



Essential Basic Protocol in Computer Aided Drug Designing: Efficiency and Challenges

Monishi Dutta Choudhury^{*1}, Rupjyoti Saikia²

¹School of Computer Science and Engineering VIT University, Vellore 632014, India

²Department of zoology, Tinsukia college, Tinsukia-786125, Assam, India

Abstract

This paper describes basic essential protocol required for computational drug design. Efficiency of the protocol, their limitation and way out have also been highlighted. It is suggested that until and unless complete simulation of protein dynamics in terms of real life situation is done achieving complete accuracy of computational tool in drug designing shall remain a challenge

Keywords: Computer Aided Drug Design, Protocol

Corresponding author: : Monishi Dutta Choudhury

School of Computer Science and Engineering, VIT University, Vellore 632014, India.

E-mail: monishi_dc@writeme.com, drmdc@bioinfoaus.ac.in

Citation: Monishi Dutta Choudhury et al. (2018), Essential Basic Protocol in Computer Aided Drug Designing: Efficiency and Challenges. Int J Biotech & Bioeng.4:4,77-80

Copyright: ©2018 Monishi Dutta Choudhury et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Received: March 30, 2018

Accepted: April 09, 2018

Published: April 30, 2018

Basic Concept of Computer Aided Drug Design (CADD)

The process of drug discovery is very complex, time and resource consuming that requires an interdisciplinary effort to design effective and commercially viable drugs. Numerous technologies have been established and applied in drug R & D to minimize the research cycle and the expenses [1]. Bioinformatics is an emerging field that can be believed as a central hub which unites several disciplines and methodologies [4]. In silico or Computer aided drug design (CADD) is such a specialized discipline that provides an in depth discussion about the computer assisted techniques used to discover, design and optimize new, effective and safe drugs that can reduce time as well as expenses associated with the process up to 50%. It uses computational methods to simulate drug-receptor interactions and are greatly dependent on bioinformatics tools, applications and databases such as over viewing tools, homology modelling and homology modelling programs, molecular dynamics, molecular docking and QSAR analysis [1][4].

Computer aided drug design has significantly extended its range of applications, covering virtually all phases in the drug discovery pipeline, ranging from identification of target to lead discovery and from lead optimization to preclinical or clinical trials. Recently, CADD has been used in designing highly selective ligands for certain targets which shares much related conformations with many other proteins. Target identification and validation are the first step in the drug discovery pipeline which is still a challenging task among thousands of candidate macromolecules. Genomic and proteomic approaches are the key tools for target identification which seems to be laborious and time consuming. Hence, complementary to the experimental methods, a sequence of computational (in silico) tools has been developed for target identification. They can be categorized into sequence based and structure based approaches. Structure based approaches are not yet applicable as the structure of the target macromolecule is unknown. In these situations, quantitative structure activity relationship (QSAR) techniques provide the best approach to rational drug design. Traditional two-dimensional QSAR methods attempt to compare biological activity with local features of atoms, whole molecular properties and substituent effects. Recently, new advances in traditional QSAR as well as focuses on three-dimensional QSAR are become more advantageous in the process of rational drug design and development [1].

Overview of tools and techniques employed in CADD

Computer aided drug design (CADD) makes use of different software, databases and web services which are categorized according to their application field that cover the whole drug design pipeline.

Databases

Some of the databases used in CADD include Protein Data Bank (PDB), NCBI PubChem BioAssay, ChEMBL, DrugBank, Binding DB and PDB Sum.

Protein Data Bank: The Protein Data Bank is a repository for the three-dimensional structural data of biological macromolecules such as proteins and nucleic acids and is a basic resource in the area of structural biology such as structural genomics. The PDB is administered by an organization Research Collaboratory for Structural Bioinformatics (RCSB). The data generated by X-ray crystallography

or NMR spectroscopy are submitted by biologists and biochemists from all over the world and are easily accessible on the internet via the websites of its member organisations [3].

NCBI PubChem BioAssay: PubChem is a free repository for biological activity data of small molecules and RNAi reagents. The main objective of PubChem is to provide free and easy access to all the deposited data and to deliver intuitive data analysis tools. PubChem's bioassay data are incorporated into the NCBI Entrez data retrieval system, thus making the data searchable and accessible by Entrez queries. Also, PubChem constantly optimizes and improves the deposition system answering various demands of both high and low volume depositors. It allows users to search, review and download bioassay description and data as well as permits researchers to collect, compare and analyze biological investigation results through web based and programmatic tools [14].

ChEMBL: ChEMBL is an Open Data database comprising binding, functional and ADMET information for a huge number of drug-like bioactive compounds. The data are abstracted manually from the primary published literature in an orderly basis and further curated and standardized to increase their quality and utility through a wide range of chemical biology and drug discovery research problems. At present, the database covers 5.4 million bioactivity data for more than 1 million compounds and 5200 protein targets [4a].

DrugBank: The DrugBank is a unique bioinformatics and cheminformatics database that combines detailed chemical, pharmacological and pharmaceutical data of drug with comprehensive sequence, structure and pathway information of drug target. The database includes 6825 drug entries comprising 1541 FDA approved small molecule drugs, 150 FDA approved biotech (protein/peptide) drugs, 86 nutraceuticals and 5082 experimental drugs. Furthermore, 4323 non-redundant protein such as drug target, enzyme, transporter and carrier sequences are also linked to these drug entries. Each DrugCard entry comprises over 150 data fields covering half of the information about drug/chemical data and the other half about the drug target or protein data.

BindingDB: BindingDB is a public, web-accessible database of measured binding efficacies, involved mainly in the interactions of proteins that considered as candidate drug-targets with small, drug-like molecules called ligands. The main aim of BindingDB is to support medicinal chemistry and drug discovery via literature awareness and progress of structure-activity relations (SAR and QSAR); authentication of computational chemistry and molecular modelling techniques such as docking, scoring and free energy methods; chemical biology and chemical genomics; and basic studies of the physical chemistry of molecular recognition.

PDBsum: PDBsum is a pictographic database that provides an overview of the contents of 3D structures deposited in the Protein Data Bank (PDB). It shows the molecule(s) that make up the structure i.e., protein chains, DNA, ligands and metal ions, and schematic diagrams of the interactions between them.

Chemical structure representation

This includes software such as ACD ChemSketch, OpenBabel, Perkin Elmer ChemBioOffice 2012 and HyperChem.

ACD ChemSketch: ACD ChemSketch freeware is a drawing package that allows drawing of chemical structures of organics, organometallics, polymers and Markush structures. It also comprises other features such as derivation of molecular properties, cleaning and viewing of 2D and 3D structures, functionality for naming structures and prediction of logP.

OpenBabel: OpenBabel is free software designed to support

molecular modelling, chemistry and interconversion of file formats and data. Due to the strong association to informatics this belongs to the category of cheminformatics and is distributed under the GNU GPL.

Perkin Elmer ChemBioOffice 2012: ChemBioOffice is a scientifically intellectual, integrated suite of personal productivity tool that enables to capture, store, retrieve and share data and information of compounds, reactions, materials and their properties. It helps to visualize and gain a deeper understanding of the results and correlate biological activity with chemical structures. It includes the applications such as ChemBioDraw, ChemBio3D and ChemBioFinder.

HyperChem: HyperChem is sophisticated molecular modelling software recognized for its quality, flexibility and ease of use.

Drug likeness and Lipinski filter screening

For drug likeness and Lipinski filter screening software like Molsoft L.L.C is used. It is a free web server used to screen the drug likeness property of a molecule and also to study whether the molecule satisfy all the Lipinski rule of 5.

ADME/Tox Screening

ADME/Tox is an abbreviation of pharmacokinetics and pharmacology for absorption, distribution, metabolism, excretion and toxicity which defines the disposition of a pharmaceutical compound within an organism. These standards have influence on the level of drug as well as kinetics of exposure of drug to the tissues and thus the pharmacological activity of the compound as a drug.

Mobyle@RPBS is an online tool especially designed for ADME/Tox screening of small molecules and is maintained by the University of Paris, France. The FAF-Drugs ADME/Tox tool is used to screen the ADME/Tox profile of the compounds. The SMILES strings of the compounds are generated and loaded into the Mobyle@RPBS server [8].

Target prediction

For any natural product or synthetic compound predicting its probable bioactivity is a challenging task. Target fishing tools are employed to identify the probable bioactivity and suitable drug target for the isolated natural product. For target fishing number of programs is used in CADD.

ReverseScreen3D is a reverse virtual screening method which can be used to predict the potential protein targets of a query compound of interest. The method uses a 2D fingerprint-based method to select a ligand template from each unique binding site of each protein within a target database. The target database contains only the structurally determined bioactive conformations of known ligands. The 2D comparison is followed by a 3D structural comparison to the selected query ligand using a geometric matching method, in order to prioritize each target binding site in the database [5].

Binding site prediction

Identification of ligand binding site on a protein is crucial for molecular docking, structural identification, de novo drug design and comparison of functional sites. Q-SiteFinder is an approach to predict the ligand binding site of a protein. It utilizes the interaction energy between the protein and a simple Van der Waals probe to determine energetically favourable binding sites. Energetically suitable probe sites are grouped on the basis of their spatial proximity and then categorized according to the sum of interaction energies for sites within each group [6].

Docking

Molecular docking is the technique of predicting and analyzing the interactions between protein receptors and ligands. It provides an in depth view of drug receptor interactions and also has created a rational approach to drug design [1]. Scoring methods are employed to rank

the affinity of ligand to bind to the active site of a receptor. In virtual high throughput screening, compounds are docked into an active site and then scored to define the more likely one that bind tightly to the target molecule [10]. Several molecular docking programs are available. Some of them are open source and some are commercial. Some of the open source docking programs include AutoDock [22], DOCK [9], Hex [7], whereas GOLD [12], FlexX [13] are some of the commercial docking packages comprising many features.

QSAR

Quantitative structure activity relationship (QSAR) is an approach to study the correlation of structural or property descriptors of compounds with activities. These physicochemical descriptors are determined either empirically or by computational methods and the activities comprise chemical measurements and biological assays. QSAR is currently applied in many disciplines including drug design and environmental risk assessment. It is a rational approach to lead optimization when the structure of the target is unknown. Since it is based on activity data, QSAR has the advantage of modelling in vivo situation [2].

STATISTICA is one of the best data analysis software designed to analyze the pertinent data. It features great tools that output data in a detailed and customizable form. STATISTICA is able to recognize the pattern of user input data as well as output detailed graphs and statistical table.

Efficiency and Challenges

Efficiency of any software intended to explain biological mechanisms depends upon sound understanding of the biological processes and attributing functional aspects of those process in relation to characteristic feature of the same. Most of the docking software mentioned above assume that the receptors are static in their conformation. Ligands are docked to the active site of the static receptors and the scores are considered as indicator of bonding. But it does not happen in living systems where the receptors are continuously changing the conformation. Catching the right conformation of the protein at a point of time when ligands reach to it through a process of drug release is the challenge. If the conformation of the protein at a future point of time is predicted and bonding efficiency of ligand is scored with respect to that future time, probably more accuracy in terms of real life situation could be achieved. Dynamics simulation of the proteins concerned was a challenge.

Software like GROMACS has come up to meet the challenge and prediction of changed conformation of protein with respect to future time in terms of nano and pico second etc are now possible through Molecular Dynamics simulation. But simulation in terms of minute and hour is still a challenge.

Groningen MACHine for Chemical Simulations (GROMACS) is a molecular dynamics package mainly designed for simulations of proteins, lipids and nucleic acids. It was originally developed in the Biophysical Chemistry department of University of Groningen, and is now maintained by contributors in universities and research centers worldwide.[16][17][18] GROMACS is one of the fastest and most popular software packages available,[19][20] and can run on central processing units (CPUs) and graphics processing units (GPUs).[21] It is free, open source software released under the GNU General Public License (GPL),[15] and starting with version 4.6, the GNU Lesser General Public License (LGPL).

References:

[1] Akhila, S., Aleykutty, N. A. and Manju, P., 2012. [Docking studies on peperomia pellucida as antidiabetic drug](#). *International Journal of Pharmacy and Pharmaceutical Sciences*. 4: 4.

[2] Azhar, S. M., Saudagar, R. B. and Daharwal, S. J., 2006. [Computer](#)

[aided drug design and bioinformatics: a current tool for designing](#). *Pharmaceutical Information*. 4(4).181-183

[3] Bramer, S. E. V., 1998. [An introduction to Mass Spectrometry](#). Revised.

[4] DBM, V., Kelmani, C., Patil, R. and Hegade, P., 2007. [Computer Aided Docking Studies on Antiviral Drugs for SARS](#). *World Academy of Science, Engineering and Technology*. 30: 297-299.

[4a] Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B. and Overington, J. P., 2012. [ChEMBL: a large-scale bioactivity database for drug discovery](#). *Nucleic Acids Res*. 40(1): 1100-1107.

[5] Kinnings, S. L. and Jackson, R. M., 2011. [ReverseScreen3D: a structure-based ligand matching method to identify protein targets](#). *J Chem Inf Model*. 51(3): 624-34.

[6] Laurie, A. T. R. and Jackson, R. M., 2005. [Q-SiteFinder: an energy based method for the prediction of protein-ligand binding sites](#). *Bioinformatic*. 21(9): 1908-1916.

[7] Macindoe, G., Mavridis, L., Venkatraman, V., Devignes, M. D. and Ritchie, D. W., 2010. [HexServer: an FFT-based protein docking server powered by graphics processors](#). *Nucleic Acids Res*. 38: 445-449.

[8] Miteva, M. A., Violas, S., Montes, M., Gomez, D., Tuffery, P. and Villoutreix, B. O., 2006. [FAF-Drugs: free ADME/tox filtering of compound collections](#). *Nucleic Acids Res*. 34: 738-744.

[9] Moustakas, D. T., Lang, P. T., Pegg, S., Pettersen, E., Kuntz, I. D., Brooijmans, N., Rizzo, R. C., 2006. [Development and validation of a modular, extensible docking program: DOCK 5](#). *J Comput Aided Mol Des*. 20(10-11): 601-619.

[10] Rahman, M. M., Karim, M. R., Ahsan, M. Q., Khalipha, A. B. R., Chowdhury, M. R. and Saifuzzaman, M., 2012. [Use of computer in drug design and drug discovery: a review](#). *International Journal of Pharmaceutical and Life Sciences*. 1(2): 1-21.

[11] Ramulu, J. and Goverdhan, P., 2012. [Computer aided drug design an emerging tool for research and drug development](#). *Pharmatutor art*. 1-16.

[12] Sato, H., Shewchuk, L. M. and Tang, J., 2006. [Prediction of multiple binding modes of the CDK2 inhibitors, anilopyrazoles, using the automated docking programs GOLD, FlexX, and LigandFit: an evaluation of performance](#). *J. Chem. Inf. Model*. 46: 2552-2662.

[13] Stahl, M., 2000. [Modifications of the scoring function in FlexX for virtual screening applications](#). *Perspectives In Drug Discovery And Design*. 20(1): 83-98.

[14] Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., Zhou, Z., Han, L., Karapetyan, K., Dracheva, S., Shoemaker, B. A., Bolton, E., Gindulyte, A. and Bryant, S. H., 2012. [PubChem's BioAssay Database](#). *Nucleic Acids Res*. 40(1): 400-412.

[15] About Gromacs". [gromacs.org](#). 16 August 2010. Retrieved 20120626.

[16] People — Gromacs". [gromacs.org](#). 14 March 2012. Retrieved 26 June 2012.

[17] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005). ["GROMACS: fast, flexible, and free"](#). *J Comput Chem*. 26 (16): 1701-18. doi:10.1002/jcc.20291. PMID 16211538.

[18] Hess B, Kutzner C, Van Der Spoel D, Lindahl E (2008). ["GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation"](#). *J Chem Theory Comput*. 4 (2): 435. doi:10.1021/ct700301q.

[19] Carsten Kutzner; David Van Der Spoel; Martin Fechner; Erik Lindahl; Udo W. Schmitt; Bert L. De Groot; Helmut Grubmüller (2007). ["Speeding up parallel GROMACS on high-latency networks"](#). *Journal*

of Computational Chemistry. 28 (12): 2075–2084. doi:10.1002/jcc.20703. PMID 17405124.

[20] Berk Hess; Carsten Kutzner; David van der Spoel; Erik Lindahl (2008). “[GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation](#)”. *Journal of Chemical Theory and Computation*. 4 (3): 435–447. doi:10.1021/ct700301q.

[21] GPUs — Gromacs”. gromacs.org. 20 January 2012. Retrieved 26 June 2012.

[22] Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J., 1998. [Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function](#). *J. Computational Chemistry*. 19: 1639-1662.