International Journal of Biotechnology and Biomedical Sciences
Volume 1, Number 1; July- December, 2015 pp. 16-20

http://www.krishisanskriti.org

In-silico Structure based, QSAR and Analogue based Studies using Dipeptidyl Peptidase 4 (DPP4) Inhibitors against Diabetes Type-2

Uma Anusha Nukala[1], P. Sahithi[2] and P. Raja Rao[3]

[1]M.Tech Final Year (Biochemical Engineering & Biotechnology), Osmania University, College of Technology
[2]Chairman BOS Biotechnology, Osmania University, College of Technology
[3]Osmania University, College of Technology
E-mail: [1]umaanusha@gmail.com, [2]sahithi.c@gmail.com, [3]prodokku_rr@yahoo.com

Abstract—*A new class of oral hypoglycemics called Dipeptidyl peptidase-4 inhibitors (DPP-4 inhibitors) block the enzyme Dipeptidyl peptidase-4 (DPP4), an intrinsic membrane glycoprotein and a serine exopeptidase that cleaves X-proline dipeptides from the N-terminus of polypeptides, by inhibiting the action of the enzyme, thus providing a useful treatment to diabetes mellitus type 2. The In-silico studies of DPP-4 inhibitors, that involved Docking studies like LIGAND FIT, C-DOCKER, LIB-DOCK and LUDI; Analogue based studies using pharmacophore generation; and finally QSAR studies, were found to be highly promising in further improvement and development of new lead compounds.*

1. INTRODUCTION

1.1 Drug discovery

Traditionally, Drug discovery and development is an intense, lengthy and an interdisciplinary endeavor involving target and lead discovery, followed by lead optimization, pre-clinical in-vitro and in-vivo studies prior to clinical development that takes approximately 12-14 years and costing up to $1.2 - $1.4 billion dollars, for the pharmaceutical industry.

In contrast to traditional methods that has resulted in high attrition rates with failures attributed to poor pharmacokinetics (39%), lack of efficacy (30%), animal toxicity (11%), adverse effects in humans (10%) and various commercial and miscellaneous factors, the process of drug discovery has been revolutionized with the advent of genomics, proteomics, bioinformatics and several efficient technologies like In-silico, rational, ligand based, structure based and analog based drug design with whose aid the candidate drug compound is developed which ideally should be “drug-like” i.e., they should possess properties that are predicted to lead to oral bioavailability, adequate chemical and metabolic stability, and minimal toxic effects. One way of estimating drug likeness is Lipinski's Rule of Five.

Often, the challenge of drug discovery is not only to find the best molecule but also to filter out the bad candidate molecules that might fail further down in drug design pipeline which is overcome by the use of complementary experimental and informatics techniques which increases the chance of success in many stages of the discovery process at a quicker rate and a lower cost.

1.2 In-silico drug design

In-silico methods can help in identifying drug targets, analyze the target structures for possible binding/ active sites, generate candidate molecules, check for their drug-likeness, dock these molecules with the target, rank them according to their binding affinities, further optimize the molecules to improve binding characteristics via bioinformatics tools.

1.3 Computer-assisted drug design

With the aid of computational chemistry that includes molecular mechanics or molecular dynamics, Semi-empirical, Ab initio quantum chemistry methods or density functional theory, the Computer-assisted drug design is employed to discover, enhance, or study drugs of their conformations, most fundamentally the binding properties and related biologically active molecules. In practice it takes several iterations of design, synthesis, and testing before an optimal molecule is discovered where in the computational methods have accelerated discovery by reducing iterance providing more novel small molecule structures.

CADD and bioinformatics together are a powerful combination in drug research and development which offers a great deal of benefits such as cost savings, time-to-market and a deep insight that researchers acquire about drug-receptor interactions**.**

1.4 Rational drug design

Rational drug design begins with a hypothesis that modulation of a specific biological target may have therapeutic value where in a biomolecule is selected as a drug target on the basis

of two essential pieces of information. The first is the evidence that modulation of the target will have therapeutic value which comes from the disease linkage studies and the other is that the target is "druggable" meaning that it is capable of binding to a small molecule and that its activity can be modulated by the small molecule.

1.5 Ligand based drug design

Ligand-based drug design (or Indirect drug design) relies on knowledge of other molecules that bind to the biological target of interest. These other molecules may be used to derive a pharmacophore which defines the minimum necessary structural characteristics a molecule must possess in order to bind to the target.

1.6 Structure based drug design

Structure-Based Drug Design (or Direct drug design) relies on knowledge of the three dimensional structure of the biological target obtained through methods such as x-ray crystallography or NMR spectroscopy. Using the structure of the biological target, candidate drugs that are predicted to bind with high affinity and selectivity to the target may be designed using interactive graphics and the intuition of a medicinal chemist.

1.7 Analog based drug design

When the structure of the biological target is unknown, analog based drug design technique is exercised which uses pharmacophore maps and QSAR to identify or modify a lead in the absence of a known receptor structure. Require experimental (based) affinities and properties of a set of active compounds, for which the structures are known.

1.8 Introduction to protein

Dipeptidyl peptidase-4 (DPP4), also known as adenosine deaminase complexing protein 2 or CD26 (cluster of differentiation 26) is an intrinsic membrane glycoprotein and a serine exopeptidase that cleaves X-proline dipeptides from the N-terminus of polypeptides is encoded by the DPP4 gene in humans. It is a rather indiscriminate enzyme for which a diverse range of substrates are known. The substrates of CD26/DPPIV are proline (or alanine) containing peptides and include growth factors, chemokines, neuropeptides, and vasoactive peptides. DPP-4 plays a major role in glucose metabolism and is responsible for the degradation of incretins such as GLP-1. It is also associated with immune regulation, signal transduction and apoptosis. Furthermore, it appears to work as a suppressor in the development of cancer and tumors.

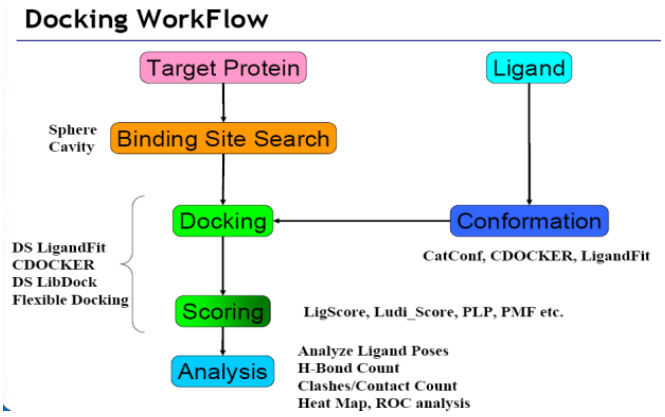
2. MATERIALS AND METHODS

2.1. Preparation of Bio Active Molecules

Bioactive compounds with the activity range 0.13 nM to 33800 nM were collected. Out of all the 120 GVKID compounds the best 10 were selected based on the optimum IC50 values and target selection was done. For this QSAR, pharmacophore and docking studies, the protein 1NU8 is loaded from RCSB protein data bank (www.rcsb.org/pdb/) and force field is applied. The RCSB PDB provides a variety of tools and recourses for studying the structures of biological macromolecules and their relationships to sequence, function and disease. So, based on the resolution and the interaction of the self-ligand in the rRNA, the PDB of 1NU8 has been selected for the study of structure based and analogue based drug design. Structure Based Design is a powerful method for rapidly identifying new lead compounds when a receptor structure is available. In the early stages of drug discovery, Virtual High Throughput Screening (VHTS) leads to efficiencies by helping to prioritize compounds in a library and by reducing library size. During the lead optimization stage, accurate docking methods, efficient de novo design methods, and accurate physics-based scoring can yield high-confidence compounds that are more likely to be active in-vivo. Structure-Based Drug Design in Discovery Studio includes: 1.A rational approach to flexible docking, 2.Fast and accurate protein ionization and pK Estimation, 3.Docking tools optimized for vHTS applications, 4.Industry-validated de novo ligand generation and optimization, 5.A comprehensive collection of ligand scoring tools. Structure-Based Drug Design is employed with the following parts: Ligand Fit, C-Docker, Lib Dock and De Novo Drug Designing.

2.2. Molecular Docking

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Two popular docking approaches exist. The conformational search approach uses a matching technique that describes the protein and the ligand as complementary surfaces. The second approach using Scoring methods simulates the actual docking process in which the ligand-protein pair wise interaction energies are calculated. These are of 3 types: Force field based, Empirical based and Knowledge based methods. Scoring is the process of evaluating a particular pose (candidate binding mode) by counting the number of favorable intermolecular interactions such as hydrogen bonds and hydrophobic contacts. Two docking methods available in Discovery Studio by Accelrys and used in the present study are Ligand fit and CDOCKER. These are summarized below:



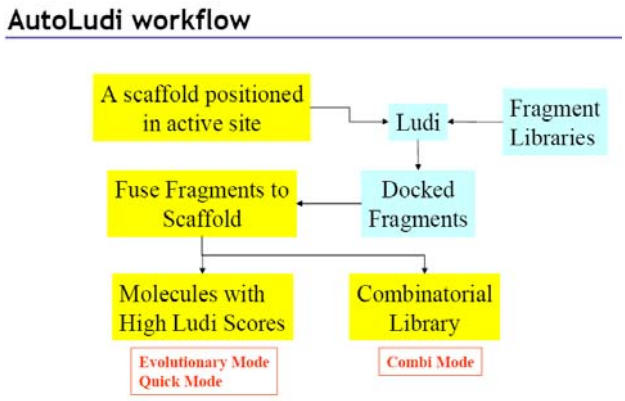


Fig. 1. Docking & Auto Ludi workflow

The ideal hip hop training set are as 2-30 compounds ideally 6 molecules, structurally diverse set of input molecules, feature rich compounds, include the most active compounds.

2.3. 3D -QSAR Pharmacophore Generation (HypoGen)

Hypogen attempts to derive SAR models for a set of molecules for which activity value (IC_{50} or K_I) on a given biological target are available. Hypogen optimizes hypothesis than that are present in the highly active compounds in the training set. But missing among the least active (or in actives) ones. It attempts to construct the simplest hypothesis that best correlates that activity (estimates vs. measured) the predicted models are created.

2.4. QSAR

The idea of quantitative structure-activity (or structure-property) relationships (QSAR/QSPR) was introduced by Hansch et al. in 1963 and was first applied to analyze the importance of lipophilicity for biological potency. This concept is based on the assumption that the difference in the structural properties of molecules, whether experimentally measured or computed, accounts for the difference in their observed biological or chemical properties.

Biological activity expressed as a reciprocal to produce a positive slope and also due to the inverse relationship between physiochemical property and biological potency. There is a positive relationship between the reciprocal of the biological activity (I/BA) and physiochemical property, because (I/BA) increases as the studies are based on the descriptors and biological activity relationship the biological activity data must be minimal and the choice of the descriptors of the descriptors must be accurate and appropriate.

2.4.1. Objective of QSAR

Drug transport/ mechanism, prediction of activity, classification of molecules as highly active, moderately active and inactive, optimization of activity by steric, electrostatic and hydrophobicity, refinement of synthetic targets, reduction and replacement of animals for the action of drugs. In our computational analysis, we have taken 41 molecules in the training set and 14 molecules in the test set. An attribute named Activ was added to both the sets and the activity values of the compounds were entered.

2.4.2. Basic Assumptions in QSAR studies

All analogues belong to generic series, all analogues exert same mechanisms of actions all analogues bind in a comparable manner, effects of iso-steric replacement can be predicted, binding affinity correlated to interaction energies, biological activities correlated to binding activity.

2.4.3. Steps in QSAR Studies

i) CSD database; ii) Molecular superimposition and alignment; iii) Choice of descriptors; iv) Statistical methods to evaluate and evolve QSAR equation; v) Validation.

2.4.4. Validation Methods

Once a regression equation is obtained it is important to determine its reliability and its significance. Internal validation uses the data set for which the model is derived and checks for internal consistency. The procedure derives a new model and is used to predict the activities of the molecules that were not included in the new model set. This is repeated until all compounds have been deleted and predicted once. Internal validation is less rigorous than external validation. External validation evaluates how well the equation generalization. The original data are divided into two groups, the training set and the best set. The training set is used to derive a model, and the mode is used to predict the activities of the test set numbers. The Calculated molecular properties protocol will calculate many properties or perform basic statistical and correlation analysis of the numeric properties as requested.

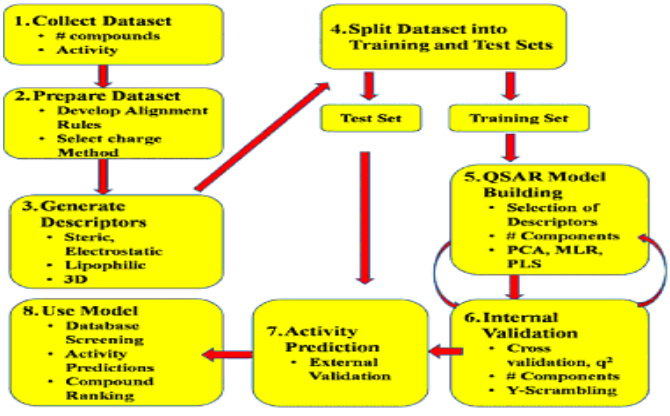


Fig. 2: Flowchart of QSAR procedure

3. RESULTS AND DISCUSSIONS

Further a graph plotted between log-Activity and log-Estimated activity of these molecules for ligand pharmacophore mapping gave an r^2 value of 0.856. These results confer that this hypothesis is a valid hypothesis.

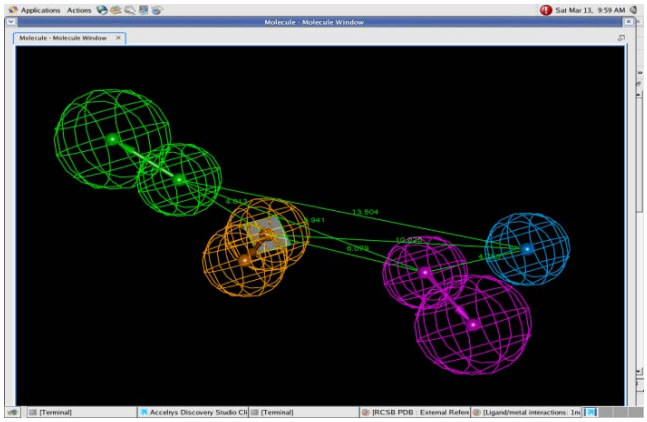


Fig. 3: Intradistances (A^0) between pharmacophore features in Hypogen

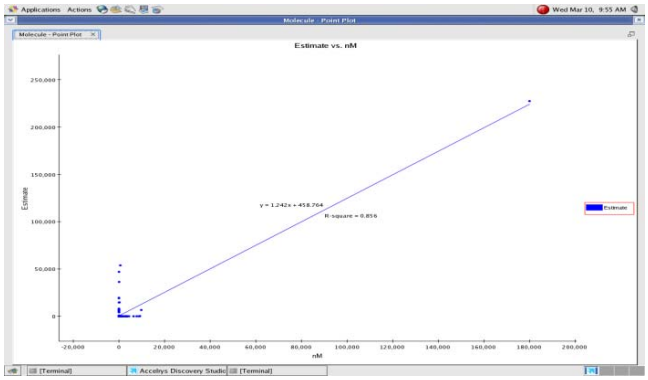


Fig. 4: Point plot representation of Hypogen test set

4. DISCUSSION

At the end of run, HypoGen generated 10 pharmacophore models. The best pharmacophore is taken as the hypothesis which has the highest cost difference, lowest error cost, lowest RMS difference and the best correlation coefficient. Pharmacophore models of *DPP 4* inhibitors are generated in HypoGen module in Discovery Studio software. Hypogen attempts to construct the simplest hypothesis that best correlates the activities (experimental vs. predicted).

The datasets was divided into training set and test set, considering both structural diversity and wide coverage of the activity range. The compounds with activity <1 nM were considered as highly actives (+++), compounds with an activity range of 10 -1100 nM as moderate actives (++) and activity of >1100 nM as least actives (+). At end of run, hypogen generated 10 pharmacophore models. All 10 hypotheses generated showed high correlation coefficient between experimental and predicted K_I values. It indicates that all the hypotheses are having true correlation between 80-95%.

5. TRAINING SET MOLECULES

The MLR method performs a search over the space of possible QSAR models using lack of fit (LOF) scores to estimate the fitness of each model. These models lead to the discovery of predictive QSAR equations. The following table contains summary statistics for the top 10 models built by the MLR learner.

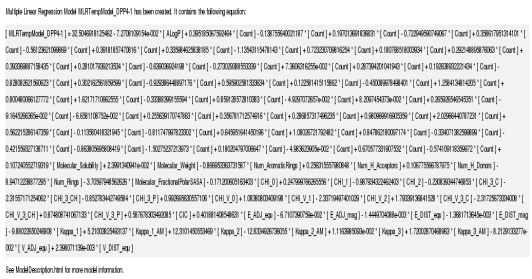


Fig. 5: r2 is the coefficient of determination and it is 0.915

From the above equations, the positive values are the reference for the presence of specific group at that point and increase the activity of molecule and the negative values indicate the presence of ionic group which reduce the activity.

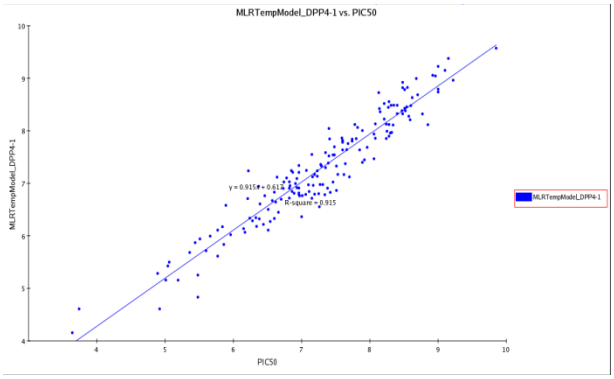


Fig. 6: Correlation between experimental and predicted activities by QSAR equation using MLR method

6. DISCUSSION

The work presented in ABDD (analogue based drug designing) study showed chemical features for a set of compounds along with their activities ranging over several orders of magnitude that can also be used to generate QSAR equation to successfully predict the activity. The models were not only predictive within the same series of training compounds but also diversified compounds of test set are also effectively mapped onto most of the features important for predicting activity. The equation identified for the antibacterial translation inhibitors can be used to evaluate how well any newly designed compound shows its biological activity before undertaking any further study including

synthesis. This application may help in identifying compounds for further biological evaluation and optimization.

The result generated from QSAR equation using MLR method, the values observed for r^2 are in specific range and there is a good correlation between experimental and MLR predicted activity as listed. Good correlation is observed between the experimental K_I and computational predicted K_I values. It has been suggested as since the predictive ability of equations is good, they can be used to develop new analogs.

7. CONCLUSION

As far as Insilco studies are concerned for DPP 4 the algorithms such as QSAR, Pharmacophore and docking were used. These algorithms showed good results.3D crystal structure of protein (1AVN) is used to understand the process of binding of its inhibitors by docking procedures Ligand Fit, C-Docker, Lib-Dock and Ludi. The force field models used for computing inter molecular interactions are sufficiently accurate and are compatible to allow the affinity of the ligand for anti- bacterial translation inhibitors which is computed with confidence. It suggests that structure based drug designing studies may be helpful in developing a new ligand molecule based on the interaction energies. The results of the works correlate with the true activity deciphered by the medicinal chemists. A MLR residual value confirms that there is not much difference in experimental activities and predicted activities. So these descriptors could predict the activity of the lethal factor inhibitors. The hydrophobic compound of a model increases the ability of the drug to pass through cell membranes. The QSAR equation generated for the TD inhibitor with new analogues possessing steric groups has shown better predictive activities. These steric groups may enhance the hydrophobic property of the molecule. The work in the catalyst presented in the study showed chemical features as hydrogen bond acceptor, hydrogen bond donor and hydrophobic. Hydrophobic aromatic and ring aromatic of set of compounds along with their activities ranging over several orders of magnitude can be used to generate pharmacophore hypothesis that can successfully predict the activity. The results obtained are used to develop new ligand molecules and find their activities Insilco and proving the same in accordance with the experimental values. Thus, the results reported can successfully employ in the rational drug designing of novel and potent lethal factor inhibitor.

REFERENCES

[1] Amit KT, Madhumita R, and Bhattacharya RK “Natural products as inducers of apoptosis: Implication for cancer therapy and prevention”.

[2] Ali IU, Schriml LM, Dean M (1999) Mutational spectra of PTEN/MMAC1 gene: a tumor suppressor with lipid phosphatase activity. J Natl Cancer Inst 91: 1922-1932.

[3] Bellacosa A, de Feo D, Godwin AK, Bell DW, Cheng JQ, et al. (1995) Molecular alterations of the AKT2 oncogene in ovarian and breast carcinomas. Int J Cancer 64: 280-285.

[4] Becker OM, Dhanoa DS, Marantz Y, Chen D, Shacham S, et al. (2006) An Integrated *in Silico* 3D Model-Driven Discovery of a Novel, Potent and Selective Amidosulfonamide 5-HT1A Agonist (PRX-00023) for the Treatment of Anxiety and Depression. J Med Chem 49: 3116-3135.

[5] Brooijmans N, Kuntz ID (2003) Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 32: 335-373.

[6] Carnero A, Blanco-Aparicio C, Renner O, Link W, Leal JF (2008) The PTEN/PI3K/AKT signalling pathway in cancer, therapeutic implications. Curr Cancer Drug Targets 8: 187-98.

[7] Chang RS, Ding L, Chen GQ, Pan QC, Zhao ZL, et al. (1991) Dehydroandrographolide succinic acid monoester as an inhibitor against the human immunodeficiency virus. Proc Soc Exp Biol Med 197: 59-66.

[8] Congreve M, Murray CW, Blundell TL (2005) Structural biology and drug discovery. Drug Discov Today 10: 895-907.

[9] Fang XK, Gao J, Zhu DN (2008) Kaempferol and quercetin isolated from Euonymus alatus improve glucose uptake of 3T3-L1 cells without adipogenesis activity. Life Sci 82: 615-22.

[10] Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, et al. (1995) Molecular Recognition of the Inhibitor AG-1343 by HIV-1 Protease: Conformationally Flexible Docking by Evolutionary Programming. Chemistry & Biology 2: 317.

[11] Gamet-Payrastre L, Manenti S, Gratacap MP, Tulliez J, Chap H, et al. (1999) Flavonoids and the inhibition of PKC and PI 3-kinase”. General Pharmacology 32: 279- 286.

[12] Gehlhaar DK, Bouzida D, Rejto PA (1999) Rational Drug Design: Novel Methodology and Practical Applications. American Chemical Society 292-311.

[13] Irwin JJ, Lorber DM, McGovern SL, Wei B, Shoichet BK (2002) Computational Nanoscience and Nanotechnology ISBN 0-9708275-6-3.