Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

LEAD OPTIMIZATION EVALUATION USING COMPUTER-AIDED PARADIGM

Z. M. Llarena

Research Center, University of Perpetual Help System – JONELTA, Binan, Laguna

Philippines

Published online: 18 April 2018

ABSTRACT

During the early part of 1970, lead compounds were optimized numerically in the absence of computers. Subsequent to that, a few computer-aided designs were developed but only to a limited extent. Combinatorial chemistry has brought short improvements with drastic failures due to low drug target rates, pharmacokinetic errors and increased toxicity. This paper aimed to apply the concepts of computer-aided drug design to formulation needed for drug development. Integration of computer-aided molecular design to perform computational approaches for investigation of quantitative structure-activity relationships (QSAR) and quantitative structure-properties relationships (QSPR) have improved the potency and selectivity of possible lead compounds prior to clinical trials and hastened the process of drug discovery. These computational algorithms had resolved previous issues and lessened the toxicities and enhanced the pharmacokinetic accountabilities. Virtual screening can assess the therapeutic activity by means of machine learning techniques. Hence, drug candidate optimization can be achieved in a timely and economical manner using computer-aided computational processes.

Keywords: computational approach; QSAR; QSPR; drug design; drug candidate

Author Correspondence, e-mail: zharllarena21@yahoo.com.ph

doi: http://dx.doi.org/10.4314/jfas.v10i3s.73

INTRODUCTION

Novel therapeutic processes start as costly and long period of ways for obtaining solutions. It has a drug discovery cycle estimation of fourteen (14) years at 800 million US dollars. This pattern starts from identification of lead by means of clinical trials. During the start of 1990s,

combinatorial chemistry and large sequencing screening technologies domains have sudden improvements in innovation. It generated matrix for fast discovery methods through permission of huge databases synthesized compounds and monitored in short time span. Unfortunately, these combined results not only failed to augment the quantity of successfully introduced new molecular compounds, but it looked to have worsened the situation. Low hit rates and its unsuccessful hits were unable to be further optimized into current leads and prior to clinical trials. 40-60% failures were at the late stage phase due to pharmacokinetic gaps. Majority of these concerns emphasize the call to alternative method developments that can promote unsuitable compound removal before the significant emission of quantitative resources [1].

A new model in drug discovery needs not only for early lead candidate evaluation for potency and selectivity but also for their possible pharmacokinetic responsibilities. This paradigm would lead to reduction of expensive late phase failures and hastens optimal innovation of novel molecular compounds. Paradigm shift is at the center of applying computational processes for the operation of novel molecular compound discovery [1].

Several computational methodologies are extensively used in the drug discovery processing of extremely difficult, long operational time, and large quantity of resources. Novel drug development usually needs more than ten (10) years and billion dollars of expenses. Early to middle stage drug discovery attempts in drug industries focus on therapeutically relevant small molecule advancements and candidate compound transport to clinical trials. Majority of computational processes are utilized during the early drug discovery stage. Fundamental research results have goals of disease-related biology translation, drug target priority, and novel chemical compound optimization leading to clinical intervention. These techniques are quickly becoming popular and its implementation are appreciated [2]. Application of various terms, namely, computer-aided drug/molecular design (CADD/CAMD), computational/rational drug design (CDD/RDD), and computer-aided molecular modeling (CAMM) were involved in several fields. There is also a term named computer-aided drug discovery and development (CADDD) applied for the total analytical method. Computational and experimental process have vital roles in complementary technique representation for discovery and development of drugs [3].

PROBLEM STATEMENT

This paper would like to discuss applications of computational algorithms for reduction of toxicity and pharmacokinetic improvement on design and discovery of drugs.

THE AIM OF RESEARCH

This paper aims to apply mathematical design applying principles used in QSAR/QSPR computational approaches for better formulation required in drug discovery and design.

METHOD OF RESEARCH

Computer-aided drug design (CADD) generally represents computational materials for the three (3) partitions in compound modeling. It involves digital repository development for the analysis of structure activity relationships (SARs). Computer applications not only utilized for designing molecular entities with determinants exploring physicochemical properties but also systematic evaluation of possible lead candidates are being employed prior to synthesis and testing. This technique is currently emphasizing its significant part for novel molecular compound discovery. Its concentration involves on enhancement of data source design and management, generation of computer guides to yield large databases of pharmacologically important molecular compounds, novel algorithm development leading to evaluation of potency and selectivity of lead candidates, and mathematical modeling design for identification of possible pharmacokinetic accountabilities [1].

In silico techniques are generally used for drug discovery and has three (3) major divisions. These partitions consist of basic computational system for design, implementation, and maintenance in order to process, classify, analyze, and keep fast increasing drug discovery data size. Processes for promoting identification, characterization, and priority of biological targets and link construction to engage targets, and expedite connections between biology and disease. Methodologies that yield better molecular entities leading to drug candidates. These three (3) classes are significant for discovery and development of drugs [2].

Dating the public usage of computers employing chemistry techniques by three (3) or four (4) decades, CAMD started in the 1980s. CAMD has the ability to combine novel processes involved in molecular modeling, thermodynamics, and mathematical optimization for optimal molecular structure development. CAMD has a great advantage in chemical modeling advancements in the last two (2) decades. Researchers have now the ability to create connections between chemical structures and properties at various accuracy levels. New techniques for more accurate process integration are becoming more apparent than semi-empirical mathematical modeling. New combinatorial optimization approaches for CAMD are gaining importance leading to optimization grant across unsteady spaces of inaccessible design [4].

Entry of data is important in attaining the purpose of drug promotion in discovery and development. Large quantities of organic compounds, its biological patterns and informative links are found in scientific publications and case studies. Databases are utilized for collection and storage of these information. Large number of scientific databases are stored every year, while computational techniques are being developed for the design execution of combinatorial sources [1].

Current modern drug discovery has combinatorial chemistry as its analytical element. Databases are usually vast for total synthesis and screening. It is generally known that these libraries may have a huge similar compound quantity owing to its physicochemical properties. The improved possible design permits database optimization on target differences or resemblances. It can promote in lessening redundancy or increasing the optimum number of identified true leads. Variances, scope and group class concepts are usually employed to promote safety in good database sampling using the required number of chemical entities. Virtual database design commonly starts with comprehensive details of all molecular versions under suitable chemical spaces. Two approaches are normally used for the enumeration of molecular analogs: (1) Markush processes which involve attachment of second functional group list to variable locations on a general platform, and (2) chemical transforms which entail reacting molecule parts that proceed to chemical reactions along with its nature. These databases may serve as a tool for optimization for molecular variance or resemblance by using descriptors, namely, composition and topology of chemicals, dimensional structures in respect to functionality purpose, or drug-likeness using experimental rules to trace pharmacokinetic gaps [1].

The pharmaceutical compound fate may be explained by its pharmacokinetic mechanisms. Pharmacological compound effect exertion must have a penetrating ability in tissues and various physiological barriers of gastrointestinal system, brain and other microcirculatory system to reach the blood circulation. It is transferred afterwards to its receptor site for tissue and organ distribution, metabolized by functionalized enzymes, and terminally eliminated from the body through excretion. Genetic diversity in metabolizing enzymes for drugs illustrates that some drug entities may proceed to metabolic activation leading to adverse effects to people. Meanwhile, the pharmacokinetic compound characteristics have direct effect on its safety and utilization [1].

Discovery and development of drugs is a composite method needing expertise from various areas. It involves extremely long working procedures, from lead optimization to clinical trials, needing great financial exertions. In regard with the drug development pattern difficulty,

bioinformatics and computational methods have become flexible facilitating and accelerating tools for design and development of drugs [5].

Hit identification is a method in which drug candidates are recognized by huge libraries involving heuristic outcomes. The linked data mining tools are beneficial for generating huge libraries, and molecular docking has the ability of performing potential candidate identification through docking drugs to protein databases. Drug targets are described to be compound groups that display desired drug-like activity during the screening stage. The target identification process can be done by high throughput screening (HTS) and virtual screening. HTS is done through a total compound library screening in reference with the target using automation; although secondary assays are needed for validation. Virtual screening is a powerful method for drug candidate search through computational methods. One commonly utilized computational approach is the molecular docking method. The crystal target protein structure is needed for binding simulation of in silico in reference to compound databases. Active targets with efficient binding affinity to the protein, given with a docking score, are known as hits and will proceed for further development. Drug targets are subsequently used for optimization to achieve an improvement on potency and pharmacokinetic characteristics and toxicity reduction. The optimization is done through structural compound modification, in which medicinal chemistry and computational methods play vital interactive roles. Quantitative structure-activity relationships (QSAR) and quantitative structure-properties relationships (QSPR) are computational approaches designed for determining the correlation between the compound chemical structures in regard with their activities or properties. Knowing these relationships is beneficial in performing structural modification by medicinal chemists in search of drug candidates [5].

Biophysical techniques developments such as X-ray crystallography and NMR techniques have resulted to protein structure availability increase. This has permitted structural information utilization to pattern discovery of drugs. Without empirical structures, computational approaches are utilized for prediction of target protein 3D structures. Comparative modeling is utilized for target compound prediction in reference of a template with an analogous sequence, by protein structure leveraging that is better than conserved sequence. Homology modeling is a particular type of comparative modeling wherein the template and target proteins have the same evolutionary source. Comparative modeling requires the following steps: (1) related protein identification to provide as template structures, (2) target sequence alignment and template proteins, (3) coordinate duplication for confidently aligned regions, (4) missing atom construction for target structure coordinate, and

(5) model evaluation and refinement [6].

An outburst development has been observed in the accessible protein structural data by X-ray crystallographic and NMR spectroscopic studies with derivation from large quantities of genomic and proteomic data through theoretical modeling. In this aspect, new drug candidate discovery depends on modeling data accuracy in rational drug design due to information from both protein structures and their ligand-binding sites might lead to exploitation. In this instance, there are two commonly use methods termed as molecular docking and molecular dynamics simulation, that take a significant impact in these methods and are commonly merged to observe small molecule interactions with the protein target at its atomic degree [5].

Molecular docking is a concept commonly utilized for the computational approach that tries to perform possible binding mode search of a ligand with its protein. Docking mathematical designs have led to growth by comprehensive generation of conformational protein-ligand complex set, which quantitatively scores them eventually, in accordance to their stability. There are various determinants that affect the ligand binding process to its protein, with the inclusion of thermodynamic and solvation yields and the charge protein and ligand molecule distributions. When the protein target structure is recognized, the rational drug design process accompanies a well-established protocol [5].

Target identification is the preliminary step throughout the novel drug discovery design. Dating 2006, 324 identified molecular targets were only identified for FDA-approved drugs. However, this entire data is under discussion, it is significantly lower than the protein numbers disseminated by the human completion and various pathogen genomes, that may be potential drug candidate. In addition to that, many molecular compounds exhibit their therapeutic activities thorough multiple target modulation, although the multi-target interactions are either enormously unknown or inadequately understood in many instances. The computational tool utilization for protein target prediction of small molecules has been receiving its worth in current period. One of the computational methods illustrated to be effective and economical in lead identification is ‘reverse’ docking. Molecular docking is a process for prediction of primary binding ligand process with a known 3D structure protein, being routinely applied in structure-based drug discovery for target identification virtual screening and drug candidate optimization. Reverse docking, opposite of the traditional method of docking utilization, is to dock a particular small molecule into the limited binding pool sites of protein structures [7].

Chemical space is the complete possible descriptor number from molecular compounds. In similarity with the spatial space degree of the universe, these descriptors are running

quantitatively infinite. In spite of the synthesis developments of organic compounds and the natural product characterization, only a small percentage of compounds has undergone for synthesis and utilization. Hence, through the origin of chemical space exploration in living organisms, new means to fight diseases will develop [5].

The QSAR/QSPR model construction is found on the main concept of machine learning. Currently, a few popular QSAR/QSPR models on the basis of machine learning techniques include multiple linear regressions (MLR), partial least square (PLS), k -nearest neighbor (k -NN), artificial neural network (ANN), support vector machine (SVM), decision tree (DT), and random forests (RF). All of these computational approaches have been discussed in numerous QSAR/QSPR modeling applications. Machine learning tasks are commonly grouped into two wide divisions involving classification and regression tasks. Classification tasks want to discriminate a variable Y into its class or property, where the Y variable could be grouped into two and more than two divisions, which are termed as binary and multi-class categorization. In the opposite, the regression task chiefly concentrates on value prediction of the variable Y with a quantitative output. The MLR, PLS, ANN, SVM, and RF approaches can be applied in both classification and regression tasks, while k-NN and DT are employed only in the classification task. Moreover, the machine learning tasks could be further separated in accordance to their coverage (supervised learning) or omission (unsupervised learning) of the variable Y [5].

Subsequent to the QSAR/QSPR model construction, the proposed model internal validation is important in the reliability model assessments and their capability to perform prediction of biological activities or chemical characteristics with accuracy. In the classification task, four measurements were mostly utilized for prediction performance evaluation of the proposed QSAR/QSPR model by cross-validation (CV), namely accuracy (ACC), sensitivity (SEN), specificity (SPEC), and Matthews correlation coefficient (MCC) [5].

ANALYSIS AND DISCUSSION

Caulerpin, a secondary metabolite of Caulerpa racemosa (Forsk.) J. Agarth, has a possible MAO-B inhibitory effect upon evaluation. Thus, efforts have been claimed to caulerpin exhibition of its possible inhibitory action as a lead compound. Five best analogs have been chosen for analysis from each algorithmic approach. A potential link between the scores gathered from the compounds CLP 012, CLP 049, CLP 068, and CLP 100 implies lead results. In chemical property evaluation from the active structural analogs, a monosubstitution has been found to be symmetrical with chlorine and methyl at positions 4 and 4’ and 5 and 5’,

respectively. The amide or acyl halide polar group presence in main bulky group and a short nonpolar allyl or butyl chain in second bulky group yielded lower docking energies. All active structural analogs showed promising values of their drug-like scores. With the CLP109 exception, all active structural analogs illustrated values superior to 0.75, emphasizing the significant leading values of structural analogs (CLP 012, CLP 049, CLP 068, and CLP100) with values equivalent to 0.87. A drug-like score is a ranging value from 0 to 1, explaining a value of 1 shows that a drug entity is a potential drug or lead candidate [9].

Generated database from the structural analogues of caulerpin done by Lorenzo, et al (2015) can be used as reference for the target lead compounds, namely, CLP068 and CLP049 (as shown below), with found drug-like score of 0.87, showing a good drug candidate for inflammation, pain and cancer.

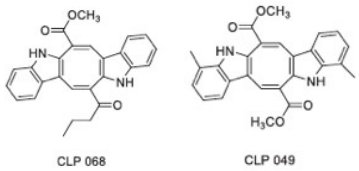

Fig.1. Caulerpin Analogs

Predicted functional groups exhibit the presence of aromatic rings, alkenes, methoxy groups, ketones, and secondary amines in the spectra. Expected molecular weights are 410 g/n (CLP068) and 426 g/n (CLP049). Furthermore, upon handling to gas chromatography factorial design, predicted response data shows that low pH and buffer concentration should be used to promote chemical stability of CLP068 and CLP049, thus, avoiding the formation of possible degradation products under various stress conditions.

CONCLUSION

Computer-aided drug design has a significant impact on drug discovery and development towards novel molecular compound optimization. Through the use of huge libraries of compounds, drug target optimization with efficient binding affinity to a receptor, prior to clinical trials, is thoroughly screened by means of computational approaches such as quantitative structure-activity relationships (QSAR) and quantitative structure-properties relationships (QSPR). These computational algorithms can reduce the expenses brought by failures due to pharmacokinetic gaps, low hit rates and toxicities. Machine learning techniques serve as a tool for computational processes to measure the potency and selectivity

of possible lead candidate. Empirical results may then be integrated to compound databases for further enhancement leading to efficient structural modification. Thus, therapeutically successful novel molecular compounds for drug discovery and development can be optimized by computer-aided computational methods.

REFERENCES

[1] Song, CM, Lim, SJ, and Tong, JC (2009): (Review) Recent advances in computer-aided drug design, Oxford University Press, Briefings in Bioinformatics, Vol 10, No 5, 579-591. doi:10.1093/bib/bbp023

[2] Bajorath, J (2016): (Review) Computer aided-drug discover, Creative Commons Attributions License, F1000 Research, 4(F1000 Faculty Rev):630. doi:10.12688/f1000research.6653.1

[3] Kapetanovic, IM (2008): (Review) Computer-Aided Drug Discovery and Development (CADDD): In Silico-Chemical-Biological Approach, NIH Public Access, Chem Biol Interact, 171(2):165-176. doi:10.1016/j.cbi.2006.12.006

[4] Austin, ND, Sahinidis, NV, and Trahan, DW (2016): (Review) Computer-aided molecular design: An Introduction And Review Of Tools, Applications, and Solution Techniques, ScienceDirect (Elsevier), Chemical Engineering Research and Design, 25 pages. doi:10.1016/j.cherd.2016.10.014

[5] Prachayasittikul, et al (2015): Computer-Aided Drug Design of Bioactive Natural Products, Current Topics in Medicinal Chemistry, 15:1780-1800

[6] Sliwoski, G, et al (2014): (Review) Computational Methods in Drug Discovery, The American Society for Pharmacology and Experimental Therapeutics, Pharmacol Rev 66:334–395

[7] Zheng, et al (2013): Computational Methods for drug Design and Discovery: Focus on China, Elsevier, Trends in Pharmacological Sciences, Vol. 34, No. 10

[8] Lorenzo, VP, et al (2015): (Review) Combined structure- and ligand-based virtual screening to evaluate caulerpin analogs with potential inhibitory activity against monoamine oxidase B, ScienceDirect (Elsevier), Revista Brasileira de Farmacognosia, 25:690-697. doi: 10.1016/j.bjp.2015.08.005

How to cite this article:
Llarena Z M. Lead optimization evaluation using computer-aided paradigm. J. Fundam. Appl. Sci., 2018, *10(3S), 846-854.*