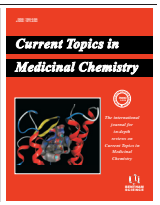
Current Topics in Medicinal Chemistry, 2020, *20*, 1761-1770

REVIEW ARTICLE



Recent Trends in Drug Design and Discovery

Devadasan Velmurugan[1,*], R. Pachaiappan[2] and Chandrasekaran Ramakrishnan[3,*]

[1]*CAS in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai - 600025, India;* [2]*Department of Biotechnology, SRM Institute of Science and Technology, Kattankulathur - 603203, Kanchipuram District, Tamilnadu, India;* [3]*Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology (IIT) Madras, Chennai - 600036, India*

Abstract: ***Introduction*:** Structure-based drug design is a wide area of identification of selective inhibitors of a target of interest. From the time of the availability of three dimensional structure of the drug targets, mostly the proteins, many computational methods had emerged to address the challenges associated with drug design process. Particularly, drug-likeness, druggability of the target protein, specificity, off-target binding, etc., are the important factors to determine the efficacy of new chemical inhibitors.

***Objective*:** The aim of the present research was to improve the drug design strategies in field of design of novel inhibitors with respect to specific target protein in disease pathology. Recent statistical machine learning methods applied for structural and chemical data analysis had been elaborated in current drug design field.

***Methods*:** As the size of the biological data shows a continuous growth, new computational algorithms and analytical methods are being developed with different objectives. It covers a wide area, from protein structure prediction to drug toxicity prediction. Moreover, the computational methods are available to analyze the structural data of varying types and sizes of which, most of the semi-empirical force field and quantum mechanics based molecular modeling methods showed a proven accuracy towards analysing small structural data sets while statistics based methods such as machine learning, QSAR and other specific data analytics methods are robust for large scale data analysis.

***Results*:** In this present study, the background has been reviewed for new drug lead development with respect specific drug targets of interest. Overall approach of both the extreme methods were also used to demonstrate with the plausible outcome.

***Conclusion*:** In this chapter, we focus on the recent developments in the structure-based drug design using advanced molecular modeling techniques in conjunction with machine learning and other data analytics methods. Natural products based drug discovery is also discussed.

ARTICLE HISTORY

Received: November 22, 2019
Revised: December 25, 2019
Accepted: January 10, 2020

DOI:
10.2174/1568026620666200622150003

Keywords: Structure-based drug design, SBDD, Machine learning, QSAR, Data analytics, Data science.

1. INTRODUCTION

Structure-based drug design (SBDD) is now in a big data era where one gains more access to perform thorough data analytics on structural data of proteins, drugs/inhibitors, enzyme kinetics, gene expression, clinical trials, etc. Moreover, one can integrate heterogeneous data to infer a particular biological mechanism associated with either normal functioning of the cell or its dysfunction. The key players in this approach are, mainly, the databases and analytical methods. Many databases are available not only as data sources, but also they have embedded with their own predictive methods. However, the choice of the scope of prediction is not limited to available resources. As it is multidimensional, most of the time, the existing analytics can serve only for customized data set. In this context, one can employ different analytical methods against different types of structural/biological databases. The following sections present the overview of the methods employed so far for SBDD and drug discovery, and also the current developments in the post-genomic and big data era.

Before discussing the data-driven drug discovery process, it is important to understand the scope of SBDD and its associated techniques. For more than two decades, the structural properties of a wide range of biological molecules (from small molecules to macromolecules such as protein, DNA, RNA, carbohydrates, etc) were studied with atomic details using molecular modeling methods. Precisely, the

*Address correspondence to these authors at the CAS in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai-600025, India; E-mail: shirai2011@gmail.com (D.V.); Protein Bioinformatics Lab, Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology (IIT) Madras, Chennai - 600036, India; E-mail: ramki.rpcr@gmail.com (C.R.)

1873-4294/20 $65.00+.00

two levels of calculations, namely, quantum mechanics (electronic properties) and molecular mechanics (force field-based) are crucial.

1.1. Three Dimensional Structure Prediction

Fold recognition and homology modeling are the two distinct methods widely used for prediction of 3D structure of proteins for which no experimentally determined structures are available (Fig. **1**). Fold recognition is basically used for assigning the fold for the protein sequence of interest and it employs SCOP, CATH, FSSP resources instead of high sequence identities/similarities [1].

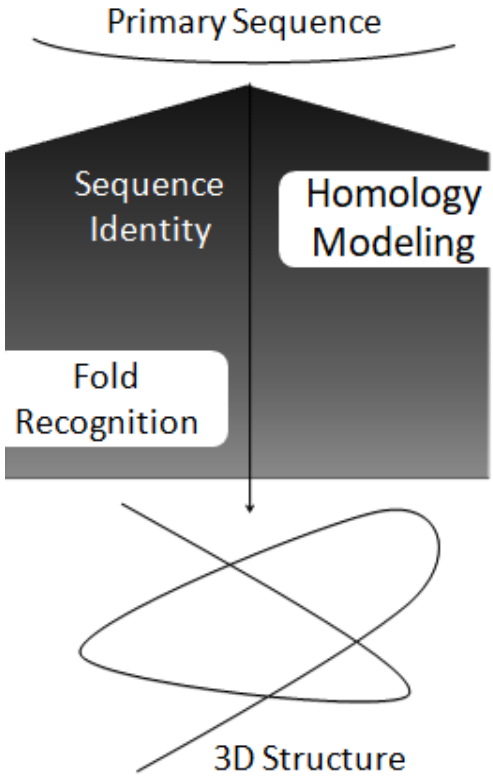


Fig. (1). Scope of Homology modeling and *ab initio* methods for fold recognition with respect to the sequence identity.

1.2. Molecular Docking & Screening

Docking tools are used to identify the biologically relevant binding mode of the two molecules. It is very effective for small molecule ligands, binding the proteins or enzymes. Docking relies on two components, the search algorithms to sample the conformations of the ligand at the rigid or flexible active site and the scoring functions to score all the poses [2]. For conformational sampling, different search algorithms like a stochastic torsional method to sample the conformations based on the rotatable bonds, genetic algorithm to identify the conformation associated with the low potential energy, molecular dynamics simulations to score the binding based on the force fields, classification and regression using machine learning, etc., are used as they are unique in their performance. To score the binding poses, mathematical functions are being used to predict the binding affinity, approximately [3]. For example, an empirical scoring function has the form:

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond}\Sigma_{il} g_1(\Delta r) g_2(\Delta\alpha) + \Delta G_{metal}\Sigma_{aM} f(r_{aM}) + \Delta G_{lipo}\Sigma_{lL} f(r_{lL}) + \Delta G_{rot} H_{rot}$$

Where the metal binding, hydrogen bond geometry, lipophilicity and bond rotations are included in the binding energy calculation [4].

1.3. Molecular Dynamics Simulations

Molecular dynamics (MD) simulation is a method to compute the physical changes in the atoms of molecules. This technique has potential applications in structure-based drug design as it efficiently samples the biological conformations of macromolecules such as proteins, enzymes, carbohydrates and nucleic acids, as well as the complex systems. The basis for the MD simulation is Newton's second law of motion ($F=ma$) to compute the atomic motions and it has the form:

$$\frac{\partial^2 x_i}{\partial t^2} = \frac{F x_i}{m_i}$$

where, Fx_i is the force and m_i is the mass along the x_i coordinate. The study conducted by McCommon *et al.,* was the first report on the application of MD simulations in the protein dynamics [5] which was preceded by the first simulation report based on the simulation of solid spheres moving at a constant velocity [6]. Later on, simulation results of proteins and other biological molecules were reported with their movements observed over time [7].

The empirical force field terms assume that atoms or particles

i to have their own arrangement of electrons and nuclei,

ii to be spherical with their atomic radii calculated experimentally or theoretically,

iii to have net charge and

iv to interact with each other and their interactions are treated with springs and potentials.

AMBER: Initially, Kollman and co-workers [8] introduced the “united atom” force field and embedded with AMBER simulation package. The "all-atom" version Amber *ff86* [9] was later introduced with effective transferability. With improved parameterization, ff99, ff02, ff03 versions were developed [10]. The atomic charges derived from the basis set, B3LYP/cc-pVTZ//HF/6-31G* and corrections in the polarization for intra-molecular interactions [11] were implemented. Additionally, parameters for organic compounds (General amber force field, *GAFF*), for phospholipids (*lipid14*), etc. [12] were also introduced.

CHARMM is a widely used force field developed by Karplus and co-workers for simulation of biomolecules. It has many versions, including the CHARMM22 [13] and CHARMM/CMAP [14] for protein simulation in explicit and implicit water environments, respectively. CHARMM27 [15] was developed for lipids and nucleic acids. The CHARMM36 [16] and its updates are the recent versions available.

GROMOS is available in both united and all-atom versions. The earlier versions GROMOS87+ and GROMOS96 [17] are widely used for protein simulations. Further, the parameterization process resulted in newer versions such as 53A6, 54A7,54B7, etc., with improved parameters for secondary structure treatments, phospholipid head groups, coarse-grained models, etc. [18].

OPLS is available in the united atom as well as in all-atom versions developed by Jorgensen *et al.,* initially for hydrocarbons [19]. Further development based on both short and long hydrocarbons provided support for simulation biomolecules in the membrane phospholipid environment. Here, the energy profiles were calculated for hydrocarbons using MP2/aug-cc-pVTZ basis set to improve the diffusion coefficient, viscosities and gauche-trans ratios.

1.4. Binding Free Energy Calculations

Binding free energy (ΔG_{bind}) calculation is important to study the affinity of the compounds towards the target of interest. The relative binding free energy is particularly applied in studies based on the congeneric molecules. The molecular dynamics simulations and statistical mechanics help to sample the conformations and aid in binding free energy calculation to be accurate. The following expression denotes the ΔG_{bind}.

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw}$$

$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA}$$

where, ΔE_{MM} is the molecular mechanics energy, $-T\Delta S$ is conformational entropy, ΔG_{sol} is solvation free energy and $\Delta E_{internal}$ includes bond, angle and dihedral energies.

The Poisson Boltzmann (PB) or Generalized Born (GB) are the methods used to calculate the polar contributions, and the solvent-accessible surface area (SASA) [20] is used to calculate the non-polar contribution. The entropy term penalizes the steric clashes and it could be calculated using (i) normal mode analysis or (ii) Quasi-Harmonic (QH) [21] methods, the low-frequency motions through the approximation of mass-weighted covariance matrix, followed by sum of the vibrational frequencies calculated for each mode. OpenMM [22] and Gromacs [23] are the open-source programs available for simulation and binding free energy calculations while AMBER [24], CHARMM [25] and NAMD [26] are academic and AceMD and Desmond are the commercial packages.

1.5. QSAR & Toxicity Prediction

The quantitative structure-activity relationship (QSAR) is the technique that relates the descriptors of the training compounds to their biological activity. This has the relation: $v = f(p)$, where v denotes the activity and p denotes the probability function of the set of descriptors, which can be represented using Taylor's series expansion. QSAR is mainly used to lead optimization based on either the 2D descriptor or 3D descriptor features. The 3D-QSAR relies on the flexible ligand alignment followed by the molecular field analyses methods such as comparative molecular field analysis (CoMFA), similarity indices based (CoMSIA), receptor surface-based (RSA) and shape of the molecule (MSA). The force field-based calculations on the 3D structures of the training set compounds involve electrostatic and van der Waals contributions correlated by partial least squares to the experimental activity data.

On the other hand, the toxicity prediction gains importance in order to eliminate the selected candidate using molecular modeling techniques. Some of the important computational methods to predict the toxicity of the given compounds are listed below.

Tools to predict drug toxicity

DEREK Nexus: www.lhasalimited.org/products/derek-nexus.htm

ToxTree: toxtree.sourceforge.net

HazardExpert: www.compudrug.com/hazardexpertpro

TOPKAT: accelrys.com/products/collaborative-science/biovia-discovery-studio/qsar-admet-and-predictive-toxicology.html

CASE Ultra: www.multicase.com/case-ultra

OECD QSAR: www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm

Toxicology databases

SIDER: sideeffects.embl.de

Comparative Toxicogenomics Database: ctdbase.org

ACToR: actor.epa.gov

FDA Adverse Event: www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects

OpenTox: www.opentox.org

PharmGKB: www.pharmgkb.org

T3DB: www.t3db.ca

TOXNET: toxnet.nlm.nih.gov

ToxBank: www.toxbank.net

SuperToxic: bioinformatics.charite.de/supertoxic/

ACToR: actor.epa.gov/actor/home.xhtml

eTOX: www.etoxproject.eu

HSDB: toxnet.nlm.nih.gov/newtoxnet/hsdb.htm

Haz-Map: hazmap.nlm.nih.gov

CEBS: tools.niehs.nih.gov/cebs3/ui/

2. QSAR AND MACHINE LEARNING

In addition to the typical QSAR methods for lead optimization, machine learning (ML) methods equally play a crucial role in the classification of active compounds from the test sets. ML is a computational study to learn the characteristics of data in the absence of experimental evidence. It has the functionalities to extract the features pertinent to the

function and classify data accordingly. The data to be handled can vary and the method of application will also vary. Supervised and unsupervised learning algorithms are commonly applied methods. The logistic and linear regression models, Naive Bayes, classification and regression trees (CART), K-nearest neighbor (KNN) [27], support vector machine (SVM) [28], random forest (RF) [29], and artificial neural networks (ANNs) are some of the supervised learning methods while the hierarchical clustering, K-means, and principal component analysis (PCA) are examples for unsupervised learning methods. Fig. (**2**) depicts the steps involved in the application of machine learning algorithms.

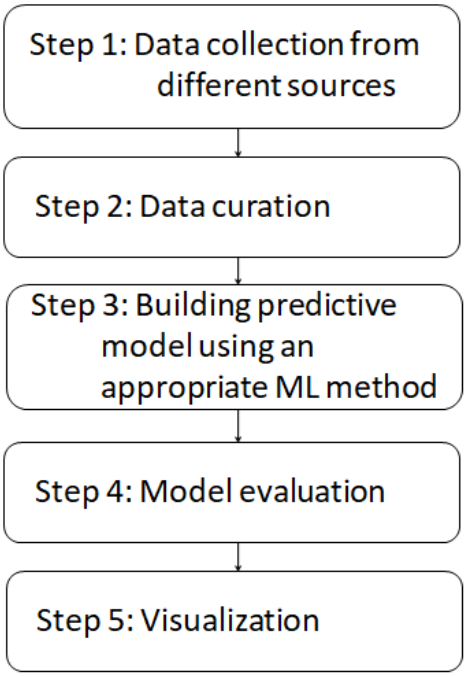


Fig. (2). Steps involved in applications of ML algorithms.

In structure-based drug design, the application of ML techniques can be understood from the literature [30]. An appropriate application of machine learning in target focussed library design will boost the successive structure-based as well as pharmacophore-based screening of large chemical libraries.

3. ARTIFICIAL INTELLIGENCE (AI) AND BIG DATA ANALYSES

Application of AI in drug discovery is at the stage of infancy now and it will become very effective as the data size keeps on increasing. The biological data is very complex in terms of the activity of the compound, say, a drug molecule. Here, the activity of the drug will be represented by many types of data heterogeneous in nature. For example, the affinity denoted by the Ki, IC_{50}, etc., and the PK/PD profile are very complex to inter-relate. Additionally, the protein interaction network, the pathway and expression data in response to drug treatment [31], disease classification, target and off-target binding, etc., are the factors that impose the complexity in multi-dimensions (Fig. **3**).

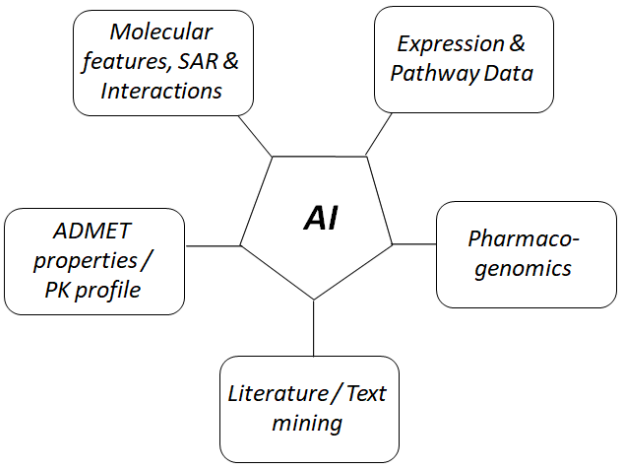


Fig. (3). Heterogeneous data involve in AI-driven SBDD.

The data set to prepare for the known and approved drugs itself involves 100-1000 data types that could cover the druggable space. Most of the time, this process will begin with the text mining of the literature database to explore the source data for further process. The integral part of ML or AI is the deep learning (DL) algorithm, which works based on the pattern matching or segmentations and is very effective when the data size is very large and when the other ML algorithms fail to handle. Hence, the big data analyses had emerged and some of the popular services available on Cloud are listed in Table **1** given below.

Table 1. List of cloud services for big data analyses in the area of precision medicine.

Service	Platform	URL
Software as a service (SaaS)	Amazon Web Services	www.aws.amazon.com
	Google Cloud Platform	cloud.google.com
	Microsoft Azure	azure.microsoft.com
	IBM Cloud	www.ibm.com/cloud/
	Alibaba Cloud	www.alibabacloud.com
Infrastructure as a service	DNAnexus	www.dnanexus.com
	Illumina Base Space Sequence Hub	basespace.illumina.com
	Seven Bridges	www.sevenbridges.com/platform
	Globus Genomics	globusgenomics.org

Further, the following section will focus on the recent findings through the application of advanced and coupled methods.

4. Identification of compounds against a target

With reference to the methods discussed above, ranging from the molecular modeling techniques to the ML algorithms, the structure-based drug design enters into the new regime with a combination of physics-based and computation-based methods. In this section, we emphasize that the conventional structure-based methods end up with more false positives as well as the failure of many candidates at different stages of the drug discovery pipeline. The new approaches address this issue in two ways: (i) drug repurposing and (ii) lead identification using ML and AI algorithms. In drug repurposing, identification of the new purpose of existing drugs is programmed using computational techniques. The drug interaction network and clinical data analyses are the key players in this domain. In the application of ML and AI algorithms to identify new chemicals with effective pharmacological activity, the conventional molecular modeling and structure-based drug design take part in the lateral stage. The AI and ML are most effective in the initial data set, such as large chemical databases. As a result, target-focussed ligand libraries are generated and successive application of SBDD will result in more promising hits with less false positives.

5. Identification of drugs from natural products

Until the colonial period (1800's) in India, Indian Medicine was completely dependent on medicinally important plants. Sometimes, the natural ones isolated were found to have serious side-effects and in these cases, modern science helped to prepare derivatives from them to overcome these side-effects. Some of the old drugs from natural products are shown in Table **2**.

Cragg and Newman (2016) reviewed the impact of medicinally important plants/natural compounds to treat various human ailments. This survey covered the period 1981-2014 [32]. Snafi from The qar University, Iraq has published a series of research papers on many medicinal plants and their effects on humans in curing various diseases. Most of these publications are reviews emphasizing the impact of herbal medicines on humans. Out of 175 small molecules with the anti-cancer activity which were introduced into therapy in western countries nearly 70 years back, around 49% were either directly obtained or were derived from natural products. Some of the important herbs included ginseng, Curcuma longa and Withania somnifera. Anti-cancer nature of some of the dietary phenols like quercitin, luteolin, genistein, apigenin and resveratrol had been discussed in many research publications. Combinatorial chemistry suppressed the impact of herbal medicine for sometimes. But recently, the focus has been switched over again to natural medicine as nature has been continuously carrying out its own version of combinatorial chemistry for 1000s of years and secondary metabolites are evolved in medicinal plants in response to the needs and challenges in the natural environment. In many cases, it has been found that nature and evolution have devised molecular structures that are far superior to even the best synthetic moieties in terms of diversity, specificity, binding efficiency and propensity in interacting with the biological targets. Compared to the compounds from combinatorial chemistry and synthetic drugs derived from natural substances, the drugs and products obtained from the natural sources exhibited more diverse and chemically complex structures. The natural product database contains a significantly larger number of scaffolds that exhibit higher structural novelty. Some of the scaffolds found naturally in the secondary metabolites cannot be synthesized at all. Herb-herb combination is used for therapeutic enhancement and advancement. Polypharmacy is a combined therapy leading to therapeutic benefit for a number of diseases. The main advantages of herbal drugs for many therapeutic activities is the involvement of low cost, complete accessibility, enhanced tolerance, more protection, high potency and efficiency with fewer side effects. At the same time, there are some drawbacks in the use of herbal medicine, namely, (i) scientific evidence for the clients by the herbalists are needed, (ii) standardization of herbal drug is needed, (iii) authenticated product test is needed, (iv) reliable clinical trials are needed and lastly, (v) reliable analytical methods are needed to reveal the phytoconstituents. As most of the herbal medicines being used now do not satisfy the above, these herbal formulations are sold as dietary supplements only without the need for the approval of FDA and some countries have formed regulations for using these herbal products. With the progress made in the areas of cellular biology, genomics and molecular mechanisms and also due to the vast development of the techniques and facilities in solving molecular structures, the number of druggable targets has increased. This allows screening for candidates of natural product libraries against an ever increasing number of potential molecular sites for therapeutic intervention. It has been reported that out of the 175 anticancer drugs developed and approved during 1940-2010 from western countries and Japan, 85 compounds (48.6%) were natural products or directly derived from natural products.

Suhitha *et al.,* detailed many herbs along with the docking results and these herbs are from the North-East region of India and are being used by the herbalists for controlling/curing various types of cancers, inflammation, diabetes and removal of kidney stones [33]. In one of the medicinal plants, *Stephania Hernandifolia* (SH), the major compound D-L.Tetrahydro palamatine (THP) was reported from the GC-MS analysis, with its presence in 57.9%. The 3-dimensional X-ray crystal structure of this compound was also elucidated and its docking with the cancer-related BCL-2 target (Pdb id: 2o2f) was compared with Fangchinoline, a compound isolated from the Chinese herb, *Stephania Tetrandra* [34] and being used for the treatment of breast cancer (Fig. **4**).

As a continuation of this work, Mohan *et al.,* carried out LCMS and LCMSMS studies and confirmed traces of two compounds whose molecular weights differed from THP isolated from *SH* (with the local name *Jabong*) [35]. These compounds were later identified as the one –OH and -2 OH group substitutions in the four -OCH3 groups containing THP (Fig. **5**). Having identified these traces and also based on the docking studies of THP using the Induced fit docking protocol in Glide module of Schrodinger LLC., which proved the anti-breast cancer nature, modeling studies were similarly carried out for -OH group substitution instead of every -OCH3 substitution in THP compound. Docking results showed better binding than THP compound when all

Table 2. List of few known natural products which had undergone chemical modifications to overcome their side effects.

Compounds Isolated	Natural Source	Clinical Use	Remarks
Paclitaxel	Pacific yew tree	Ovarian Cancer	Poor water solubility. Docetaxel is an analog with improved solubility
Morphine	Opium, Poppy plant	Analgesic and sedative effects	-
Salicin	Willow trees	Pain relief	Acetyl salicylic acid (Aspirin) is the synthetic derivative, causes gastric complications.
Berberine	Berberi fermontii	Anti-microbial activity	100 fold more active in combination with methoxyhydrocarpin (an inactive compound from the same plant)
Ginsenoide	Ginseng	Anti-breast cancer and anti-diabetic activity	Cardiovascular dysfunction
Aristolochic acid	Aristolochia	Anti-inflammatory activity	Renal and kidney failure
Camptothecin	Camptotheca acuminata	Anti-ovarian and small cell lung cancer	Bladder toxicity
Toptecan & Irinotecan	Semisynthetic derivative of Camtothecin	Anti-varian and colon cancers	
Fatty acids	Sterculiafoetida L. Seeds	Anti-oxidant, anti-microbial	-
Podophyllotoxin	Podophyllum species	-	Anologs: Etoposide
Vinblastine, Vincristine and their semi-synthetic analogs vinorolbine and vindesine.	Cantheranthus roseus	Anti-bladder and breast cancers and leukemia	Vindesine is the chemically modified one.
Saponins, Sapogenol A,B Naringin, rutin, baicalin	Soybean	Anti-microbial and immunomodulator	Colon cancer
Crude Extract from Ethylacetate	Garcinia mangostana L	Anti-inflammatory activity	Xanthones are major constituents
Areal part	Caralluma diffusa	Anti-oxidant and anti-viral	-
Silimarin	Seeds of Silibum maranum	Anti-viral activity	-
Leave juice	Moringa	Blood pressure, head and tooth aches	-

the four -OCH3 groups were substituted by four -OH groups. A simple reaction of THP compound experimentally converted THP into the above new compound, namely, tetrahydroprotoberberine (THPB). In the cell line studies, THP showed good IC_{50} for colon, lung and breast cancers. The cell line studies confirmed IC_{50} of 10.5 µg/mL for the breast cancer cell line MCF-7 and IC_{50} of 27 µg/mL for the lung cancer cell line A-549. For the colon cancer cell line HCT -116, the IC_{50} was 17 µg/mL for THP and for the four -OH substituted compound, Tetrahydroprotoberberine the IC_{50} was 7 µg/mL suggesting that the new compound, tetrahydroprotoberberine is the best for colon cancer. THP is found to be good for treating lung and breast cancers [36].

Table **3** shows better docking score and glide energy for the newly obtained compound, tetrahydroprotoberrerine (THPB) compared to both stepholidine (two -OH group substitutions) and the native ligand, 3-Hydroxybenzisoxazole suggesting that THPB compound is a better candidate as an anti-breast cancer compound. All the hydrophobic interactions of the native compound are also found with the binding of THPB compound along with the better hydrogen bonds. Fig. (**5**) shows the binding of the THPB compound at the active site of the human 3-phosphoglycerate dehydrogenase enzyme along with the co-crystal compound and also with stepholidine.

Mohan *et al.,* also identified the major presence of two compounds from the mixtures of two herbal formulations being used by a herbalist in Meghalaya for curing breast cancer (Mohan et.al., 2018b). These compounds were found to inhibit the TNFα and NFkB targets of inflammation. Five compounds from the herb 'Sivakarandai', being used in south India to control various types of cancers, have so far been identified and reported [37]. Details of one of these compounds, Decahydro-6-(imonomethyl)-4a-methylnaphthalen-

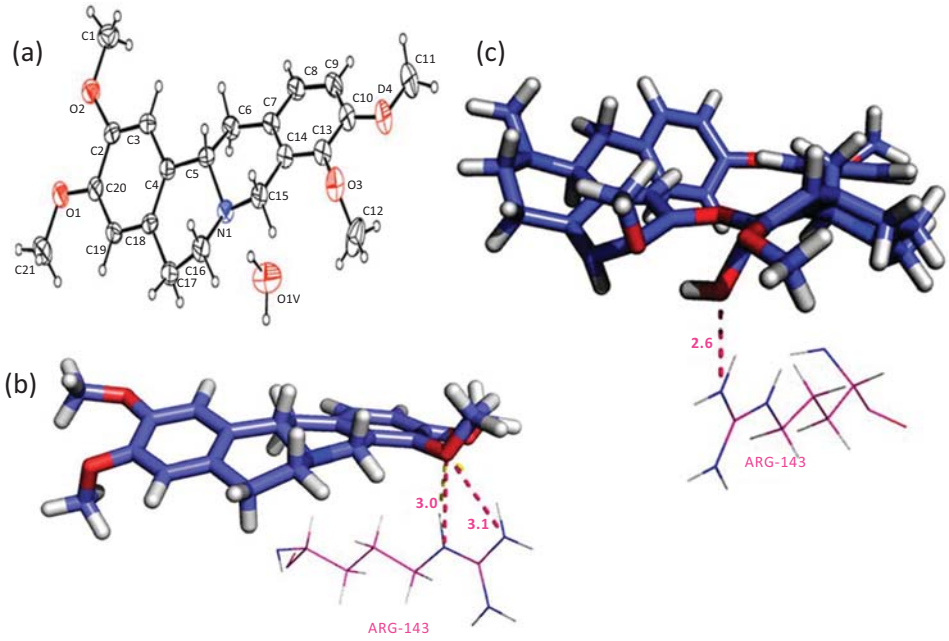
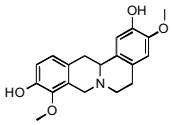
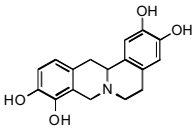
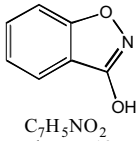


Fig. (4). ORTEP diagram of THP (**a**) and its interactions with the breast cancer target, BCL-2, PDB id: 2O2F (**b**). (**c**) shows the binding of fangchinoline isolated from *Stephania Tetrandra*.

Table 3. Results of docking studies with Breast Cancer Target PDB ID: 5NZP.

Ligand	Docking Score (kcal/mol)	Glide Energy (kcal/mo)	Interaction	Distance Å D…A	Hydrophobic Interaction
 Stepholidine	-6.724	-38.241	-	-	Leu151, Tyr174, Pro176, Leu193, Trp197, Pro208, Leu210 and Leu216
 Tetrahydroprotoberberine (MW: 299Da)	-6.824	-41.946	(O-H...O)Asp175 Ser212(O-H...O) (O-H...O)Ser212	3.06 2.94 2.98	Leu151, Leu153, Tyr174, Pro176, Ile177, Leu193, Pro208, Leu210 and Leu216
 $C_7H_5NO_2$ Mol. Wt.: 135 Native benzo[d]isoxazol-3-ol	-6.256	-26.262	Asp175(O-H...O)	3.15	Leu151, Leu153, Tyr174, Pro176, Ile177, Pro208, Leu210 and Leu216

2-ol (MW 195 Da) were presented in a meeting in the USA [38, 39]. The crystallographic structure and anticancer activity of one of the compounds detected by us [37] were later published [40]. Modeling studies with the breast cancer target phosphoglycerate dehydrogenase (PDB ID 5NZP) showed that one compound with molecular weight 516 Da had better binding than the co-crystal ligand and also compared to the two isomers (MW 195 Da) and the two other compounds identified [37].

Subasri *et al.,* (2016) examined the phytoconstituents from four herbs (being used as folk medicine) and their use as antidotes for snake bites [38].

There are many more examples validating the use of herbs to cure/control various human ailments, but these herbs should be used after consulting with experienced herbalists as there may be some side effects due to herb-herb interactions also.

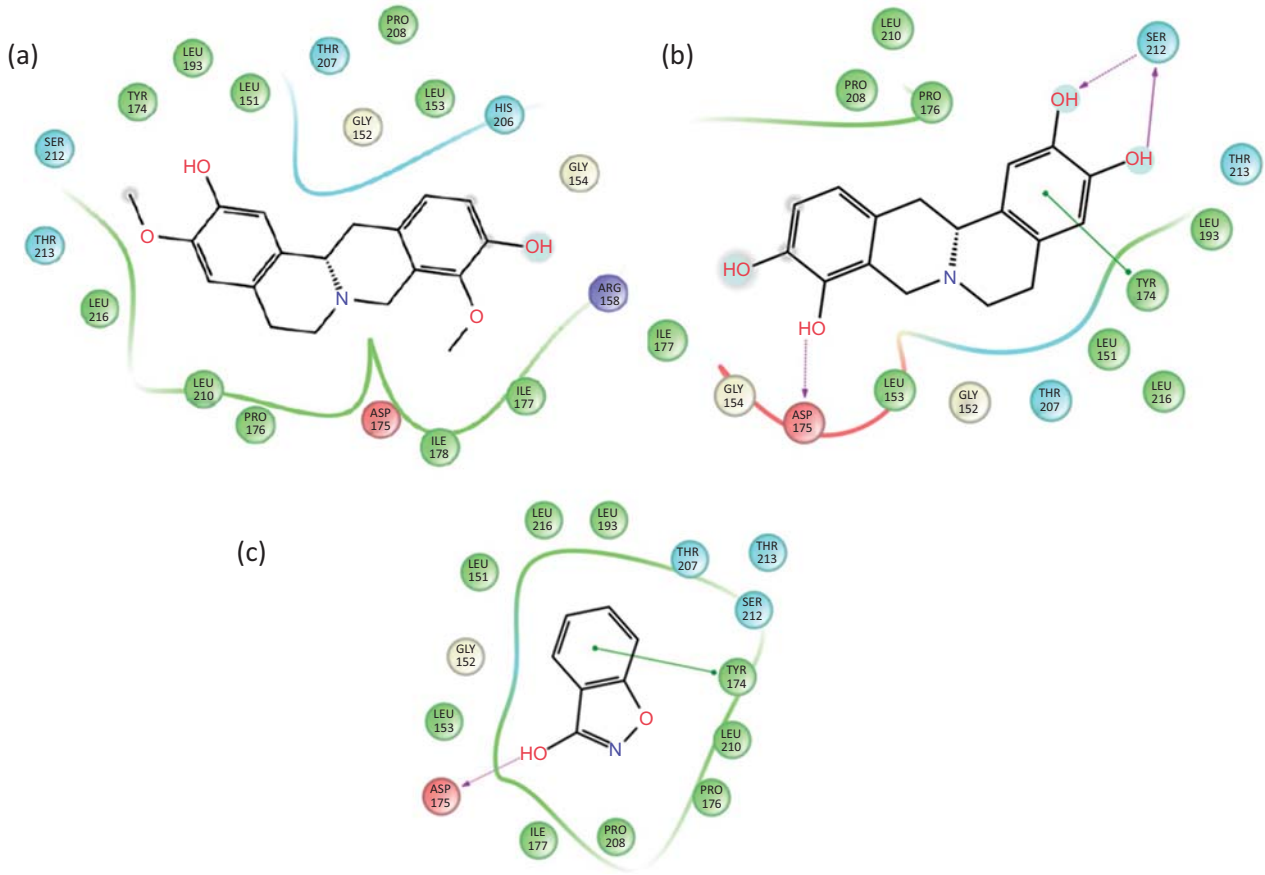


Fig. (5). The ligand interaction diagram of (**a**) Stepholidine, (**b**) tetrahydro protoberberine and (**c**) the cocrystallized ligand, 3-Hydroxybenzisoxazole with Breast Cancer Target, human 3-phosphoglycerate dehydrogenase (PDB ID: 5NZP).

CONCLUSION

Continuous growth in the biological data such as protein and chemical structure data, biochemical data related to catalytic/inhibitory activities, *etc.,* in parallel to the growth of pharmacological/clinical data provides new avenues for advanced drug discovery and development of low-risk compounds. But, the conventional methods do not handle all these data together and particularly have limitations to make use of clinical data in SBDD. Hence, there is an increased interest in the application of AI in drug discovery as it paves the path to access very complex data, which is not feasible for humans or a routine method to handle. This includes many divisions such as biomarkers discovery, drug repurposing, identification of new chemical entities, generation of data sets and predictive models, drug-drug interactions, etc. As of now, nearly 150 pharma startups initiated AI-driven drug discovery projects and it will reach 200 very soon. As medicinal plants and their phytoconstituents are reported as a cure for many human diseases, it is the right time to validate these herbs and their products so that they can also be exported. Attempts have to be made to overcome the losses of the valuable natural source due to deforestation, environmental pollution and global warming. One has to be very careful while choosing the herbs for treatment as herb-pharma drug interactions and also herb-herb interactions are already reported. One must also be very careful in identifying the level of heavy metals in the herbal preparations, if added any. With the upcoming research and development in science and technology, improvements in the quality, efficacy and safety of herbal medicines are immediately needed. Regulatory information like indications, contents, precautions, usage details, particular side effects if any, storage, etc., should be made available for all herbal products. Toxicity should be completely removed in the preparation of herbal drugs. Safety, efficiency, quality control and compatibility with other medicines are also to be cautiuosly observed while using herbal medicines. Computational methodologies, computer-aided drug design, bioinformatics tools, machine learning, along with modern wet-lab technologies will soon give a big boom to herbal medicines as these have been in practice for thousands of years.

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] (a) Jones, D.T.; Taylor, W.R.; Thornton, J.M. A new approach to protein fold recognition. *Nature,* **1992**, *358*(6381), 86-89. http://dx.doi.org/10.1038/358086a0 PMID: 1614539
(b) Bowie, J.U.; Lüthy, R.; Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science,* **1991**, *253*(5016), 164-170. http://dx.doi.org/10.1126/science.1853201 PMID: 1853201

[2] (a) Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.,* **1996**, *6*(3), 402-406. http://dx.doi.org/10.1016/S0959-440X(96)80061-3 PMID: 8804827
(b) Morris, G.M.; Lim-Wilby, M. Molecular docking. *Methods Mol. Biol.,* **2008**, *443*, 365-382. http://dx.doi.org/10.1007/978-1-59745-177-2_19 PMID: 18446297

[3] Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? *J. Chem. Inf. Model.,* **2009**, *49*(6), 1568-1580. http://dx.doi.org/10.1021/ci8004308 PMID: **1944**5499

[4] Eldridge, M.D.; Murray, C.W.; Auton, T.R.; Paolini, G.V.; Mee, R.P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.,* **1997**, *11*(5), 425-445. http://dx.doi.org/10.1023/A:1007996124545 PMID: 9385547

[5] McCammon, J.A.; Gelin, B.R.; Karplus, M. Dynamics of folded proteins. *Nature,* **1977**, *267*(5612), 585-590. http://dx.doi.org/10.1038/267585a0 PMID: 301613

[6] Alder, B.J.; Wainwright, T.E. Phase transition for a hard sphere system. *J. Chem. Phys.,* **1957**, *27*(5), 1208-1209. http://dx.doi.org/10.1063/1.1743957

[7] Karplus, M.; McCammon, J.A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.,* **2002**, *9*(9), 646-652. http://dx.doi.org/10.1038/nsb0902-646 PMID: 12**1984**85

[8] Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.,* **1984**, *106*(3), 765-784. http://dx.doi.org/10.1021/ja00315a051

[9] Weiner, S.J.; Kollman, P.A.; Nguyen, D.T.; Case, D.A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.,* **1986**, *7*(2), 230-252. http://dx.doi.org/10.1002/jcc.540070216 PMID: 29160584

[10] (a) Fox, T.; Kollman, P.A. Application of the RESP methodology in the parametrization of organic solvents. *J. Phys. Chem. B,* **1998**, *102*(41), 8070-8079. http://dx.doi.org/10.1021/jp9717655
(b) Bayly, C.I.; Cieplak, P.; Cornell, W.; Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.,* **1993**, *97*(40), 10269-10280. http://dx.doi.org/10.1021/j100142a004
(c) Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Kollmann, P.A. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.,* **1993**, *115*(21), 9620-9631. http://dx.doi.org/10.1021/ja00074a030
(d) Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A Second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.,* **1995**, *117*(19), 5179-5197. http://dx.doi.org/10.1021/ja00124a002

[11] (a) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.,* **2003**, *24*(16), **1999**-2012. http://dx.doi.org/10.1002/jcc.10349 PMID: 14531054
(b) Cieplak, P.; Caldwell, J.; Kollman, P. Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. *J. Comput. Chem.,* **2001**, *22*(10), 1048-1057. http://dx.doi.org/10.1002/jcc.1065
(c) Wang, J.; Cieplak, P.; Kollman, P.A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.,* **2000**, *21*(12), 1049-1074. http://dx.doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F

[12] (a) Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.,* **2004**, *25*(9), 1157-1174. http://dx.doi.org/10.1002/jcc.20035 PMID: 15116359
(b) Dickson, C.J.; Madej, B.D.; Skjevik, A.A.; Betz, R.M.; Teigen, K.; Gould, I.R.; Walker, R.C. Lipid14: the amber lipid force field. *J. Chem. Theory Comput.,* **2014**, *10*(2), 865-879. http://dx.doi.org/10.1021/ct4010307 PMID: 24803855

[13] MacKerell, A.D.; Bashford, D.; Bellott, M.; Dunbrack, R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B,* **1998**, *102*(18), 3586-3616. http://dx.doi.org/10.1021/jp973084f PMID: 24889800

[14] Mackerell, A.D., Jr; Feig, M.; Brooks, C.L., III Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.,* **2004**, *25*(11), 1400-1415. http://dx.doi.org/10.1002/jcc.20065 PMID: 15185334

[15] MacKerell, A.D., Jr; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers,* **2000**-**2001**, *56*(4), 257-265. http://dx.doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W PMID: 11754339

[16] Huang, J.; MacKerell, A.D., Jr CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.,* **2013**, *34*(25), 2135-2145. http://dx.doi.org/10.1002/jcc.23354 PMID: 23832629

[17] Stocker, U.; van Gunsteren, W.F. Molecular dynamics simulation of hen egg white lysozyme: a test of the GROMOS96 force field against nuclear magnetic resonance data. *Proteins,* **2000**, *40*(1), 145-153. http://dx.doi.org/10.1002/(SICI)1097-0134(20000701)40:1<145::AID-PROT160>3.0.CO;2-Y PMID: 10813839

[18] (a) Schmid, N.; Eichenberger, A.P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A.E.; van Gunsteren, W.F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.,* **2011**, *40*(7), 843-856. http://dx.doi.org/10.1007/s00249-011-0700-9 PMID: 21533652
(b) Lin, Z.; van Gunsteren, W.F. Refinement of the application of the GROMOS 54A7 force field to β-peptides. *J. Comput. Chem.,* **2013**, *34*(32), 2796-2805. http://dx.doi.org/10.1002/jcc.23459 PMID: 24122968
(c) Marzuoli, I.; Margreitter, C.; Fraternali, F. Lipid Head Group Parameterization for GROMOS 54A8: A consistent approach with protein force field description. *J. Chem. Theory Comput.,* **2019**, *15*(10), 5175-5193. http://dx.doi.org/10.1021/acs.jctc.9b00509 PMID: 31433640
(d) Eichenberger, A.P.; Huang, W.; Riniker, S.; van Gunsteren, W.F. Supra-atomic coarse-grained gromos force field for aliphatic hydrocarbons in the liquid phase. *J. Chem. Theory Comput.,* **2015**, *11*(7), 2925-2937.

http://dx.doi.org/10.1021/acs.jctc.5b00295 PMID: 26575730

[19] (a) Jorgensen, W.L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.,* **1988**, *110*(6), 1657-1666.
http://dx.doi.org/10.1021/ja00214a001 PMID: 27557051
(b) Jorgensen, W.L.; Maxwell, D.S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.,* **1996**, *118*(45), 11225-11236.
http://dx.doi.org/10.1021/ja9621760

[20] Connolly, M. Analytical molecular surface calculation. *J. Appl. Cryst.,* **1983**, *16*(5), 548-558.
http://dx.doi.org/10.1107/S0021889883010985

[21] Brooks, B.R.; Janežič, D.; Karplus, M. Harmonic analysis of large systems. I. Methodology. *J. Comput. Chem.,* **1995**, *16*(12), 1522-1542.
http://dx.doi.org/10.1002/jcc.540161209

[22] Eastman, P.; Swails, J.; Chodera, J.D.; McGibbon, R.T.; Zhao, Y.; Beauchamp, K.A.; Wang, L.P.; Simmonett, A.C.; Harrigan, M.P.; Stern, C.D.; Wiewiora, R.P.; Brooks, B.R.; Pande, V.S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Comput. Biol.,* **2017**, *13*(7), e1005659.
http://dx.doi.org/10.1371/journal.pcbi.1005659 PMID: 28746339

[23] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: fast, flexible, and free. *J. Comput. Chem.,* **2005**, *26*(16), 1701-1718.
http://dx.doi.org/10.1002/jcc.20291 PMID: 16211538

[24] Case, D.A.; Cheatham, T.E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.,* **2005**, *26*(16), 1668-1688.
http://dx.doi.org/10.1002/jcc.20290 PMID: 16200636

[25] Brooks, B.R.; Brooks, C.L., III; Mackerell, A.D., Jr; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A.R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R.W.; Post, C.B.; Pu, J.Z.; Schaefer, M.; Tidor, B.; Venable, R.M.; Woodcock, H.L.; Wu, X.; Yang, W.; York, D.M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.,* **2009**, *30*(10), 1545-1614.
http://dx.doi.org/10.1002/jcc.21287 PMID: 19444816

[26] Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.,* **2005**, *26*(16), 1781-1802.
http://dx.doi.org/10.1002/jcc.20289 PMID: 16222654

[27] Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.,* **1992**, *46*(3), 175-185.

[28] Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.,* **1995**, *20*(3), 273-297.
http://dx.doi.org/10.1007/BF00994018

[29] Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods,* **2009**, *14*(4), 323-348.
http://dx.doi.org/10.1037/a0016973 PMID: 19968396

[30] (a) Chiba, S.; Ikeda, K.; Ishida, T.; Gromiha, M.M.; Taguchi, Y.H.; Iwadate, M.; Umeyama, H.; Hsin, K.Y.; Kitano, H.; Yamamoto, K.; Sugaya, N.; Kato, K.; Okuno, T.; Chikenji, G.; Mochizuki, M.; Yasuo, N.; Yoshino, R.; Yanagisawa, K.; Ban, T.; Teramoto, R.; Ramakrishnan, C.; Thangakani, A.M.; Velmurugan, D.; Prathipati, P.; Ito, J.; Tsuchiya, Y.; Mizuguchi, K.; Honma, T.; Hirokawa, T.; Akiyama, Y.; Sekijima, M. Identification of potential inhibitors based on compound proposal contest: Tyrosine-protein kinase Yes as a target. *Sci. Rep.,* **2015**, *5*, 17209.
http://dx.doi.org/10.1038/srep17209 PMID: 26607293
(b) Ramakrishnan, C.; Mary Thangakani, A.; Velmurugan, D.; Anantha Krishnan, D.; Sekijima, M.; Akiyama, Y.; Gromiha, M.M. Identification of type I and type II inhibitors of c-Yes kinase using in silico and experimental techniques. *J. Biomol. Struct. Dyn.,* **2018**, *36*(6), 1566-1576.
http://dx.doi.org/10.1080/07391102.2017.1329098 PMID: 28589758

[31] (a) Li, J.; Zheng, S.; Chen, B.; Butte, A.J.; Swamidass, S.J.; Lu, Z. A survey of current trends in computational drug repositioning. *Brief. Bioinform.,* **2016**, *17*(1), 2-12.
http://dx.doi.org/10.1093/bib/bbv020 PMID: 25832646
(b) Delavan, B.; Roberts, R.; Huang, R.; Bao, W.; Tong, W.; Liu, Z. Computational drug repositioning for rare diseases in the era of precision medicine. *Drug Discov. Today,* **2018**, *23*(2), 382-394.
http://dx.doi.org/10.1016/j.drudis.2017.10.009 PMID: 29055182

[32] Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.,* **2016**, *79*(3), 629-661.
http://dx.doi.org/10.1021/acs.jnatprod.5b01055 PMID: 26852623

[33] Suhitha, S.; Devi, S.K.; Gunasekaran, K.; Pakyntein, H.C.; Bhattacharjee, A.; Velmurugan, D. Phytochemical analyses and activity of herbal medicinal plants of North- East India for anti-diabetic, anti-cancer and anti-tuberculosis and their docking studies. *Curr. Top. Med. Chem.,* **2015**, *15*(1), 21-36.
http://dx.doi.org/10.2174/1568026615666150112104344 PMID: 25579573

[34] (a) Xing, Z.; Zhang, Y.; Zhang, X.; Yang, Y.; Ma, Y.; Pang, D. Fangchinoline induces G1 arrest in breast cancer cells through cell-cycle regulation. *Phytother. Res.,* **2013**, *27*(12), 1790-1794.
http://dx.doi.org/10.1002/ptr.4936 PMID: 23401195
(b) Sun, Y. F.; Wink, M. Tetrandrine and fangchinoline, bisbenzylisoquinoline alkaloids from Stephania tetrandra can reverse multidrug resistance by inhibiting P-glycoprotein activity in multidrug resistant human cancer cells *Phytomedicine,* **2014**, *21*(8-9), 1110-1119.

[35] Mohan, K.; Pakyntein, H.C.; Bhattacharjee, A.; Viswanathan, V.; Velmurugan, D. Discovery of novel-anti-cancer compound from the identified phytoconstituent of "Jabung" an herbal medicinal plant. *RJBPCS,* **2018**, *4*(6), 612-628.

[36] Mohan, K.; Marthong, B.; Atanu, B.; Wadhwani, A.; Gayathri, D.; Velmurugan, D. Identification and *in-silico* analysis of anti-cancer compounds from herbal mix of North-East India. *RJLBPS,* **2018**, *4*(5), 485-497.

[37] Mohan, K.; Rangasamy, K.; Viswanathan, V.; Velmurugan, D. Identification of potential anti-cancer leads from Sivakanthai-Sphaeranthus Amaranthoides- Using tandem mass and their *in silica* studies. *RJBPCS,* **2018**, *4*(6), 335-356.

[38] Subasri, S.; Viswanathan, V.; Manish, K.; Velmurugan, D. Phytochemical analysis, molecular docking and molecular dynamics simulations of selected phytoconstituents from four herbs as anti-dotes for snake bites. *Clin Proteom Bioinform,* **2016**, *1*(3), 1-13.
http://dx.doi.org/10.15761/CPB.1000117

[39] Suhitha, S.; Mohan, K,; Rampriya, U.; Manohar, V.; Kesavan, M.; Rangasamy, KS.; Velmurugan, D.; In: *Tandem mass spectrometry enables characterization of the major phytoconstituents of "Sivakaranthai"- Sphaeranthus Amaranthoides - An Indian rejuvenator herb.* Proceedings of 62nd ASMS Conference on Mass Spectroscopy and Allied Subjects, June 15-19 , 2014. Baltimore, Maryland, USA - WP 328.

[40] Gayatri, S.; Suresh, R.; Reddy, CU.; Chitra, K. Isolation and characterization of chemopreventive agent from sphaeranthus amaranthoides. *Burm. F. Phcog. Res.,* **2016**, *8,* 61-65.