Advancing Computer-Aided Drug Discovery (ACADD): *In-Silico* Approach towards Nuclear Receptors by Big Data

Abstract:

The progression of drug discovery and development is time consuming and costly. Advancing Computer-aided drug discovery (ACADD) is an effective tool in reducing the time and cost of research and development. This study deals with the evaluation of the nuclear receptors for the *in-silico* biological activity using ligand betulinic acid and dexamethasone. Docking results showed that binding energy was -74.190 kcal/mol when compared with that of the standard (-51.551 kcal/mol). Interaction energy -44.16 & -25.14 kcal/mol) of the ligands also coincide with the binding energy. These ligands have shown the best ligand-receptor interaction based on their structural parameters.

Key words: Drug discovery, docking, *in-silico*, betulinic acid, ligands, binding energy.

1 Introduction

With the tremendous increase in cost of drug discovery, efforts were initiated to identify and employ tools to reduce the associated costs. ACADD technologies is a tool in this direction and has the potential to reduce the number of ligands to be screened in experiments. ACADD can be broadly divided into two segments, namely ligand-based (LB) and structure-based (SB) screening. In-silico high throughput screening (in-silico HTS) is one such tool that is widely used for screening of ligands. In addition, use of ACADD ensures identification of an optimized ligand for non-clinical studies, thereby expediting the whole process of taking a molecule to market [1-2]. Utilization of ACADD is growing leaps and bounds and assuming a significant role in drug discovery and development.

Nuclear hormone receptors (NR) are ligand-activated transcription factors that regulate cell growth, reproduction, and metabolism [3]. Structurally, all nuclear receptors is a super family of highly conserved topology of structural motives though they bind with different ligands. NR consists of a membrane-bound, 7-membered helical ring structure. Upon binding with the ligands of extracellular domain, the C-terminal domain (LBD) locks with the zinc fingers to induce post-transcriptional activity. Similarly, the N-terminal region is responsible for ligand-independent post-transcriptional activation (AF-1) [4].

All these research efforts have generated enormous amounts of data for drugs and drug candidates and moved modern drug discovery into an era of ‘big data’. The term ‘big data’ refers to massive data, which have large, varied and complex data structures, with associated difficulties of storing, analysing,

and visualizing them using traditional computational approaches [7–14]. Being mostly used in the information technology field, big data is now expanding in all science and engineering domains, including drug discovery.

2 Proposed Work

In the current era of big data, developments in computational tools and the rapid growth of public data sources have advanced the CADD. The proposed approach have been applied to the data generated along various stages of drug discovery and development and affirmed the value of big data by reducing the drug attrition. The challenges brought by the multiple Vs feature of big data require the development of appropriate computational approaches and algorithms.

There are ten characteristics that are intrinsic for big data in drug discovery in table 1. : including volume, velocity, variety, veracity, validity, vocabulary, venue, visualization, volatility, and value.

Table 1. Ten Characteristics of Big Data in Drug Discovery

Ten characteristics of big data in drug discovery		
1.	Volume	Size of data.
2.	Velocity	Speed of new data generation.
3.	Variety	Various formats of data.
4.	Veracity	Quality of data.
5.	Validity	Authenticity of data.
6.	Vocabulary	Terminology of data.
7.	Venue	Platform of data generation.
8.	Visualization	View of data.
9.	Volatility	Duration of data usefulness.
10.	Value	Potential of data usefulness to reduce the cost of drug discovery and development

Compilation of large amounts of data generated daily and shared through public databases, such as Enamine REAL Database, ChEMBL, PubChem, and so on[15] This work selected these databases for high resolution structures of nuclear receptor complexes to study the quality of our docking procedure. It represents the volume and velocity of available data. Currently, most data depository portals (e. g., PubChem) gather data from diverse sources, which define the variety of data.

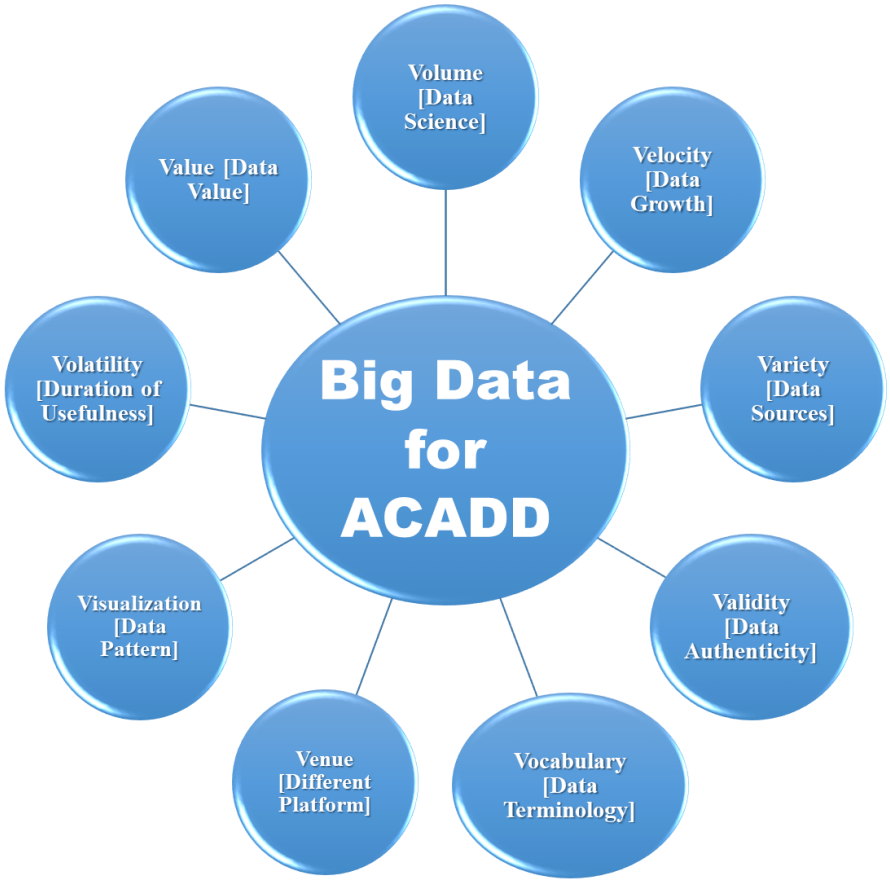


Fig 1. Big Data for Drug Discovery

The application of big data approaches in drug discovery and development, particularly during early stages, has proved valuable as shown fig 1. The challenges brought by the multiple Vs feature of big data require the development of appropriate computational approaches and algorithms. In addition to the progress in ML applications in drug discovery described earlier, the multiple Vs features, such as volume, velocity, variety, vocabulary, and volatility, require better database management, data curation, and web portal design. The variety, veracity, validity, and venue features also require further refinements of experimental protocols, better quality controls, and more transparent data reporting.

In present work an application of a recent docking algorithm to a set of protein complexes from the nuclear receptors (NR) super family for those known three-dimensional structures. Starting point for the simulations assumed that both molecules (receptor and coactivator) in their native states were shifted apart from each other to an arbitrary distance of 40°A between their centres of gravity. Ligand molecules were not explicitly present in simulations, but their influence on the system was incorporated in structural restraints imposed on receptor molecules.

Present experimental studies were carried out using the tools Discovery Studio 3.5, Ligand Fit, C-Docker.

Steps followed for ligand fit

1. Potent inhibitor molecules which can inhibit the action of spla2 were taken.
2. The pharmacophore were selected from through literature.
3. The molecules which are to be docked in a receptor site are created in a SD file so as all molecules are processed for the docking score at a site.
4. The active site of a protein was identified from receptor cavities which is identified by the flood flow algorithm.
5. The active site was placed by the dock ligand
6. The SD file was selected and docking score is determined.
7. Thus, the docking score for a set of molecules are calculated through ligand fit.

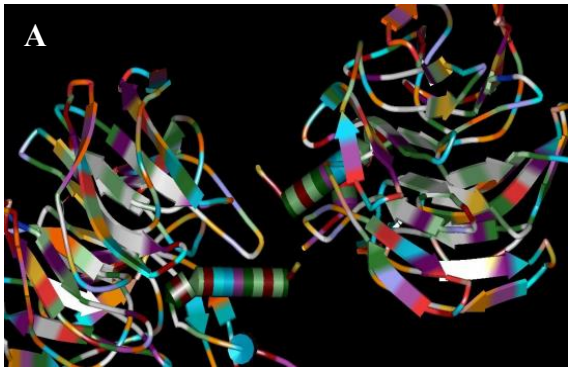
Steps for C-Docker

1. Open the receptor protein and apply the force field
2. The selected molecule works as receptor after that ligand define bubble from selection.
3. Open the C-docker protocol and set the parameters.
4. Run the protocol.

3 Experimental Results

To study the binding mode of betulinic acid compound in the binding site of apoptotic proteins, intermolecular flexible docking simulations were performed and energy values were calculated from the docked conformations of the protein-inhibitor complexes. Docking studies yielded decisive information regarding the course of the inhibitors in the binding the target proteins. Several potential inhibitors have been identified through the docking simulation. The binding affinity of the apoptotic proteins with the betulinic acid compound was measured by kcal/mol. The docking scores were Nuclear Factor NFκB, NrF2 protein (Fig.2. & Fig. 3.). Docked pose of nuclear receptors with the ligand betulinic acid clearly demonstrated the binding positions of the ligand with the enzyme. Hydrogen bond formation was good in all the seven proteins, when docked with betulinic acid. The protein-ligand interaction plays an important role in structural-based designing (Fig. 4.).

Fig. 2. Structure of Nuclear Receptors (NR)



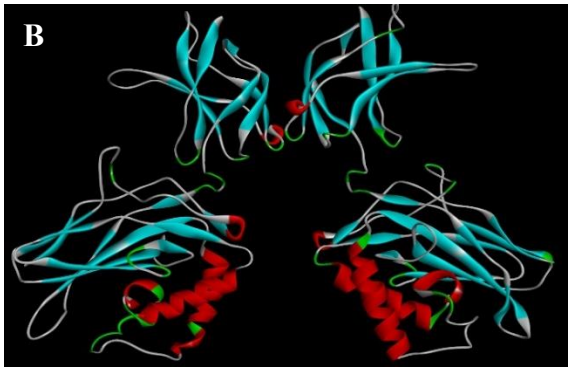


Fig. 3. Ligand binding domain for Nuclear Receptors (NR)

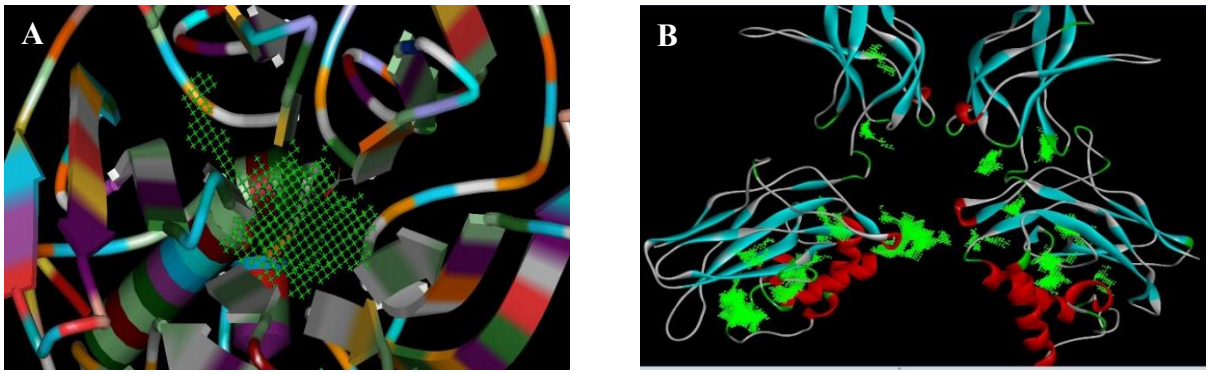
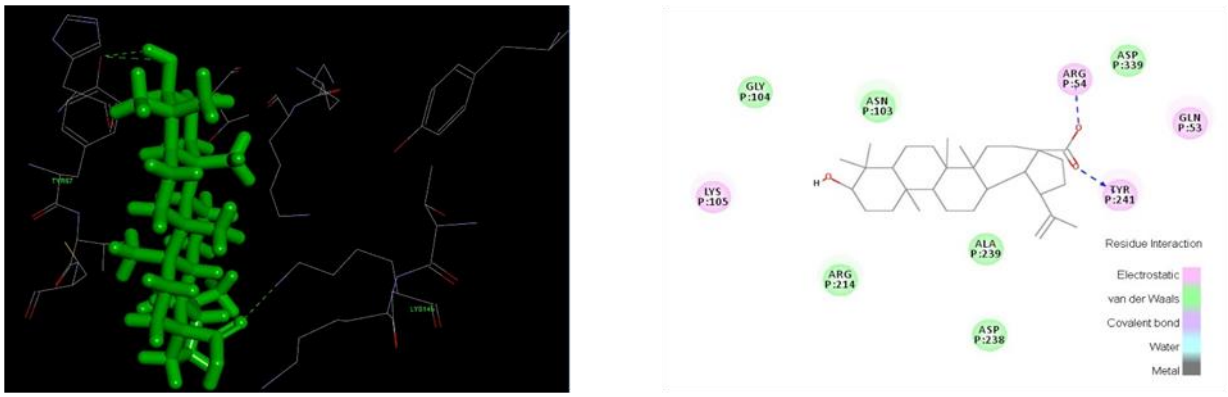


Fig. 4. Ligand Betulinic acid structure and energy value



The ligand scores were evaluated with the help of Lipinski's rule. The docking studies were performed by the use of C-docker. In the docking studies, lesser binding energy has higher the predicted biological activity. The following table enlists the docking energy and the corresponding minimization energies obtained for the beast conformer for each molecule. The activity of each molecule may be contributed by the best lowest energy obtained in the c- docker with the corresponding dock in energy was mentioned in Table 2.

Table 2. C-Dock energy after subjecting to ligand fit.

Name	Van der Waals Energy	Electrostatic energy	-C docker energy	-C docker interaction energy
1SVC	-712.141	-111,645		
Betulinic acid	-12.682	-21.719	74.190	44.965
Dexamethasone	-4.297	7.490	51.551	25.144

Every molecule in the prepared bio active compound SD file will be docked into the binding site chosen, the fits will be automatically processed according to the preferences chosen and saved into the output SD file. The docking score is the negative values of the non-bonded inter molecular energy; if the ligand

atom has partial charge on it, the electrostatic grid is used to estimate electrostatic energy. The activity of each molecule may be contributed by the best lowest energy obtained in the ligand fit with the corresponding dock scores mentioned in Table 3.

Table 3. Docking scores of molecules obtained after subjecting to ligand fit

Name	Dock Score
Betulinic acid	19.14
Dexamethasone	39.72

4 Discussion

Drug discovery encompasses identification of targets, discovery and optimization of leads through various in-vitro and in-vivo studies which eventually would lead to drug development process. ACADD assumes an important role in this whole process of discovery and development. Conventional methods of drug development attracts huge costs and is a time-intensive process [5]. ACADD would assist in identifying potential leads via bioinformatics tools. Bioinformatics tools in turn are capable of predicting the possible binding sites, active drug molecules through docking studies. *In silico* approach of drug discovery, part of structure based drug design, would simulate an active drug candidate [6]. In the current study, the docking results obtained through C-docker showed a good binding interaction between the proteins NRf2 and NFkb with betulinic acid ligand which was evident from the formation of hydrogen bond between the proteins and ligand, in conformation with the Lipinski rule of drug discovery. Results of the current study clearly endorses the use of *in silico* approach for molecular docking of NRs with betulinic acid, thereby showing the potential of betulinic acid for the treatment of various diseases.

5 Conclusion

Computer based approaches are useful tools to infer and guide experiments to expedite the antibiotic drug design process This paper proposed Advancing Computer-Aided Drug Discovery (ACADD): In-Silico Approach Towards Nuclear Receptors. Nuclear receptors account for a sizable portion of therapeutic targets. Current research on NRs have not yielded potential candidates despite various approaches. The current results indicate that from the betulinic acid have better binding sites and interactions with nuclear receptor for its biological activity.

Competing Interests Disclaimer:

Authors have declared that no competing interests exist. The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the authors and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the producing company rather it was funded by personal efforts of the authors.

6 References

1. J. Clark, P. Tiwary *et al*., "Prediction of Protein–Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations"**,** *Journal of Chemical Theory and Computation*, 2016, vol. 12, no. 6, pp. 2990 – 2998.
2. W. R. Chao, D. Yean, *et al.*, "Computer-Aided Rational Drug Design: A Novel Agent (SR13668) Designed to Mimic the Unique Anticancer Mechanisms of Dietary Indole-3-Carbinol to Block Akt Signaling"**,** *Journal of Medicinal Chemistry***,** 2007, vol. 50, issue 15, pp. 3412 – 3415.
3. J. P. Overington, B. Al-Lazikani and A. L. Hopkins, How many drug targets are there?, *Nature Reviews Drug Discovery*, 2006, vol. 5, pp. 993–996.
4. A. R. Clark and M. G. Belvisi, Maps and legends: the quest for dissociated ligands of the glucocorticoid receptor, Pharmacology & Therapeutics, 2012, vol. 134, pp. 54–67.
5. Tran, Nhut et al. "Identification of Novel Compounds against an R294K Substitution of Influenza A (H7N9) Virus Using Ensemble Based Drug Virtual Screening." *International Journal of Medical Sciences*, 2015, vol. 12, issue 2, pp. 163–176.
6. A. R. Laurie, R. M. Jackson; Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites, *Bioinformatics*, 2005, vol. 21, issue. 9, pp. 1908–1916.
7. Palaniammal, K., & Vijayalakshmi, S. (2013). Semantic Web Based Efficient Search Using Ontology and Mathematical Model. International Journal of Engineering and Technology.https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.641.9009&rep=rep1&type=pdf
8. Palaniammal, K & Vijayalakshmi, S 2014, 'QoS based ranking methodology for semantic search system', Australian Journal of Basic and Applied Sciences, vol. 8, no. 18, pp. 374-382. http://www.ajbasweb.com/old/ajbas/2014/December/374-382.pdf.
9. K.Palaniammal, M.Indra devi, S.Vijayalakshmi,"An Unfangled approach to semantic search for e-tourism domain" ICRTIT-12 ,IEEE Explorer, PP: 130-135,2012. doi: 10.1109/ICRTIT.2012.6206758.
10. K.Palaniammal, S.Vijayalakshmi, "Ontology Based Meaningful Search Using Semantic Web and Natural Language Processing Techniques" ICTACT Journal on Soft Computing,2013, Volume: 04, Issue: 01, PP: 662-666. https://core.ac.uk/download/pdf/25558297.pdf.
11. BL Nithiasree, K Palaniammal 2017. Aspect Ranking: Identifying Important Product Aspects by Exploring Online Consumer Reviews, Data Mining and Knowledge Engineering 9 (5),pp. 93-96. http://www.ciitresearch.org/dl/index.php/dmke/article/view/6717.
12. Palaniammal, K & Vijayalakshmi, S 2014, 'Improving search performance for tourism domain using semantic web and bayesian network', Information-An International Interdisciplinary

Journal, vol. 17, no. 8, pp. 3675-3682. http://www.information-iii.org/information_journal.html.

13. Palaniammal, K & Vijayalakshmi, S 2013, 'Semantic web based spatial search for tourist attractions using google maps', International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 10, pp. 1-5. https://www.tce.edu/sites/default/files/PDF/MCA/MCA_Publications.pdf.
14. Palaniammal, K & Vijayalakshmi, S 2013, 'A survey on semantic web search and technologies', International Journal on Computer Science and Technology, vol. 4, no. 4, pp. 334-340. http://ijcst.com/vol44/3/kpalaniammal.pdf.
15. Zhao L, Ciallella HL, Aleksunes LM, Zhu H. Advancing computer-aided drug discovery (CADD) by big data and data-driven machine learning modeling. Drug Discovery Today. 2020 Jul 11.
16. Hoque I, Chatterjee A, Bhattacharya S, Biswas R. An approach of Computer-Aided Drug Design (CADD) tools for in silico pharmaceutical drug design and development. Int. J. Adv. Res. Biol. Sci. 2017;4(2):60-71.
17. Yu W, MacKerell AD. Computer-aided drug design methods. InAntibiotics 2017 (pp. 85-106). Humana Press, New York, NY.
18. Olasupo SB, Uzairu A, Shallangwa GA, Uba S. Computer-aided drug design and in silico pharmacokinetics predictions of some potential antipsychotic agents. Scientific African. 2021 Jul 1;12:e00734.
19. Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 2: a discussion of chemical and biological data used for AI in drug discovery. Drug Discovery Today. 2021 Jan 27.