Academic Sciences International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 4, Issue 4, 2012

Research Article

VIRTUAL SCREENING OF HETEROBASED LIGAND LIBRARY FOR PROTEIN KINASE INHIBITOR FOR ANTICANCER ACTIVITY

ANAND M DAVE[1], VISHAL A MEVADA[2], NIKHIL S BHATT[1*], PRADIP B ACHARYA[1], RAJESH K PATEL[2]

[1*]Post Graduate Department of Microbiology, Gujarat Vidyapeeth, Sadra 382320, Dist. Gandhinagar, Gujarat, India, [2]Department of Life Sciences, Hemchandracharya North Gujarat University, Patan 384265, Gujarat. Email: bhattnikhil2114@gmail.com

Received: 16 April 2012, Revised and Accepted: 23 May 2012

ABSTRACT

Incorporating receptor flexibility is considered crucial for improvement of docking based virtual screening with an abundance of crystallographic structure are freely available, docking with multiple crystal structure is believed to be a practical approach to cope with protein flexibility.2SRC structures were first compared in a single structure docking by predicting the binding mode and recovering known ligands. Combination of different protein structures were compared by recovery of known ligands and an optional ensemble of 2SRC structure were selected. The chosen structure was used in virtual screening of over 2700 diverse compounds for 2SRC inhibitors. Six novel drugs ranked at the top of the hits list were tested experimentally proved. Further study indicated that achieving a better enrichment and identifying more diverse compounds was more likely using multiple structures than using only a single structure ever when protein structure were randomly selected. Taking into account conformational energy difference did not help to improve enrichment in the top ranked list.

Keywords: Virtual screening, 2SRC, Docking

INTRODUCTION

Cancer is a group of diseases in which there is an uncontrolled multiplication and spread of the body's own cells within body in abnormal forms. Cancer may affect almost any tissue of the body and may metastasize to other tissues within the body. If the spread is not controlled, it can result in death. According to the American Cancer Society's 'Global Cancer Facts and Figures 2007' and 'Cancer Facts and Figures 2008 cancer was the second leading cause of deaths after heart disease in developed countries and third leading cause of death in developing countries shown in Fig.. Over the last couple of decades, research was revealed considerable information about the molecular biology, pathobiochemistry and the intricate pathways involved in cancer (Ltadani *et al.,* 2008)

Cancer

Cancer (Medical Term: Malignant Neoplasm) is a disease that begins in the cells of the body. In normal situations, the cells grow and divide as the body needs them. In other words, Cancer is a group of diseases characterized by uncontrolled growth and spread of abnormal cells. This orderly process is disturbed when new cells form that the body was not needed and old cells don't die when they should.

Type of tumor

Benign tumor: They are non-cancerous, usually can be removed and generally don't grow back once they are removed. Moreover, benign tumor don't spread and it's rare for the benign tumor to be life threatening.

Malignant tumor: They are cancerous one which behave aggressively and attack the tissue surrounding them. Moreover, the cells can also jump from the malignant tumor and invade the blood stream or lymphatic system to form new tumor in other parts of the body. This process is known as Metastasis

Radiation: Up to 10% of invasive cancers are related to radiation exposure, including both ionizing radiation and non-ionizing radiation(Anand *et al.,* 2008).Additionally, the vast majority of non-invasive cancers are non-melanoma skin cancers caused by non-ionizing ultraviolet radiation.

Infection: Worldwide approximately 18% of cancers are related to infectious diseases (Hanahan and Weinberg, 2000). A virus that can cause cancer is called an oncovirus. These include human papilloma virus (cervical carcinoma), Epstein-Barr virus (B-cell lymph proliferative disease and nasopharyngeal carcinoma), Kaposi's sarcoma herpes virus (Kaposi's Sarcoma and primary effusion lymphomas), hepatitis B and hepatitis C viruses (hepato cellular carcinoma), and Human T-cell leukemia virus-1 (T-cell leukemia's).

Drug

A drug, broadly speaking, is any substance that, when absorbed into the body of a living organism, alters normal bodily function (World Health Organization, 1969). In pharmacology, a drug is "a chemical substance used in the treatment, cure, prevention, or diagnosis of disease or used to otherwise enhance physical or mental well-being (Drug Dictionary.com, 2007).

Drug discovery pipeline and *in silico* approach

In the fields of medicine, biotechnology and pharmacology, drug discovery is the process by which drugs are discovered or designed. In the past most drugs had been discovered either by identifying the active ingredient from traditional remedies or by serendipitous discovery

Target identification and validation *in silico*

Target identification and validation is the first key stage in the drug discovery pipeline. However, identification and validation of druggable targets from among thousands of candidate macromolecules is still a challenging task (Sheikh *et al.,* 2007).Numerous technologies for addressing the targets have been developed recently. Genomic and proteomic approaches are the major tools for target identification.

In silico ADMET (absorption, distribution, metabolism, excretion, toxicity)

Studies indicate that poor pharmacokinetics and toxicity are the most important causes of costly late-stage failures in drug development and it has become widely appreciated that these areas should be considered as early as possible in the drug discovery process. Combinatorial chemistry and high-throughput screening have significantly increased the number of compounds for which early data on absorption, distribution, metabolism, excretion (ADME) and toxicity (T) are needed.

Virtual screening

Virtual screening involves the docking of selected lead molecules against the biological target. This is followed by a scoring pattern. There is a number of software available for this. Some are commercially available and some are free to use (Amin Rostami-Hodjegan, Geoff Tucke, 2004).

Quantitative structure activity relationship (QSAR)

A major goal in *in silico* drug discovery research is to predict the behavior of new molecules using knowledge derived from the analysis of the properties of previously tested molecules. Origination of the modern QSAR formalism is attributed to the works of (Hansch and Fujita, 1964). The QSAR methodology is based on the concept that the differences observed in the biological activity of a series of compounds can be quantitatively correlated with differences in their molecular structure (Kubinyi, 1999).

Ligand library

Ligand library is the database of the chemical compounds having several calculations of chemical and biological properties of the compounds. Such libraries were very much useful for carryout research in the field of drug discovery. It provides baseline line data to generate pharmacophore, molecular descriptors for the QSAR studies and also docking studies related information. Currently there are several such chemical compound database is shown below.

Drug bank

The Drug Bank database, available at the University of Alberta, is a bioinformatics and cheminformatics resource that combines detailed data with comprehensive drug information (Wishart *et al*., 2008; Knox *et al*., 2011). The database contains nearly 4800 drug entries including: more than 1480 FDA-approved small molecule drugs, 128 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals, 3200 experimental drugs, 2500 protein (i.e., drug target, non-redundant) sequences are linked to these drug entries (Wishart *et al.,* 2006).

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 (Kitchen *et al.,* 2004). Docking is a process needed to predict how new hypothetical or existing compounds will bind to the protein.

Tyrosine kinase as a receptor

A tyrosine kinase is an enzyme that can transfer a phosphate group from ATP to a protein in a cell. It functions as an "on" or "off" switch in many cellular functions. The phosphate group is attached to the amino acid tyrosine on the protein. Tyrosine kinase are a subgroup of the larger class of protein kinase which attach phosphate groups to other amino acids (serine and threonine). Tyrosine Kinase (PTKs) have been found in the human genome. They are divided into two classes, receptor and non-receptor PTKs. At present, 58 receptor tyrosine kinases (RTKs) are known, grouped into 20 subfamilies. They play pivotal roles in diverse cellular activities including growth, differentiation, metabolism, adhesion, motility, death. Many RTKs are involved in oncogenesis, either by gene mutation or chromosome translocation or simply by over-expression. Tyrosine kinases are particularly important today because of their implications in the treatment of cancer. A mutation that causes certain tyrosine kinases to be constitutively active has been associated with several cancers. Imatinib (brand names Gleevec and Glivec) is a drug able to bind the catalytic cleft of this tyrosine kinase, inhibiting its activity. (Weinberg and Robert, 2004).

MATERIALS AND METHODS

Receptor preparation

Target selection

There are several well known anticancer targets are available right now; i.e.; vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), and platelet-derived growth factor (PDGF) etc. Tyrosine kinase inhibitors (TKIs) are the largest class of therapeutic agents in clinical use and under development that target angiogenesis.

Protein 3D structure

Protein databank is one of the best databases for protein 3d structure with experimentally confirmation. 3D protein structure file of Human Tyrosine-Protein Kinase C-SRC having PDBID: 2SRC was downloaded from PDB. This protein structure contains some of other interacting ions and metal ions. Such part is interfering with the docking procedure. We had removed these ions by using Molgrow molecular structure viewer software.

Ligand preparation

Heterobase (Heterocyclic chemical structure database)

Heterobase is a Heterocyclic chemical structure database of Gujarat with very extensive QSAR studies. Currently contains the 2700 compound were available in database. There were more than 60 different parameters were being checked during the submission of molecule. i.e.: Lipinsky rule of Five, ADME-TOX, carcinogenicity testing, mutagenicity and several more. After applying very specific and stringent condition to database only 197 compounds were passed for our study. We had downloaded all these compounds and converter then to pdb format using open babel software.

Docking

After both receptor and ligand molecules were prepared for the docking studies. The docking was done using patch dock, web server for small ligand-protein docking. Patch dock server only accepts the structure in pdb files only. Results of the docking were sent to the email of the corresponding submitter.

Result analysis based on the scoring

Score is the mathematical value of the docking results. It is combination of several parameters including free energy ΔG^0, RMSD value of the docking pose, hydrogen bonding etc. Binding patter of the drug binding is dependent on h-bond found in the final docked results. So interaction map of ligand and receptor was compared with the standard well known drugs. The complete procedure for the carry out the drug discovery study in current research study is shown below in flow chart like manner Figure- Fig 6.

RESULTS AND DISCUSSION

Drug like compound library (Heterobase)

It is generally recognized that drug discovery and development are very time and resources consuming processes. Virtual screening (*in silico*) technology is gaining increasing importance in the discovery process because it is a reliable and inexpensive method for identifying lead molecules. To facilitate the incorporation of the huge amount of the chemical data available in the field of synthetic medicinal chemistry of Heterocyclic, a database of heterocyclic compounds named HETEROBASE was created by Student of i-Life Academy that centralizes the structural, pharmacokinetic and toxic properties of 2700 compounds contained in the database

QSAR study

During the study of all the 2700 molecules of Heterobase, upon applying very stringent condition with Lipinky rule of five, ADME-TOX, carcinogenicity, BBB permeability, Mutagenicity etc. only 197 compound can passed out, which had like nature. In our current study these 197 molecules had been used for the docking procedure.

Ligand preparation

The smile file was extracted from the database for the all 197 compounds and structure was prepared for the running Patch dock. Patch dock required the PDB file as an input, So interconvnersion of the molecules from smile to pdb was done using Molegro Molecular viewer in Figure: 7.

Receptor selection

After carry out extensive literature survey for the best cancer target, we had found Tyrosine kinase as the best receptor in current time. Receptor tyrosine kinases (RTK)s are the high-affinity cell surface receptors for many polypeptide growth factors, cytokines, and hormones. It had been shown not only to be key regulators of normal cellular processes but also to have a critical role in the development and progression of many types of cancer. 3D structure for tyrosine kinase was downloaded from Protein data bank with

accession number2SRC. Below information indicated the detailed information related of this structure.

Molecule: TYROSINE-PROTEIN KINASE

Polymer: 1

Type:protein

Length: 452

Chains: A

EC#: 2.7.1.112

Fragment: RESIDUES 86-836, CONTAINING SH2, SH3, KINASE 2 DOMAINS AND C-TERMINAL TAIL

Organism: Homo sapiens

UniProtKB: P12931

Docking using Patch dock

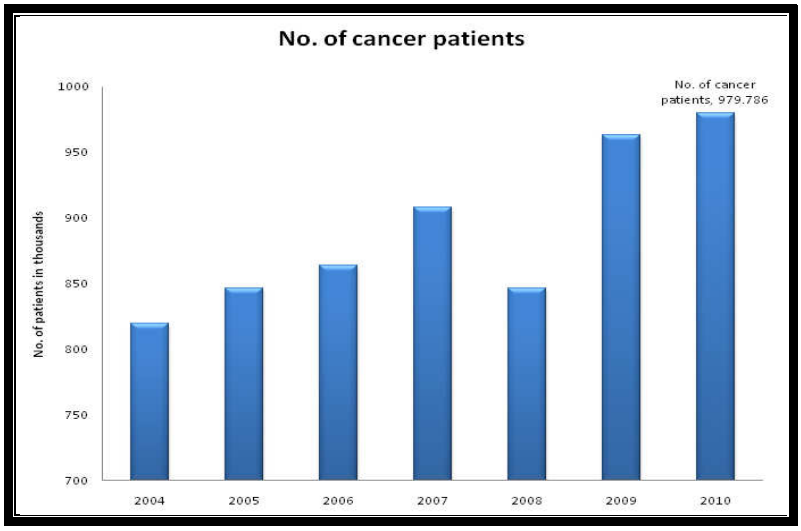
Patch dock is a tool for protein-protein, protein-small ligand binding study. It is only accept input files in pdb only. We had added previously prepared ligand and receptor file as an input and carry out the docking procedure with all the default parameter, except molecule type to small ligand- protein mode. The screenshot of front end was displayed in the Figure: 9.

Analysis

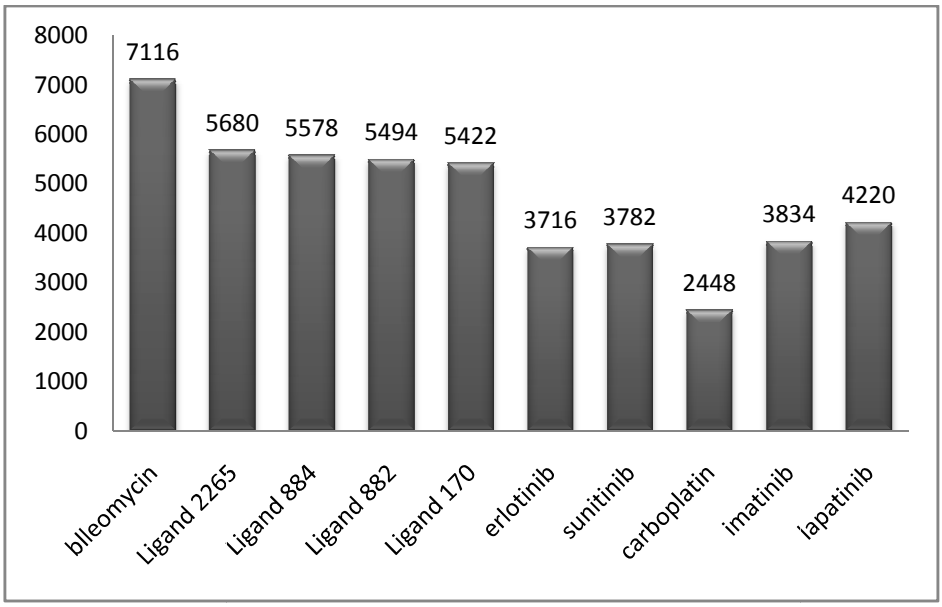
Result of Patch dock is based on delta G value, Number of H-bond, Van der Waals forces and stability of binding. But, for better understanding of such complex details, patch dock provides all in one result as a score value. Higher score value provide better results. After docking study of all 197 compounds and some standard drugs, Bleomycin had given highest score followed by ligand number 2265, 884, 882, 170, and 1527. As per the above listed list of Docking score and its comparison with the standard drugs, Ligand number 2265 is the best molecule which can work as the drug or lead molecule (Graph: 1)

Ligand 2265

The total chemo informatics based analysis was done using Neural Network based for ware name Sarchitech Miner.It provides several QSAR based analysis, detailed information related to the Ligand is available in Table: 1



Graph 1: Death in developing countries



Graph 2: Standard drug data

Table 1: Details List of cancers

S. No.	Type of cancer
1	Anal cancer
2	Basal cell carcinoma
3	Bladder cancer
4	Brain tumor
5	Breast cancer
6	Cervical cancer
7	Colon Cancer
8	Childhood cancers
9	Endometrial cancer
10	Oesophageal cancer
11	Gastric (Stomach) cancer
12	Gastric carcinoid
13	Head and neck cancer
14	Heart cancer
15	Kaposi sarcoma
16	Kidney cancer (renal cell cancer)
17	Leukaemia's
18	Lymphomas
19	Lung Cancer, Non-Small Cell
20	Lymphoma, AIDS-related
21	Melanoma
22	Myeloid Leukemia, Adult Acute
24	Non-small cell lung cancer
25	Oral Cancer
26	Ovarian germ cell tumor
27	Pancreatic cancer
28	Pharyngeal cancer
29	Prostate cancer
30	Rectal cancer
31`	Renal cell carcinoma (kidney cancer)
32	Sarcoma, Ewing family of tumors
33	Small cell lung cancer
34	Throat cancer
35	Thyroid cancer
36	Uterine sarcoma
37	Vaginal cancer
38	Vulvar cancer
39	Visual pathway and hypothalamic glioma, child hood
40	Waldenströmmacroglobulinemia

Table 2: Available web servers for docking

S. No.	Name	Algorithm	Limitations	License	Website
1	Rosetta dock	Monte carlo	Input is based on command line surface	Free	http://rosettadock.graylab.jhu.edu/
2	PRUNE and PROBE	Monte carloalgoritham	Input is based on command line surface	commercial	http://www.prune.server.jsp
3	FiberDock	Protein–protein docking algorithms and montecarlo.	Command line surface	commercial	http://www.scfbio-iitd.res.in/dock/fiberdock.jsp
4	AUDockervina LE:	Genetic algoritham	there is a limit on the number of atoms in the ligand	commercial	http://www.scripps.edu/pub/olson-web/doc/autodock/
5	Rosetta FlexPepDock web server	Genetic algoritham	Flexible docking	Free	http://rosettadock.ee.edu.tw/
6	MEDock	Monte carlo	Flexible docking	Free	http:// medock.ee.ncku.edu.tw/)
7	Meta dock	Genetic algorithm	Flexible docking		(http://dock.bioinfo.pl/metadock
8	Par dock	Monte Carlo	Flexible docking	Free	http://www.scfbio-iitd.res.in/dock/pardock.jsp

Table 3: Full report for Ligand 2265

Identifier	Analysis result
ID	2265
No	2265
Code	GP-100
SMILES	S=C(Nn3c(Cc2cccc1ccccc12)nnc3S)Nc4ccc(Cl)cc4
IUPAC Name	1-(4-chlorophenyl)-3-[3-mercapto-5-(1-naphthyl methyl)-4H-1,2,4-triazol-4-yl]thiourea
Molecular Formula	Molecular Formula = C20H16ClN5S2
Basic Moiety	Thiadiazoles(Thiouredo)
Structure	[H]Sc1nnc(n1N([H])C(=S)N([H])c2c([H])c([H])c(Cl) c([H])c2[H])C([H])([H])c3c([H])c([H])c([H])c4c([H]) c([H])c([H])c([H])c34
MW	425.9569
MLOGP	3.7098925
nHAcc	3
nHDon	2
No of Violations	0
Aquatic_Toxicity_Model	-1.4294249
Chromosomal_Aberration_Model	POS
pKi (FXa)	-1.2486961
Absorption Rate Constant	6.684911
Aqueous Solubility	-7.1371956
BBB Permeability	P
Bioavailability	47.654255
Half-life in Plasma	0.1586046
Plasma Protein Binding	99.603096
Volume of Distribution	0.60503334
Carcinogenicity Mouse Female	NEGATIVE
Carcinogenicity Mouse Male	NEGATIVE
Carcinogenicity Rat Female	NEGATIVE
Carcinogenicity Rat Male	NEGATIVE
hERG Channel Binding 50 uM	NONBINDER
Mutagen city	nonmutagen

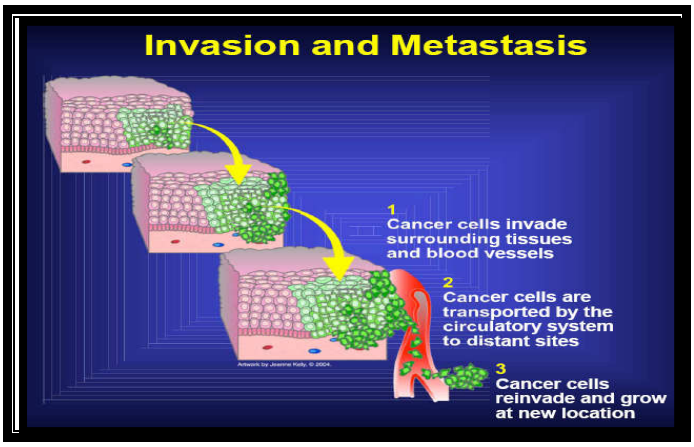


Fig. 1: Invasion and metastasis

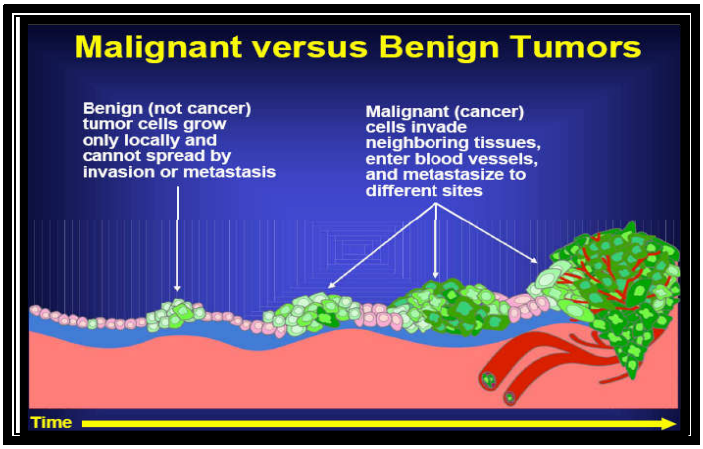


Fig. 2: Showing Invasion and metastasis

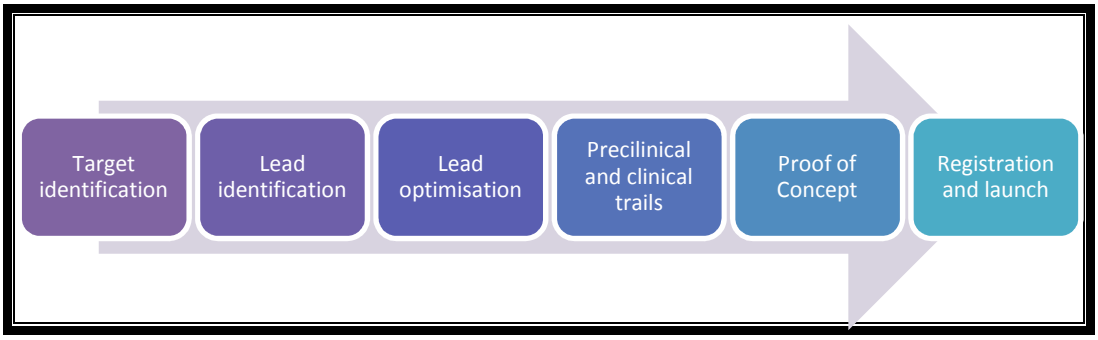


Fig. 3: Drug Discovery Pipeline

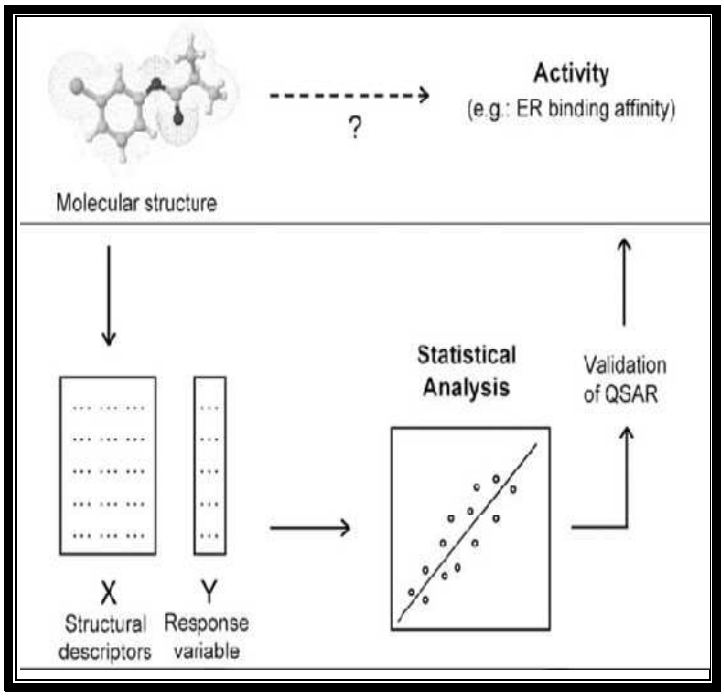


Fig. 4: Quantitative structure activity relationships (QSAR)

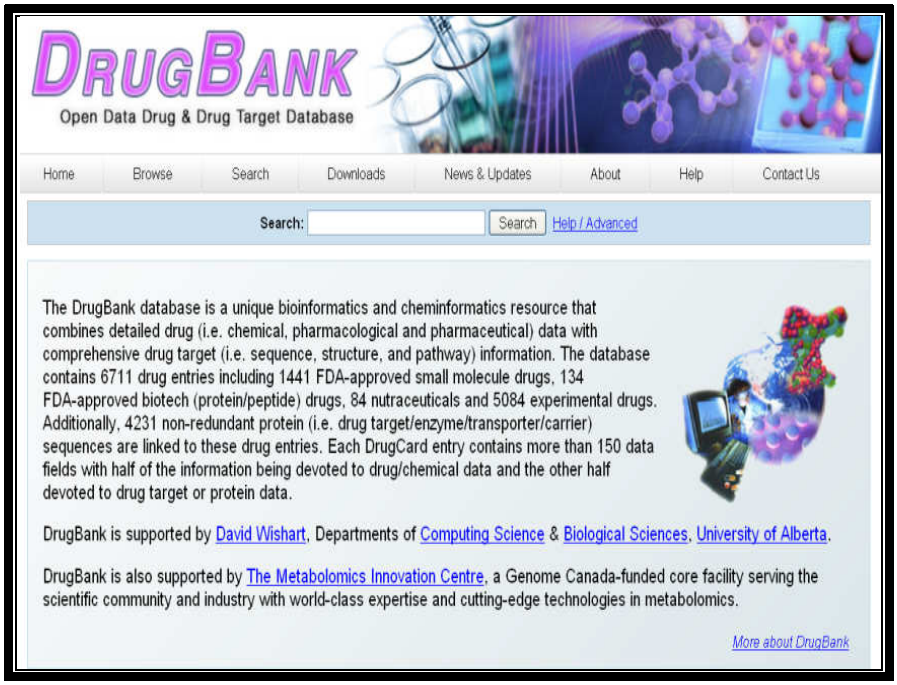


Fig. 5: Web Server of Drug Bank

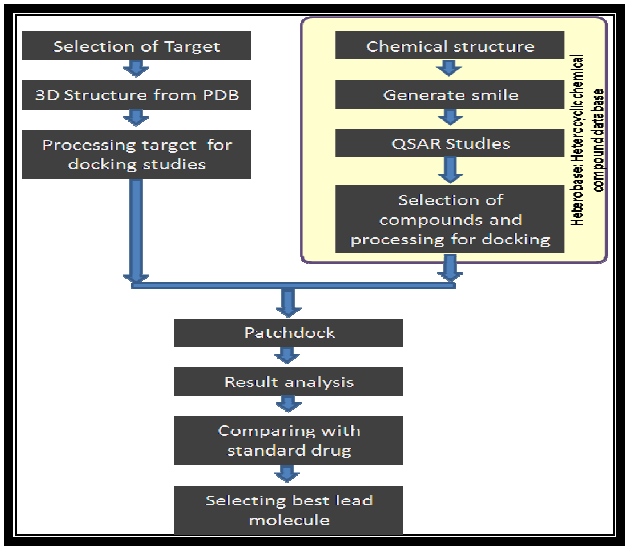


Fig. 6: Flow chart of drug discovery

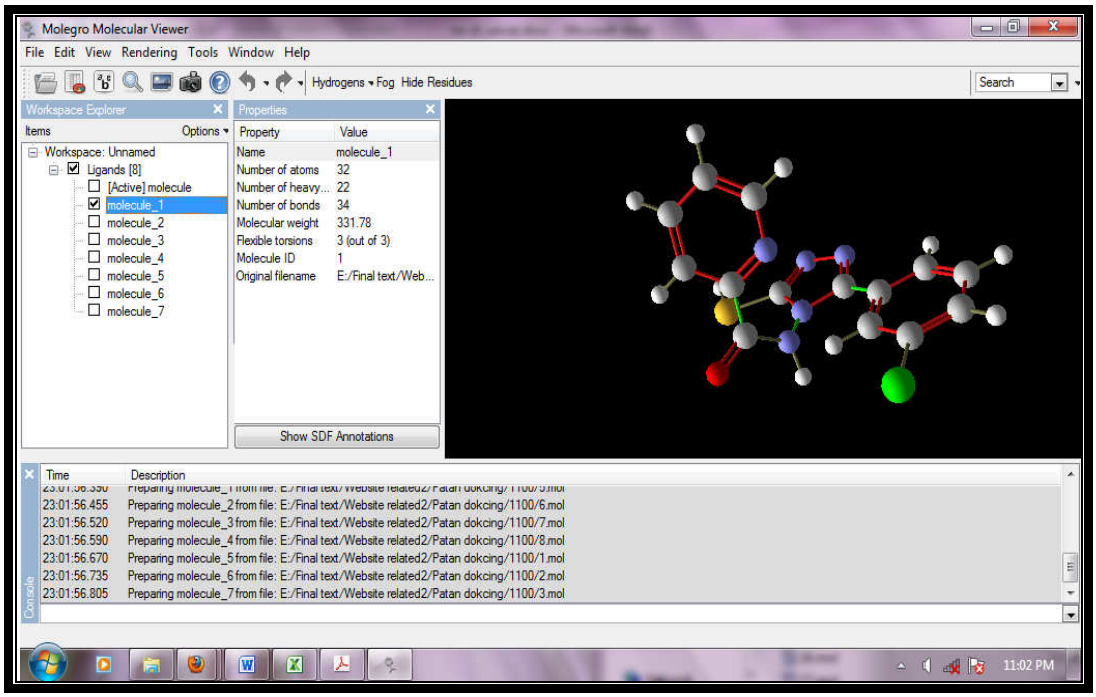


Fig. 7: Discovery Studio

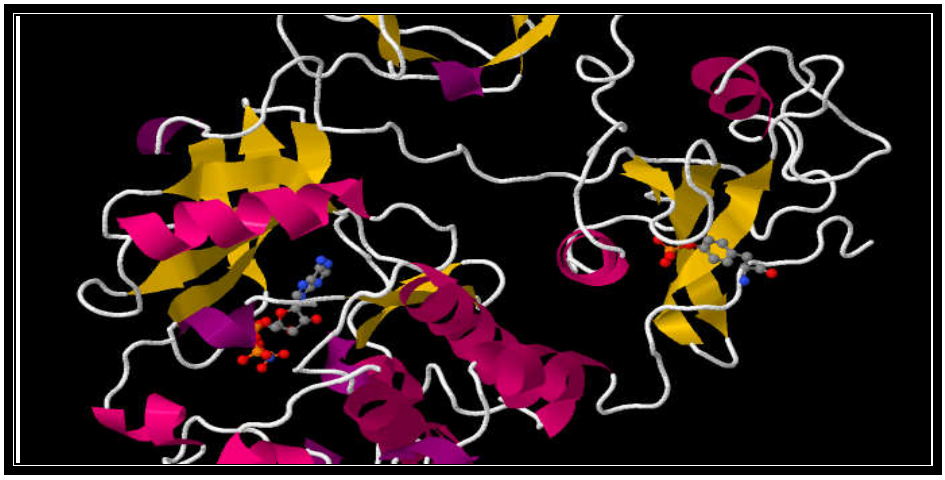


Fig. 8: Protein 3-D Structure

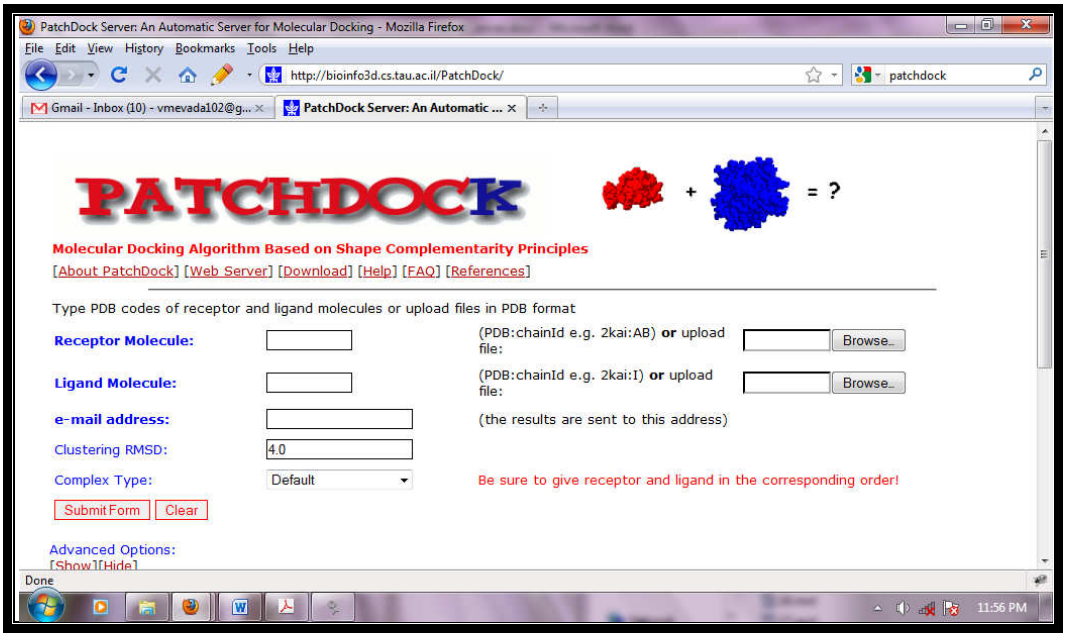


Fig. 9: Patch dock result

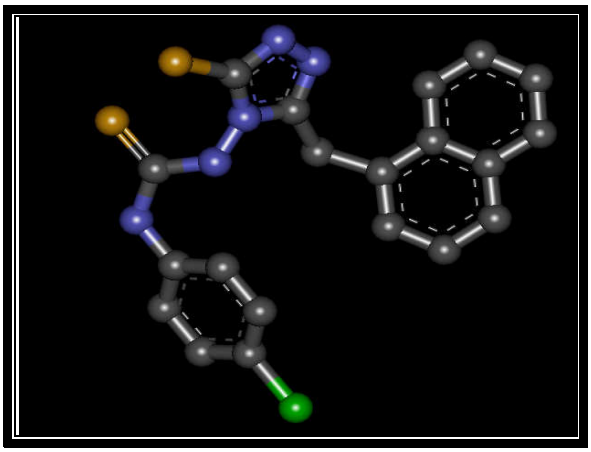


Fig. 10: Ligand 2265

CONCLUSION

Currently, several thousand molecules are yet to be screened. Virtual screening is more effective method to cope up with present world of drug discovery. As like heterobase, there were several other databases can be screened using High-throughput screening method with effective identification of the lead molecule. Patch dock is online free server for drug discovery of protein-protein and protein-ligand binding approach. Virtual screening with combinational chemistry provides several aspects for lead optimization and also helpful to clinical research trails at the last stage of drug discovery. Present work is prototype for Drug discovery using fixed receptor and multiple ligand based drug discovery called structure based drug discovery and several other aspects like flexible docking, Ligand based drug discovery need to be checked yet in future time.

REFERENCES

1. Ltadani, H.; Mizuarai, S.; Kotani, H.Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation. *Current Genomics* 2008; *9*: 349-360
2. Anand P, Kunnumakara AB, "Cancer is a Preventable Disease that Requires Major Life style Changes".*Pharmaseutical. Residence 2008;25(9): 2097–116*
3. Hanahan, D.; Weinberg, R. A. The hallmarks of cancer. Cell 2000;100: 57-70.
4. World Health Organization.(1969). WHO Expert Committee on Drug Dependence. Sixteenth report. (Technical report series. No. 407).Geneva:World Health Organization.
5. "Drug."*Dictionary.com Unabridged (v 1.1)*, Random House, Inc., via dictionary.com. Retrieved on 20 2007.
6. Kubinyi, H. Mann hold, R., Kroogsgard-Larsen, P., Timmerman, H., QSAR: Hansch Analysis and Related Approaches. In Methods and Principles in Medicinal Chemistry; Eds.; Wiley-VCH: Weinheim, Germany 1993; 1: 240.
7. Wishart D.S, Knox C, Guo A.C, "Drug Bank: a comprehensive resource for *in silico* drug discovery and exploration". Nucleic Acids Research 2006; 34 (Database issue): D668–D672
8. Kitchen, D. B. Decornez, H.Furr, J. R. Bajorath, J. Docking and scoring invirtual screening for drug discovery: methods and applications. National Review Drug Discovery 2004; 3: 935-949.
9. Weinberg, Robert A. *The Biology Of Cancer*. New York: Garland Science, Taylor and Francis Group, 2007; LLC.757–759.
10. Sheikh S.A, Sandhu G, Latha N, Jayaram B: From target to leads sketching A physiochemical pathway for lead molecule *in silico*.current pharmaceutical design 2007; 13:3454-3470.
11. Amin Rostami-Hodjegan, Geoff Tucker. '*In silico'* simulations to assess the'in vivo' consequences of 'in vitro' metabolic drug–drug interactions. Drug Discovery Today: Technologies 2004; 1 suppl(4):441-448.