information about the binding affinities is available. The empirical scoring function is built using a training set to obtain weights for each term present in the scoring function. Multivariate regression analysis is then applied to find the best fit between the predicted and experimental protein-binding affinities. Although empirical scoring functions can be interpreted based on physical principles underlying binary complex formation, much of the intermolecular interaction detail is lost.

In spite of many problems in the understanding of the structural features important for binding affinity, most of the experimental available data indicates that additive functions for protein-ligand interactions might be a good approach for the development of empirical scoring functions. These functions may be expressed in Gibbs free energy of binding, as in the PEARLS [109] or in pK_d as used for XSCORE [77] and POLSCORE [80]. With atomic coordinates (x,y,z) available for protein-ligand complexes, the analysis of the binding can be estimated as a sum of interactions multiplied by weighting coefficients (c_j), as indicated by the following equation:

$$pK_d = c_0 + \sum_{j=1}^{N} c_j f_j(x,y,z)$$

where pK_d is the $-\log K_d$ where K_d could also be either K_M or K_i, c_0 is a regression constant, f_j's are functions that accounts for intermolecular interactions.

Evaluation of protein-ligand affinity by means of computational methodologies has demonstrated to be extremely useful for virtual screening projects. Their applications are not limited to the evaluation of protein-ligand affinity only, but also to the identification of best docking results. Application of empirical scoring function to rank best results obtained in molecular docking simulations showed the ability of these empirical functions to predict the correct position for the ligand for several different ligands [28, 81, 82, 90].

Recent developments opened the possibility to test several empirical scoring functions implemented in a single program. In addition, the use of the cross-term empirical scoring function implemented in the program POLSCORE was able to predict the orientation of ligands for complexed structures with results better than a popular empirical scoring function implemented in the program XSCORE [77] and DrugScore [110]. Furthermore, the use empirical scoring function that ranks better for each crystallographic structure, to analyze molecular docking results, indicates that the use of this methodology gives better results than XSCORE and Drug-

Score [110]. One of the weaknesses of empirical scoring methods is the low reliability for unknown systems. Strictly speaking, neither traditional scoring function nor the cross-term function present a direct physical basis. Further developments in the experimental determined protein-ligand affinities through precise methods, such as Isothermal Titration calorimetry (ITC), especially organized in databases [111], may furnish addition experimental information to help in the development of more realistic empirical scoring functions.

FINAL REMARKS

Parasite genome projects are rich sources of biological information that could be employed to identify potential protein targets. Once a protein target is identified the next steps involve massive use of bioinformatics tools to elucidate three-dimensional structure, identify binding pockets, select small molecule databases, and carry out virtual screening. The success of a virtual screening procedure depends not only on the docking simulation capability of posing a ligand inside a given protein binding site, but also on a set of other processes. One of the first steps that are used in most of successful virtual screening procedures consists in a pre processing of virtual screening library.

When there is enough information about a given target binding site properties, such as how it interacts with substrate or which amino acids are the most important for binding and specificity, the setup of a new specific library of molecules can be carried out by filtering a larger library, selecting the molecules that have features that provide a higher chance to bind to the target. The criteria for selecting a wide spectrum different molecule of subsets of compounds lie in the similarity principle. When we consider that structurally-similar small molecules are prone to exhibit similar activity the maximum coverage of the activity space should be reached by choosing a structurally diverse set of molecules. A diverse subset of molecules of a diverse combinatorial library should be more likely to present molecules with a wider spectrum of biological activities and should reduce the number of molecules that are structurally similar and have the same activity, which is called "redundant" molecules [112]. After this first setup, library can be filtered again. It can be carried out by a fast rigid-body docking screening, which may eliminate the geometrically inappropriate ligands by shape complementarities search. Following this strategy, MS-DOCK, a rigid docking tool based on DOCK, proposes a shape complementarity algorithm as a first filtering procedure for a multi-step virtual screening.

Development of a new generation of antiparasitic drugs will benefit greatly from the application of bioinformatics tools, from protein target identification to molecular docking simulations. Post-processing of docking results is also a key factor of virtual screening procedure. Thus, after rigid-body docking search, the remaining compounds can be screened by flexible docking search methods and analyzed with scoring functions. A combination of different scoring functions can be applied in order to improve the accuracy. Although being more CPU-time demanding, this methodology may return results with considerable higher accuracy. Since thousands of compounds must be tested against the target in a short period of time, experimental methodology would be

impracticable and molecular docking combined with binding affinity evaluation has been established as a key method for virtual screening [113-122].

ACKNOWLEDGEMENTS

ABBREVIATIONS

DG	=	Distance geometry
EP	=	Evolutionary programming
GA	=	Genetic algorithms
IC	=	Incremental construction
ITC	=	Isothermal Titration calorimetry
MC	=	Monte Carlo
PDB	=	Protein Data Bank
SA	=	Simulated annealing
SM	=	Shape matching
VDW	=	Van der Waals
TS	=	Tabu search (TS)
$\Delta G_{binding}$	=	Gibbs free energy of binding

REFERENCES

[1] Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z. and Woolsey, J. (2006) *Nucleic Acids Res*., **34**, 668-672.

[2] Schneider, G. and Baringhaus, K.H. *Molecular Design – Creating the Design*, WILLEY-VCH Verlag GmbH & Co. KGaA, Weinheim: Germany, 2008.

[3] Bleicher, K.H., Böhm, H.J., Muller, K., Alanine, A.I. (2003) *Nat. Rev. Drug Discov*., **2**(5), 369-378.

[4] Berman, H.M., Westbrook, J., Bourne, P., Weissig, H., Gilliland, G. and Bhat, T.N. (2000) *Abstracts of Papers of the American Chemical Society*, **219**(U), 112-112.

[5] Westbrook, J.; Feng, Z.; Chen, L.; Yang, H. and Berman, H. M. (2003) *Nucleic Acids Res*., **31**(1), 489-491.

[6] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) *Nucleic Acids Res.*, **28**(1), 235-242.

[7] Westbrook, J.; Ito, N.; Nakamura, H.; Henrick, K. and Berman, H. M. (2005) *Bioinformatics*, **21**(7), 988-992.

[8] Berman, H., Henrick, K., Nakamura, H. and Markley, J.L. (2007) *Nucleic Acids Res.*, **35**(Database issue), D301-303.

[9] Pereira J.H., Canduri F, de Oliveira J.S, da Silveira N.J.F., Basso L.A., Palma M.S., de Azevedo W.F., Santos DS.(2003) *Biochem. Biophys. Res. Commun.,* **312**(3), 608-614.

[10] Pereira, J.H.; Oliveira, J.S.; Canduri, F.; Dias, M.V.B.; Palma, M.S.; Basso, L.A.; Santos, D. S. and De Azevedo Jr., W.F. (2004) *Acta Crystallogr. Sect. D. Biol. Crystallogr*., **60**(Pt 12 Sp. Iss. 2), 2310-2319.

[11] De Azevedo Jr., WF. (2007) *Curr. Drug Targets*, **8**(3), 387-387.

[12] Marques, M.R.; Pereira, J.H.; Oliveira, J.S.; Basso, L. A.; De Azevedo Jr., W. F., Santos, D.S. and Palma, M.S. (2007) *Curr. Drug Targets*, **8**(3), 445-457.

[13] Da Silveira, N.J.; Uchoa, H.B.; Canduri, F.; Pereira, J.H.; Camera Jr., J. C.; Basso, L.A.; Palma, M.S.; Santos, D.S. and De Azevedo Jr., W.F. (2004) *Biochem. Biophys. Res. Commun*., **322**(1), 100-104.

[14] Marques, M.R.; Vaso, A.; Neto, J.R.; Fossey, M.A.; Oliveira, J.S.; Basso, L.A.; dos Santos, D.S.; De Azevedo Jr., W.F. and Palma, M.S. (2008) *Biochemistry*, **47**(28), 7509-7522.

[15] Pauli, I., Caceres, R.A., de Azevedo J.r., W.F.(2008) *Bioorg. Med. Chem.*, **16**(17), 8098-8108.

[16] Canduri, F., Perez, P.C., Caceres, R.A. and de Azevedo, W.F., Jr. (2007) *Curr. Drug Targets*, **8**(3), 389-398.

[17] De Azevedo, W.F., Jr., Canduri, F., Fadel, V., Teodoro, L.G., Hial, V. and Gomes, R.A. (2001) *Biochem. Biophys. Res. Commun.*, **287**(1), 277-281.

[18] De Azevedo, W. F., Canduri, F., Santos, D. M., Silva, R. G., Oliveira, J. S., Carvalho, L.P.S., Basso, L. A., Mendes, M. A., Palma, M.S., and Santos, D.S. (2003) *Biochem. Biophys. Res. Commun.,* **308**(3), 545-552.

[19] Santos, D.M., Canduri, F., Pereira, J.H., Dias, M.V.B, Silva, R.G., Mendes, M. A., Palma, M. S., Basso, L. A., De Azevedo, W. F., and Santos, D. S. (2003) *Biochem. Biophys. Res. Commun.,* **308**(3), 553-559.

[20] De Azevedo, W.F., Canduri, F., Dos Santos, D.M., Pereira, J.H., Dias, M. V.B., Silva, R.G., Mendes, M.A., Basso, L.A., Palma, M.S., and Santos, D.S. (2003) *Biochem. Biophys. Res. Commun.,* **309**(4), 917-922.

[21] De Azevedo, W.F.; Canduri, F.; dos Santos, D.M.; Pereira, J.H.; Dias, M. V. B.; Silva, R. G.; Mendes, M.A.; Basso, L.A.; Palma, M.S. and Santos, D.S. (2003) *Biochem. Biophys. Res. Commun.*, **312**(3), 767-772.

[22] Canduri, F., dos Santos, D.M., Silva, R.G., Mendes, M.A., Basso, L.A., Palma, M.S., De Azevedo, W.F. and Santos, D.S. (2004) *Biochem. Biophys. Res. Commun.,* **313**(4), 907-914.

[23] Canduri, F.; Fadel, V.; Dias, M.V.B.; Basso, L.A.; Palma, M.S.; Santos, D. S. and De Azevedo, W.F. Jr. (2005) *Biochem. Biophys. Res. Commun.*, **326**(2), 335-338.

[24] Canduri, F., Fadel, V., Basso, L.A., Palma, M.S., Santos, D.S., and De Azevedo, W.F. Jr. (2005) *Biochem. Biophys. Res. Commun.,* **327**(3), 646-649.

[25] Canduri, F., Silva, R.G., dos Santos, D.M., Palma, M.S., Basso, L.A., Santos, D.S., De Azevedo, W. F. Jr. (2005) *Acta .Crystallogr. D .Biol .Crystallogr.,* **61**(7), 856-862.

[26] Silva, R.G., Pereira, J.H., Canduri, F., de Azevedo, W.F.J.r., Basso, L.A., Santos, D.S. (2005) *Arch. Biochem. Biophys.,* **442**(1), 49-58.

[27] De Azevedo, Jr., W.F.; Canduri, F.; Basso L.A.; Palma M.S. and Santos, D.S. (2006) *Cell Biochem. Biophys*., **44**(3), 405-411.

[28] Timmers, L.F., Caceres, R.A., Vivan, A.L., Gava, L.M., Dias, R., Ducati, R.G., Basso, L.A., Santos, D.S. and de Azevedo, W.F., Jr. (2008) *Arch .Biochem. Biophys*., **479**, 28-38.

[29] Vivan, A.L., Caceres, R.A., Beltran Abrego, J.R., Borges, J.C., Neto, J.R., Ramos, C.H., de Azevedo, W.F. Jr., Basso, L.A., Santos, D.S. (2008) *Proteins,* **72**(4), 1352-1362.

[30] Arcuri, H.A., Borges, J.C., Fonseca, I.O., Pereira, J.H., Neto, J.R., Basso, L.A., Santos, D.S., de Azevedo, W.F. Jr. (2008) *Proteins,* **72**(2), 720-730.

[31] Da Silveira, N.J.F.; Uchoa, H. B.; Pereira, J.H.; Canduri, F.; Basso, L.A.; Palma, M.S.; Santos, D.S. and De Azevedo, Jr., W.F. (2005) *J. Mol. Model*., **11**(2), 160-166.

[32] Manhani, K.K.; Arcuri, H.A.; Da Silveira, N.J.; Uchôa, H.B.; De Azevedo Jr., W.F. and Canduri, F. (2005) *J. Mol. Model*., **12**(1), 42-48.

[33] Da Silveira, N.J.F., Bonalumi, C.E., Uchoa, H.B., Pereira, J.H., Canduri, F., de Azevedo Jr., W. F. (2006) *Cell. Biochem. and Biophy.,* **44**(3), 366-374.

[34] Uchoa, H. B., Jorge, G. E., da Silveira, N.J., Camera, J.C., Canduri, F., and De Azevedo, W.F. (2004) *Biochem. Biophys. Res. Commun.,* **325**(4), 1481-1486.

[35] Lazo, J.S. (2006) *Mol. Interv.,* **6**(5), 240-243.

[36] Xie, X.Q. and Chen, J.Z. (2008) *J. Chem. Inf. Model.,* **48**(3), 465-475.

[37] Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Miller, V., Ostell, J., Pruitt, K.D., Schuler, G.D., Shumway, M., Sequeira, E., Sherry, S.T., Sirotkin, K., Souvorov, A. Starchenko, G., Tatusov, R.L.,

Tatusova, T.A., Wagner, L., Yashcenki, E. (2008) *Nucleic Acids Res.*, **36**, 13-21.

[38]   eMolecules Online Service: eMolecules, Del Mar, CA, USA (2008). *http://www.emolecules.com*

[39]   Available Chemical Directory: Santa Clara, California, USA (2008). *http://www.mdli.com/products/experiment/available_chem_dir/index.jsp*

[40]   ChemCats: Chemical Abstract Services, Columbus, OH, USA (2006). *http://www.cas.org/expertise/cascontent/chemcats.html*

[41]   ChemACX: CambridgeSoft Corp, Cambridge, MA, USA (2008). *http://www.cambridgesoft.com/databases/details/?db=12*

[42]   Williams, A.J. (2008) *Curr. Opin. Drug Discov. Devel.*, **11**(3), 393-404.

[43]   Wishart, D.S., Knox, C., Guo, A.C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., Hassanali, M. (2008) *Nucleic Acids Res.,* **36**, D901-906.

[44]   Chang, C., Bahadduri, P.M., Polli, J.E., Swaan, P.W., Ekins, S. (2006) *Drug Metab. Dispos.*, **34**(12), 1976-1984.

[45]   Chong, C.R. and Sullivan D.J. Jr (2007) *Nature*, **448**(7154), 645–646.

[46]   Li, H., Gao, Z., Kang, L., Zhang, H., Yang, K., Yu, K., Luo, X., Zhu, W., Chen, K., Shen J., Wang X., Jiang H. (2006) *Nucleic Acids Res*., **34**, W219-224.

[47]   Jolivette, L.J. and Ekins, S. (2007) *Adv. Clin. Chem*., **43**, 131-176.

[48]   Wishart, D.S. (2007) *Comp. Biochem. Phys*., **2**(1), 9-17.

[49]   Irwin, J.J. and Shoichet, B.K. (2005) *J. Chem. Inf. Model*., **45**(1), 177-182.

[50]   Ertl, P. and Jacob, O. (1997) *J. Mol. Struct*., **419**, 113-120.

[51]   Ewing, T.J., Makino, S., Skillman, A.G. and Kuntz, I.D. (2001) *J. Comput. Aided Mol. Des.*, **15**(5), 411-428.

[52]   Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998) *J. Comput. Chem.*, **19**(14), 1639-1662.

[53]   Goodsell, D.S., Morris, G.M. and Olson, A.J. (1996) *J. Mol. Recognit*., **9**(1), 1-5.

[54]   Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. and Taylor, R.D. (2003) *Proteins*, **52**(4), 609-623.

[55]   Joy, S., Nair, P.S., Hariharan, R. and Pillai, M.R. (2006) *In Silico Biol.*, **6**(6), 601-605.

[56]   Rarey, M., Kramer, B., Lengauer, T. and Klebe, G. (1996) *J. Mol. Biol.*, **261**(3), 470-489.

[57]   Kramer, B., Rarey, M. and Lengauer, T. (1999) *Proteins*, **37**(2), 228-241.

[58]   Chen, R., Li, L. and Weng, Z. (2003) *Proteins*, **52**(1), 80-87.

[59]   Pierce, B., Tong, W. and Weng, Z. (2005) *Bioinformatics*, **21**(8), 1472-1478.

[60]   Sauton, N., Lagorce, D., Villoutreix, B.O. and Miteva, M.A. (2008) *BMC. Bioinform.*, **9**, 184-196.

[61]   Jain, A.N. (2003) *J. Med. Chem.*, **46**(4), 499-511.

[62]   Liu, M. and Wang, S. (1999) *J. Comput. Aided. Mol. Des.*, **13**(5), 435-451.

[63]   Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R. (1997) *J. Mol. Biol.*, **267**(3), 727-748.

[64]   Kuntz, I.D.,Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E. (1982) *J. Mol. Biol.*, **161**(2), 269-288.

[65]   Moré, J.J. and Wu, Z. (1999) *J. Global Optimization*, **15**(3), 219-234.

[66]   Eiben, A.E. and Smith, J.E. (2003) Intro. to Evo. Comput. Springer,

[67]   Fogel, L.J., Owens, A.J. and Walsh, M.J. (1966) Artificial Intelligence through Simulated Evolution. John Wiley.

[68]   Glover, F. and Laguna, M. (1993) Tabu Search. In: Reeves, C.R. Ed, Modern Heuristic Techniques for Combinatorial Problems, John Wiley & Sons.

[69]   Baxter, C.A., Murray, C.W., Clark, D.E., Westhead, D.R. and Eldridge, M.D. (1998) *Proteins*, **33**(3), 367-382.

[70]   Kirkoatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983) *Science*, **220**, 671-680.

[71]   Goodsell, D.S. and Olson, A.J. (1990) *Proteins-Structure. Function Genet.,* **8**(3), 195-202.

[72]   Betzi, S., Suhre, K., Chetrit, B., Guerlesquin, F. and Morelli, X. (2006) *J. Chem.Inf. Model.*, **46**(4), 1704-1712.

[73]   Yang, C.Y., Wang, R. and Wang, S. (2006) *J. Med. Chem.*, **49**(20), 5903-5911.

[74]   Krammer, A., Kirchhoff, P.D., Jiang, X., Venkatachalam, C.M. and Waldman, M. (2005) *J. Mol. Graph Model.*, **23**(5), 395-407.

[75]   Velec, H.F., Gohlke, H. and Klebe, G. (2005) *J. Med. Chem.*, **48**(20), 6296-6303.

[76]   Wang, R.X., Liu, L., Lai, L.H. and Tang, Y.Q. (1998) *J.Mol. Model.*, **4**(12), 379-394.

[77]   Wang, R., Lai, L. and Wang, S. (2002) *J. Comput. Aided Mol. Des*, **16**(1), 11-26.

[78]   Zhao, X., Liu, X., Wang, Y., Chen, Z., Kang, L., Zhang, H., Luo, X., Zhu, W., Chen, K., Li, H., Wang, X. and Jiang, H. (2008) *J. Chem. Inf. Model.*, **48**(7), 1438-1447.

[79]   Muryshev, A.E., Tarasov, D.N., Butygin, A.V., Butygina, O.Y.; Aleksandrov, A.B. and Nikitin, S.M. (2003) *J. Comput. Aided Mol. Des.*, **17**(9), 597-605.

[80]   De Azevedo, W.F. and Dias, R. (2008) *Bioorg. Med. Chem.*, **16**(20), 9378-9382.

[81]   Wang, R.; Lu, Y.; Fang, X. and Wang, S. (2004) *J. Chem. Inf. Comput. Sci.*, **44**(6), 2114-2125.

[82]   Wang, R., Lu, Y. and Wang, S. (2003) *J. Med. Chem.*, **46**(12), 2287-2303.

[83]   Graves, A.P., Brenk, R. and Shoichet, B.K. (2005) *J. Med. Chem.*, **48**(11), 3714-3728.

[84]   Rossi, A., Marti-Renom, M.A. and Sali, A. (2006) *Protein Sci.*, **15**(10), 2366-2380.

[85]   Friberg, M., von Rohr, P. and Gonnet, G. (2005) *BMC Bioinform.*, **6**(84), 1-11.

[86]   Liang, S., Zhang, C., Liu, S. and Zhou, Y. (2006) *Nucleic Acids Res.*, **34**(13), 3698-3707.

[87]   Huang, S.Y. and Zou, X. (2008) *Proteins*, **72**(2), 557-579.

[88]   Bernauer, J., Aze, J., Janin, J. and Poupon, A. (2007) *Bioinformatics*, **23**(5), 555-562.

[89]   Chen, R. and Weng, Z. (2003) *Proteins*, **51**(3), 397-408.

[90]   Caceres, R.A., Saraiva Timmers, L.F., Dias, R., Basso, L.A., Santos, D.S. and de Azevedo, W.F., Jr. (2008) *Bioorg. Med. Chem.*, **16**(9), 4984-4993.

[91]   Miteva, M.A., Lee, W.H., Montes, M.O. and Villoutreix, B.O. (2005) *J. Med. Chem.*, **48**(19), 6012-6022.

[92]   Buzko, O.V.; Bishop, A.C. and Shokat, K.M. (2002) *J. Comput. Aided. Mol. Des.*, **16**(2), 113-127.

[93]   Vaque, M., Arola, A., Aliagas, C. and Pujadas, G. (2006) *Bioinformatics*, **22**(14), 1803-1804.

[94]   Sato, H., Shewchuk, L.M. and Tang, J. (2006) *J. Chem. Inf. Model*, **46**(6), 2552-2562.

[95]   Muller, W. and Sticht, H. (2007) *Proteins*, **67**(1), 98-111.

[96]   Rajamani, R. and Good, A.C. (2007) *Curr. Opin. Drug. Discov. Devel.*, **10**(3), 308-315.

[97]   Gilson, M.K. and Zhou, H.X. (2007) *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 21-42.

[98]   Bohm, H.J. (1994) *J. Comput. Aided Mol. Des.*, **8**(3), 243-256.

[99]   Laurie, A.T. and Jackson, R.M. (2006) *Curr. Protein Pept. Sci.*, **7**(5), 395-406.

[100] Vajda, S. and Guarnieri, F. (2006) *Curr. Opin. Drug Discov. Devel.*, **9**(3), 354-362.

[101] Cozzini, P., Fornabaio, M., Marabotti, A., Abraham, D.J., Kellogg, G.E. and Mozzarelli, A. (2004) *Curr. Med. Chem.*, **11**(23), 3093-3118.

[102] De Azevedo, W.F., Jr., Mueller-Dieckmann, H.J., Schulze-Gahmen, U., Worland, P.J., Sausville, E. and Kim, S.H. (1996) *Proc. Natl. Acad. Sci. U. S. A.*, **93**(7), 2735-2740.

[103] De Azevedo, W.F., Leclerc, S., Meijer, L., Havlicek, L., Strnad, M. and Kim, S.H. (1997) *Eur. J. Biochem.*, **243**(1-2), 518-526.

[104] Dias, M. V., Vasconcelos, I.B., Prado, A.M., Fadel, V., Basso, L.A., De Azevedo, W.F. Jr, Santos, D.S. (2007) *J. Struct. Biol.,* **159**(3), 369-380.

[105] Borges, J.C.; Pereira, J.H.; Vasconcelos, I.B.; dos Santos, G.C.; Olivieri, J.R.; Ramos, C.H.I.; Palma, M.S.; Basso, L.A.; Santos, D.S. and De Azevedo Jr., W.F. (2006) *Arch. Biochem. Biophys.*, **452**(2), 156-164.

[106] Krystof, V.; Cankar, P.; Hajduch, M.; De Azevedo Jr., W. F.; Paprskarova, M.; Orsag, M.; Rolcik, J.; Latr, A.; Fischer, P. M. and Strnad, M. (2006) *J. Med. Chem.*, **49**(22), 6500-6509.

[107] Oliveira, J.S.; Pereira, J.H.; Canduri, F.; Rodrigues, N. C.; de Souza, O. N.; De Azevedo Jr., W.F.; Basso, L.A. and Santos, D. S. (2006) *J. Mol. Biol.*, **359**(3), 646-666.

[108] Dias, M.V.B.; Canduri, F.; da Silveira, N.J.F.; Czekster, C.M.; Basso, L.A.; Palma, M.S.; Santos, D.S. and De Azevedo, Jr. WF. (2006) *Cell Biochem. Biophys.*, **44**(3), 375-384.

[109] Han, L.Y., Lin, H.H., Li, Z.R., Zheng, C.J., Cao, Z.W., Xie, B. and Chen, Y.Z. (2006) *J. Chem. Inf. Model.*, **46**(1), 445-450.

[110] Chrencik, J.E., Staker, B.L., Burgin, A.B., Pourquier, P., Pommier, Y., Stewart, L. and Redinbo, M.R. (2004) *J. Mol. Biol.*, **339**(4), 773-784.

[111] Li, L., Dantzer, J.J., Nowacki, J., O'Callaghan, B.J. and Meroueh, S.O. (2008) *Chem. Biol. Drug Des.*, **71**(6), 529-532.

[112] Leach, A.R.; Gillet, V.J. An Introduction to Chemoinfomatics, Springer: Dordrecht, 2007.

[113] De Azevedo Jr., W. F. (2008) *Curr. Drug Targets*, **9**(12), 1030-1030.

[114] De Azevedo Jr., W. F. and Dias, R. (2008) *Curr. Drug Targets*, **9**(12), 1031-1039.

[115] Dias, R. and De Azevedo Jr., W. F. (2008) *Curr. Drug Targets*, **9**(12), 1040-1047.

[116] Canduri, F. and De Azevedo Jr., W. F. (2008) *Curr. Drug Targets*, **9**(12), 1048-1053.

[117] Pauli, I.; Timmers, L. F. S. M.; Caceres, R. A.; Soares, M. B. P. and De Azevedo Jr., W. F. (2008) *Curr. Drug Targets*, **9**(12), 1054-1061.

[118] Dias, R.; Timmers, L. F. S. M.; Caceres, R. A. and De Azevedo Jr., W. F. (2008) *Curr. Drug Targets*, **9**(12), 1062-1070.

[119] De Azevedo Jr., W. F. and Dias, R. (2008) *Curr. Drug Targets*, **9**(12), 1071-1076.

[120] Caceres, R. A.; Pauli, I.; Timmers, L. F. S. and De Azevedo Jr., W. F. (2008) *Curr. Drug Targets*, **9**(12), 1077-1083.

[121] Barcellos, G. B.; Pauli, I.; Caceres, R. A.; Timmers, L. F. S.; Dias, R. and De Azevedo Jr., W. F. (2008) *Curr. Drug Targets*, **9**(12), 1084-1091.

[122] Timmers, L. F. S.; Pauli, I.; Caceres, R. A. and De Azevedo Jr., W. F. (2008) *Curr. Drug Targets*, **9**(12), 1092-1099.