Research Article **Pharmacoinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

CHEMINFORMATICS TOOLS IN DRUG DISCOVERY AND DOCKING STUDY USING PARALLEL COMPUTING

SREENIVASA REDDY P E[*1], T. SREENIVASULU REDDY[2] AND RAHUL SAGAJKAR[3]

[1] Department of chemistry, Sri Krishnadevaraya University, Anantapur, AP, India - 515003
[2] Department of chemistry, Sri Krishnadevaraya University, Anantapur, AP, India - 515003
[3] Department of cheminformatics, Accelrys, Bangalore, India – 560066

ABSTRCT

This is an attempt to highlight the tools being used by cheminformaticians, drug discovery methodology and challenges in structure based drug design using Insilico techniques with high performance computing. In the current pharmaceutical & biotechnology industry research scientist are extensively using software to reduce the time period of their research activities. Insilico techniques are useful for the screening and selection of lead compound for a target receptor. Parallel computing methods are available which support to dock[1] large ligand data set in protein active site. Docking study is performed against a kinase protein PDB ID 2C69.pdb[7] with randomely selected data set of 249,072 ligands[13]. It was observed that some of the ligands show better docking scores than the reference ligand which is already available in the active site of protein.

KEYWORDS:Bioinformatics, cheminformatics, drug discovery, in-silico methods, rational drug design, docking and parallel computing.

SREENIVASA REDDY P E
Department of chemistry, Sri Krishnadevaraya University, Anantapur, AP, India - 515003

***Corresponding author**

INTRODUCTION

Traditional Drug Discovery

Regular Pharmaceutical industry would require 12-14 years and costing up to $1.2 - $1.4 billion to bring a drug from discovery to market[11], in this approach drugs were discovered by synthesizing compounds in a time-consuming multi-step processes 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.

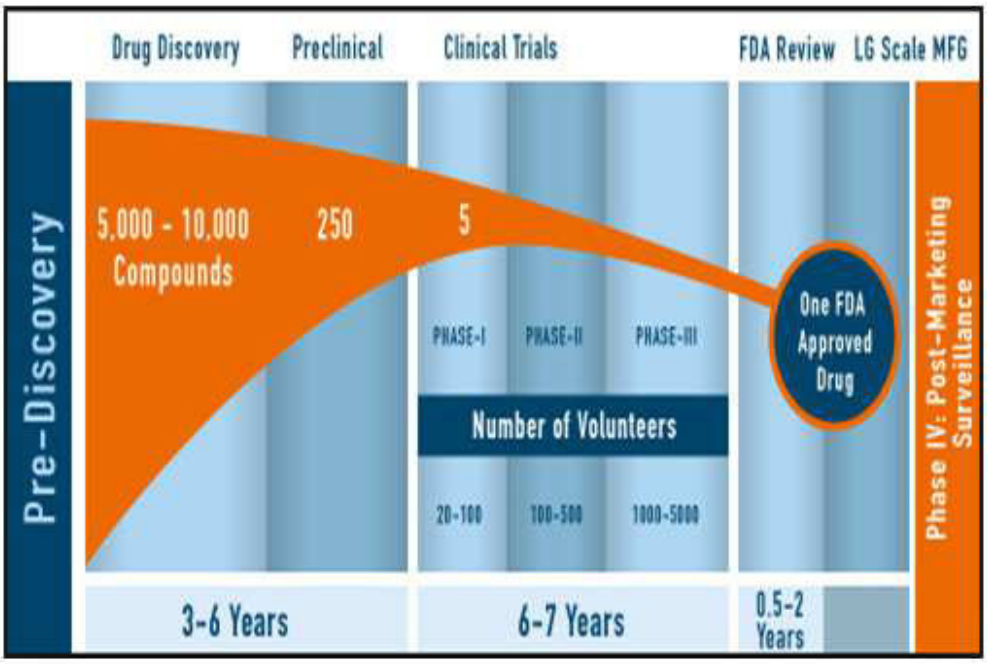


Figure1
Milestones of Traditional Drug Discovery.

Modern Drug Discovery

Current drug discovery has been revolutionized with the advent of genomics, proteomics, bioinformatics and efficient technologies like, combinatorial chemistry, cheminformatics, high throughput screening[9] (HTS), virtual screening, de novo design, in vitro, in silico ADMET screening, Quantitative structure-activity relationship (QSAR) and structure-based drug design[8].The majority of pharmaceutical drug discovery programs currently, begins with a known macromolecular targets[2], and seek to identify a suitable small molecule modulator.The process of optimizing the lead molecule into a candidate drug is usually the longest and most expensive stage in the drug discovery process. Following the selection of the candidate molecule from limited chemical space of analogs of original lead compounds, to the drug development, scientists develop large scale production methods, and conduct the preclinical animal safety studies.

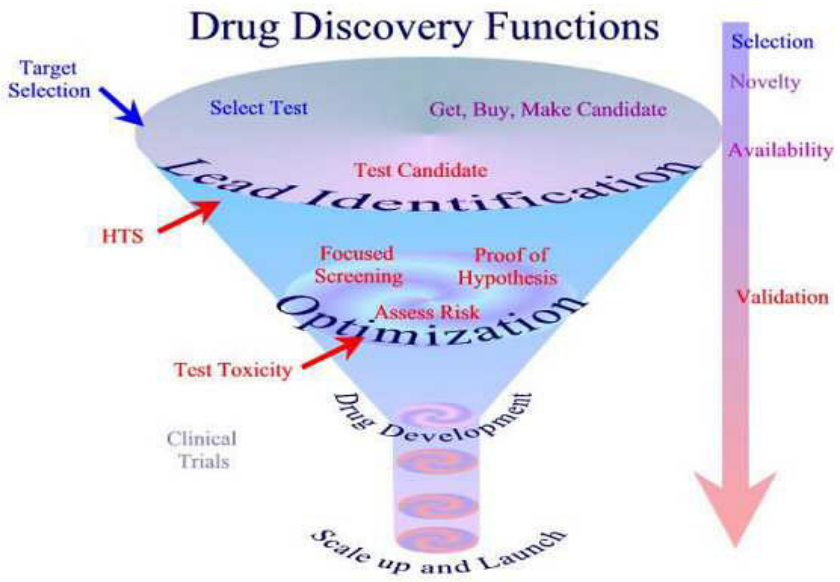


Figure 2
Milestones of Modern Drug Discovery.

Cheminformatics in Drug Discovery

Cheminformatics consists of several in-silico techniques which are widely used in pharmaceutical companies in the process of drug discovery. Investigate and test hypotheses in silico prior to costly experimental implementation, thus reducing the time and expense involved in bringing products to market The primary application of cheminformatics is in the storage of information related to the drug molecules and the efficient presentation of such stored information during the process of lead optimization. Cheminformatics tools also enable companies to accelerate the identification of genetic information for gene-based drug targets, validate the same through drug-target interaction studies, and interpret existing genetic information to enable decision making in the product development cycle.

Figure 3
Life Cycle of Drug Discovery

Cheminformatics Tools

Depictions of molecular compounds and storing a chemical structure or reaction is the primary objective of cheminformatics. Representation of Stereochemistry is key while presenting structures, 3-D structures, lone pairs of electrons and inorganic compounds are also part of structure drawing tools. Several File formats are being used to represent the molecule or reaction eg: .cml, .sdf, .mol, smiles etc.

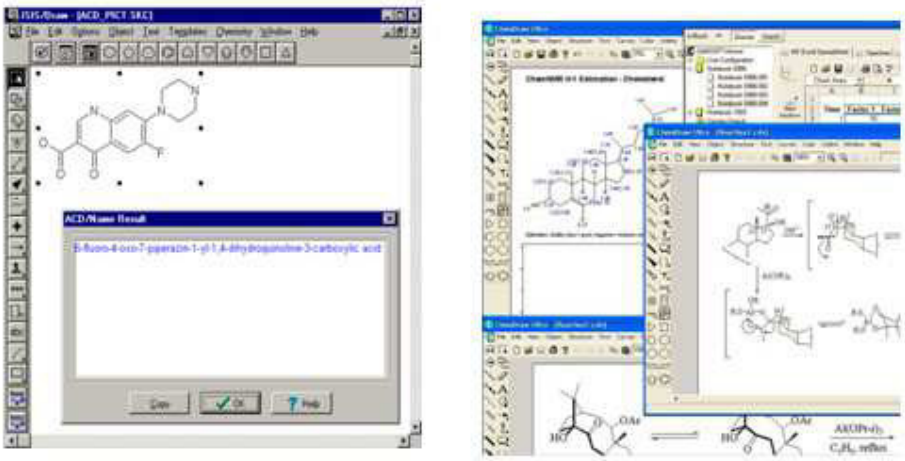


Figure4
Cheminformatics tools (ISIS Draw and Chem Draw)

Tools are being used in many ways throughout the life cycle of drug discovery, some of them are depicted below

Structure to Name Convert chemical structures into standard IUPAC names.
Import and export structures or reactions.

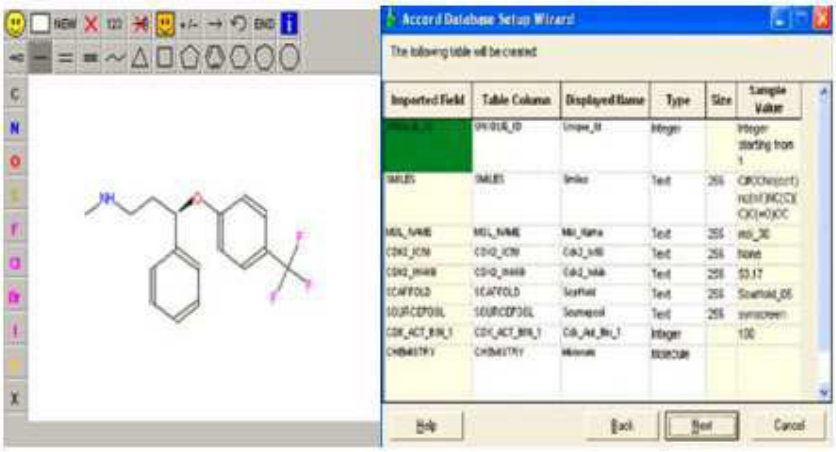


Figure5
Tools (JME & Accord) to calculate chemical structure properties

Chemical Property Caliculations and Analysis Tools

Cheminformatic tools ensure that the drug discovery process is fully automated, including design, synthesis, QC, and testing of compounds. Data bases: Substructure and similarity searches, fragment-based virtual screening, bioactivity prediction and data visualization Aid pharmaceutical and biotechnology industries by generating huge combinatorial libraries, access and search, and manipulate data to identify analogues.

These systems and tools compute,sample, visualize and validate compounds in relation to their chemical Properties.

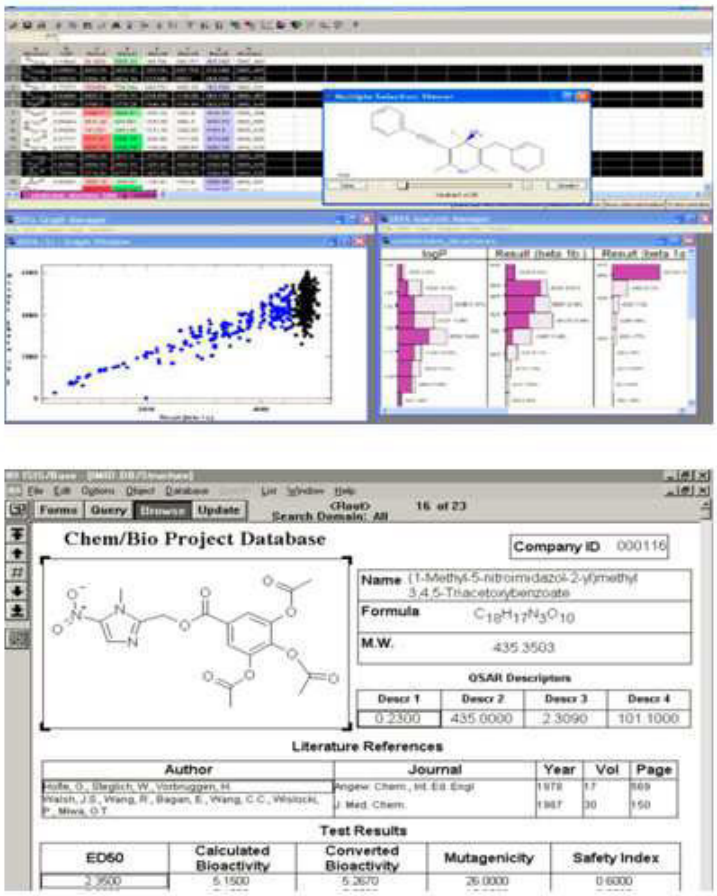


Figure6

Tools (DIVA 3.0 and ISIS Base) to calculate chemical structure properties

Pharmacophore Model Development

A range of parameters are used to assess the quality of a compound, or a series of compounds, as proposed in the Lipinski's Rule of Five[6]. LogP to estimate lipophilicity Rotatable bonds to estimate molecular flexibility Molecular weight of the ligand Hydrogen Bond Acceptors and Donors to estimate Pharmacophoric Properties olar surface area and measured properties such as potency, in-vitro measurement of enzymatic clearance etc. Descriptors such as ligand efficiency (LE) and lipophilic efficiency (LiPE) combine such parameters to assess the druglikeness.

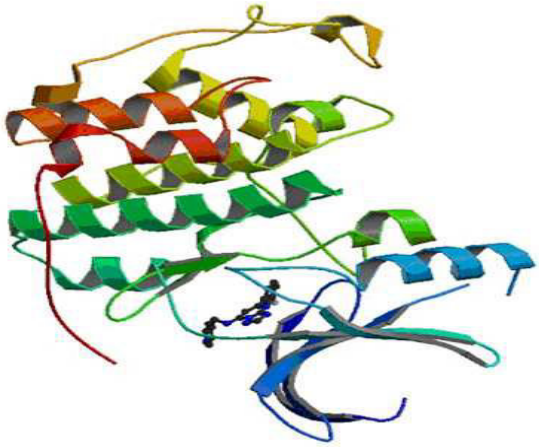
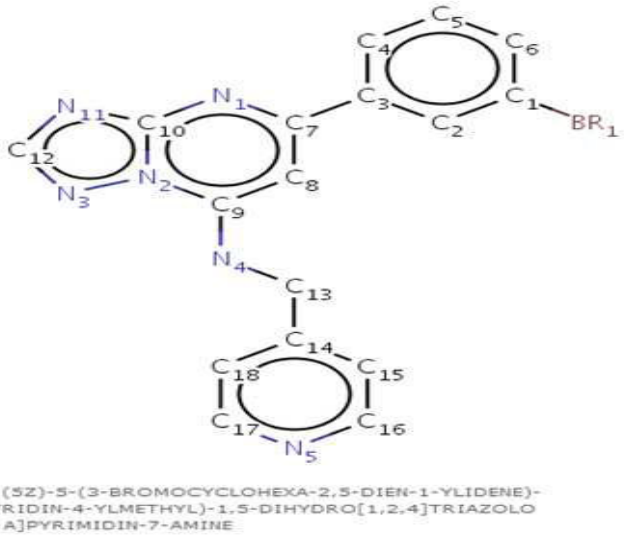


Figure7
Pharmacophore Model

Current research mainly focused on to identify a lead molecule by using cheminformatics tools and docking study using parallel computing. It will help chemists at the stage where lead optimization takes place. The ultimate goal of docking is the prediction of the three dimensional structure of the macromolecular complex of interest as it would occur in a living organism. Docking itself only produces plausible candidate structures. These candidates must be ranked using methods such as scoring functions to identify structures that are most likely to occur in nature.



CT8: (5Z)-5-(3-BROMOCYCLOHEXA-2,5-DIEN-1-YLIDENE)-N-(PYRIDIN-4-YLMETHYL)-1,5-DIHYDRO[1,2,4]TRIAZOLO[1,5- A]PYRIMIDIN-7-AMINE

Figure8
Chemical structure of Pharmacophore Model

Terminology

Receptor *or* ***host*** *or* ***lock*** – The "receiving" molecule, most commonly a protein or other biopolymer.

Ligand *or* ***guest*** *or* ***key*** – The complementary partner molecule which binds to the receptor. Ligands are most often small molecules but could also be another biopolymer.

Docking – Computational simulation of a candidate ligand binding to a receptor.

Binding mode – The orientation of the ligand relative to the receptor as well as

the conformation of the ligand and receptor when bound to each other.

Pose – A candidate binding mode.

Scoring – The process of evaluating a particular pose by counting the number of favorable intermolecular interactions such as hydrogen
bonds and hydrophobic contacts.

Ranking – The process of classifying which ligands are most likely to interact favorably to a particular receptor based on the predicted free-energy of binding[17].

Pharmacophore - The group of atoms in the molecule of a drug responsible for the drug's action.

MATERILS AND METHODS

Docking study " the process of finding the best relationship between two molecules" is performed against a kinase protein (receptor) PDB ID 2C69 with randomly selected data set of 249,072 ligands (potential candidate drugs) .Docking study is performed by using Accelrys Discovery Studio 3.5[1] to find the "best" lowest energy state of the receptor - ligand complex.Pharmacophore was developed based on Lipinski "Rule of Five" [12]Pharmacophore Drug likeness score and bio activity were calculated by using molinspiration tools[18].

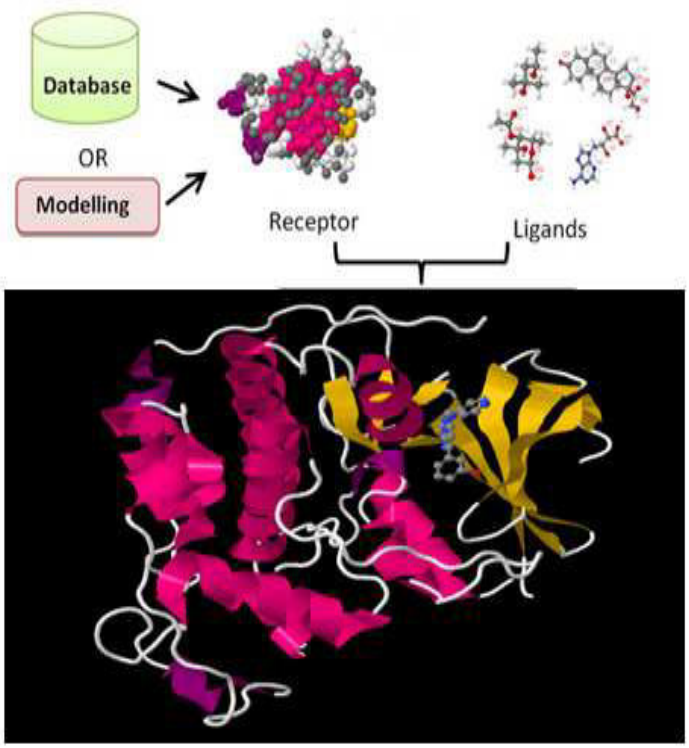


Figure9
Process of identifying Receptor and docking with Ligands.

For Docking the forcefield used is dreiding & scoring scores are LigScore1, LigScore2, PLP1, PLP2, Jain, PMF, Ludi Energy Estimate 1, Ludi Energy Estimate 2 & Ludi Energy Estimate 3. The parallel computing is performed on two different computational servers. The chunk size used is 20000. Conformation Search maximum internal energy is 10,000.Conformer results & analysis show that number of output poses in docked.sd are 678,235. Some of the output poses show better dock score than reference ligand.

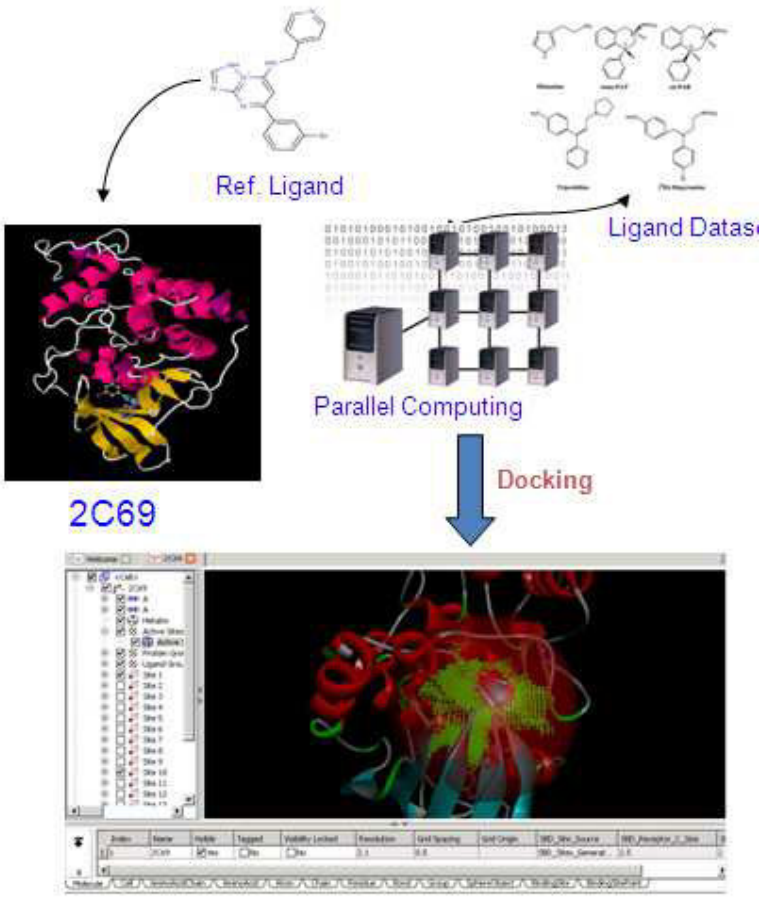


Figure 10
Process of docking through parallel computing.

Scoring Functions

Name	Function	Contents
LigScore1	Empirical scoring function	VDW + Protein-Ligand Attraction / Repulsion
LigScore2		LigScore1 + Parameter Improvement
PLP1		H-Donor, H-Acceptor, Both, Non-polar
PLP2		PLP1 + Atomic Radii, Repulsion
Jain		Very short range - Lipophilic, Polar Interaction, Solvation
Ludi		Ionic, Hydrogen, Lipophilic Interaction
PMF	Knowledge based scoring function	Interatomic pair interaction in data-set
PMF04		PMF + Larger Data set, Several Metal and Halogen

Tabale1
Scoring Functions used in Accelrys Discovery Studio

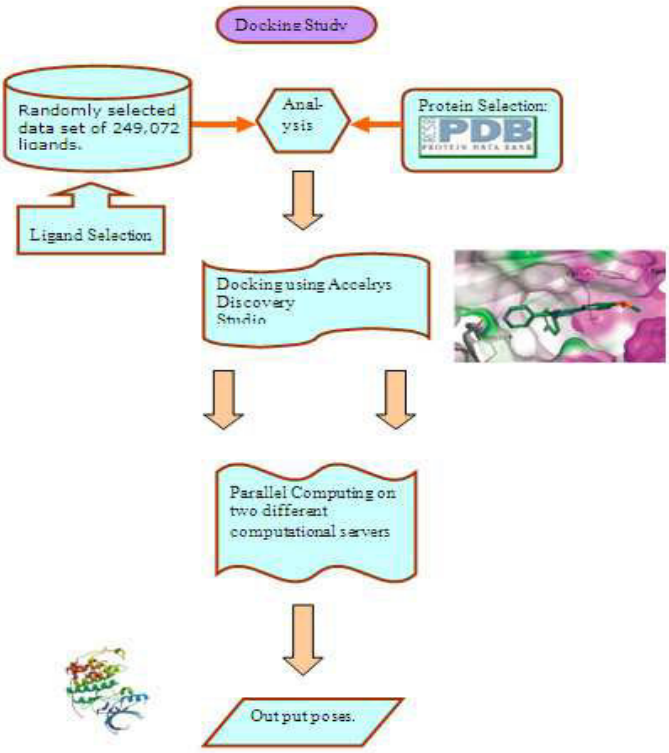


Figure 11
Docking study using parallel computing tools – Flow representation.

7. RESULTS AND DISCUSSION

By using Accelrys Discovery Studio 3.5 Docking study is performed against a kinase protein PDB ID 2C69 (a human cyclin dependent kinase 2 which plays a major role in cell cycle regulation) with randomly selected data set of 249,072 ligands .For Docking the forcefield used is dreiding & scoring functions are LigScore1, LigScore2, PLP1, PLP2, Jain, PMF, Ludi Energy Estimate 1, Ludi Energy Estimate 2 & Ludi Energy Estimate 3.The parallel computing is performed on two different computational servers. The chunk dataset. size used is 20000. Conformation Search maximum internal energy is 10,000.Conformer results & analysis show that the number of output poses in results.sd are 1098. Some of the output poses show better dock score than reference ligand.Further study of the poses having a good score show better fit in active site of the receptor & can be considered as potential drug candidates.Parallel computing is a better option for efficient screening of large & varied

Docked Ligand Properties[18]
Physical Properties

LogP	3.149
TPSA	93.969
natoms	32.0
MW	466.592
nON	8
nOHNH	2
nviolations	0
nrotb	9
Volume	399.506

Table 2 ***Physical Properties***

Bio Activity Score

GPCR Ligand	**0.63**
Ion channel modulator	**0.88**
Kinase inhibitor	**0.49**
Nuclear receptor ligand	**1.06**
Protease inhibitor	**0.63**
Enzyme inhibitor	**0.62**

Table3
Bio Activity Properties

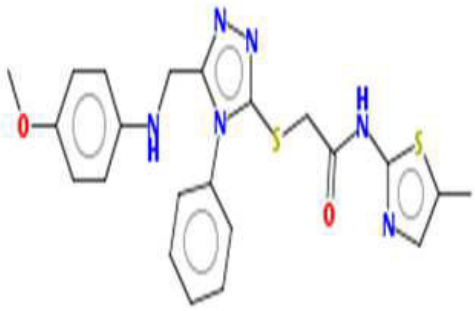


Figure 12
Chemical Structure of Docked Ligand.

Reference Ligand Properties
Physical Properties

miLogP	-1.129
TPSA	70.597
natoms	24.0
Molecular Weight	382.245
nON	6
nOHNH	2
nviolations	0
nrotb	4
volume	289.534

Table 4
Physical Properties
Bio Activity Score

GPCR Ligand	-0.18
Ion channel modulator	-0.05
Kinase inhibitor	0.08
Nuclear receptor ligand	-0.53
Protease inhibitor	-0.35
Enzyme inhibitor	-0.10

Table 5
Bio Activity Properties

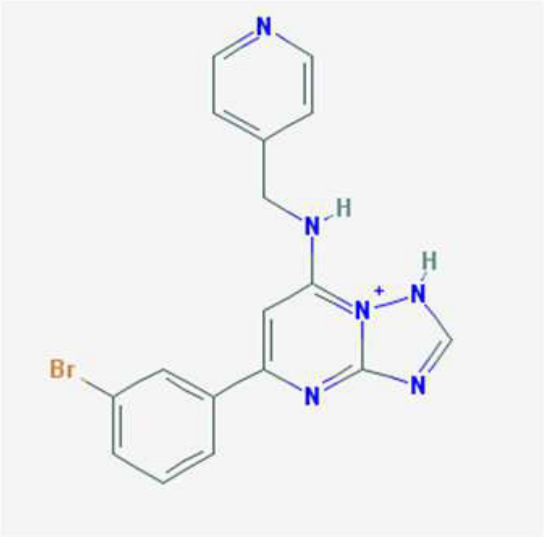


Figure13
Chemical Structure of reference ligand.

Potential drugs from the study can be taken in to the next phases of drug development and clinical trails may result desirable drug[19]. Hence Parallel computing is a better option for efficient screening of large and varied dataset, also cost and time effective during the critical phase of lead optimization[20].Cheminformatics is evolving into an important facilitator that accelerates as well as improvises the efficiency of the drug discovery processes. With improvements in computing power as well as technology, cheminformatics is all set to play a vital role in the drug discovery processes.

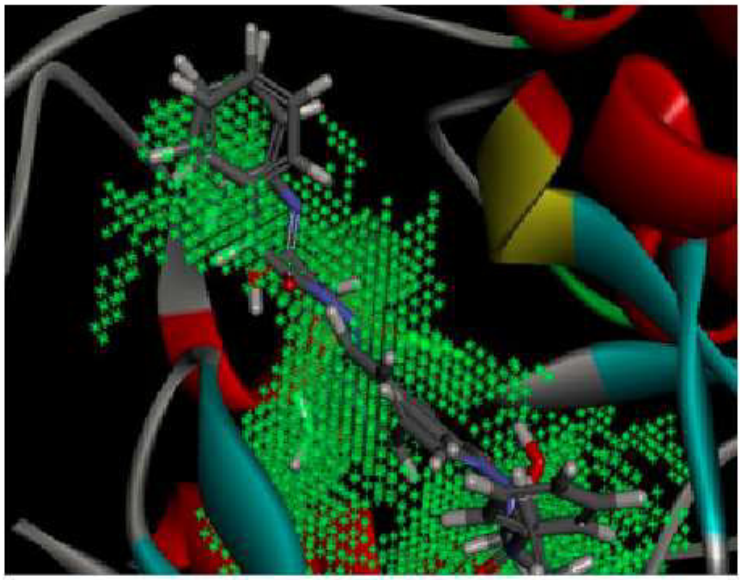


Figure14
Docked Ligand

The docking results were ranked according to the ascent of the docking energies of the conformers for each of the ligands, ranking the energy results according to the Binding Energy which included the Intermolecular Energy and the Torsion terms. It was found that most of the ligands interacted quite well with the receptor in the pocket.A comparison

of the results between the top 12 ranked conformers suggests that some large sized ligands suffer from more loss of Torsional freedom upon binding.The top three compounds were selected based on the scoring function, which predicts the ranking of different ligands in approximate order of ligand size, order of affinity and allows selectivity. It was found that compound no 1, 2, 6 and 7 had lowest Ligand Internal energy.Docking perform rapid rigid body docking with the well-published ZDOCK algorithm[14].Which employed a pair-wise shape complementarity function for identifying docked conformations and scores hits based on atomic contact energies. Increase the accuracy of docked poses using the ZRANK scoring function11. Used the RDOCK[15] algorithm12 to refine ZDOCK hits based on a CHARMm energy minimization and score poses by CHARMm energy and desolvation energy[16]. Narrow the search and identify poses of interest with advanced clustering methods.

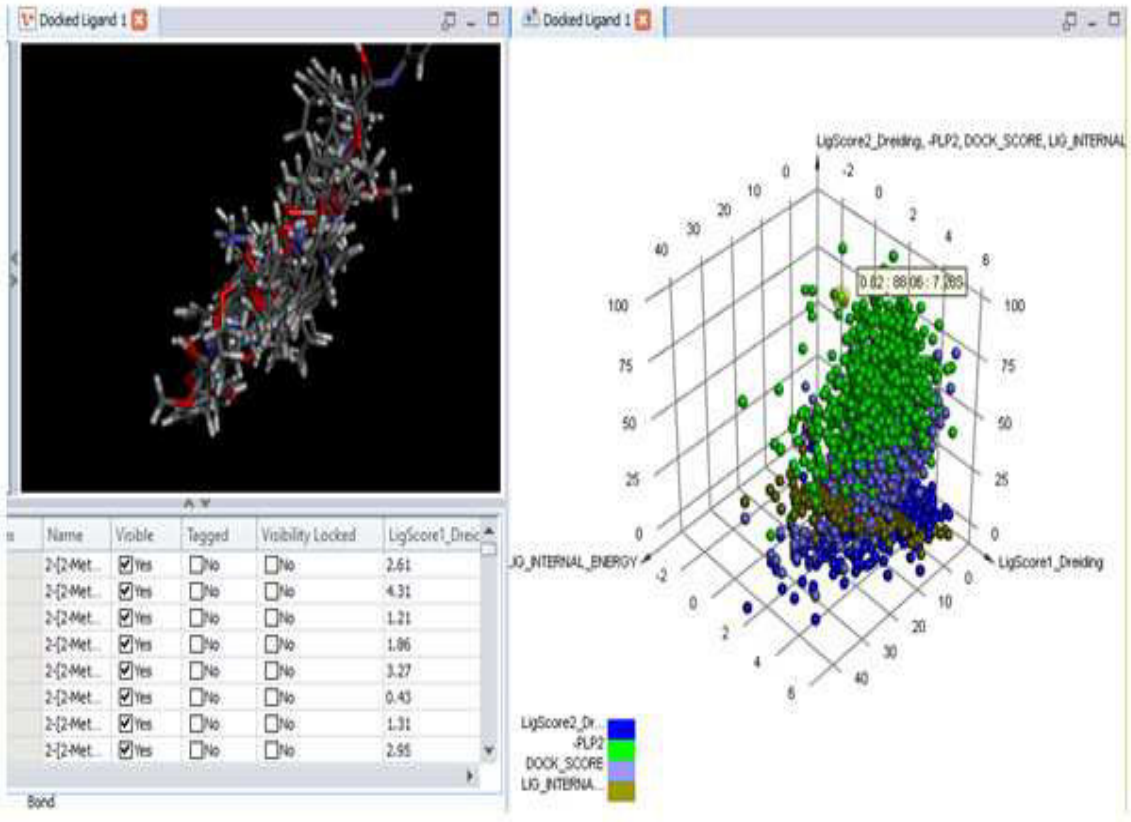


Figure 15
Cluster representation.

Docked Ligand Properties

	POSE	Structure	Name	DOCK_SC	LigScore1	LigScore2	-PLP1	-PLP2	ALogP	Jain	ADME	LIG_INTE	Molecu	Num	ADMET_
1	1		2-(5-[(4-...	75.603	5.31	3.26	30.85	41.18	3.408	0.8	0	-6.063	466.579	2	-5.004
2	1		5-(4-Buto...	69.736	4.33	5.27	94.66	82.49	4.48	2.24	0	-5.998	400.473	1	-5.553
3	1		2-(5-[(2-...	62.788	2.05	3.73	34.16	24.12	4.929	-1.21	0	-3.729	479.982	2	-5.872

Table 6 ***Docked Ligand properties (Pose 1-3)***

	POSE	Structure	Name	DOCK_SC	LigScore1	LigScore2	-PLP1	-PLP2	ALogP	Jain	ADME	LIG_INTE	Molecu	Num	ADMET_Solubility
4	1		5-(4-Buto...	59.953	3.36	4.95	93.56	81.7	4.48	1.3	0	-1.375	400.473	1	-5.553
5	2		5-(4-Buto...	58.736	3.27	4.67	81.21	69.37	4.48	0.95	0	2.773	400.473	1	-5.553
6	1		2-(2-[5-(...	57.868	4.98	3.19	69.51	66.96	3.98	1.02	1	-6.926	459.539	2	-5.363

Table 7 ***Docked Ligand properties (Pose 4-6)***

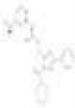
	POSE	Structure	Name	DOCK_SCO	LigScore1	LigScore2	-PLP1	-PLP2	ALogP	Jain	ADME	LIG_INTE	Molecu	Num	ADMET_Solubility
7	1		2-(2-Cyd...	57.144	3.2	4.86	71.2	59.58	4.94	0.58	2	-7.096	465.525	1	-5.938
8	1		5-Acetyl-...	57.082	5.04	4.38	66.61	53.07	2.206	-0.29	0	-1.997	446.5	1	-3.706
9	1		5-Acetyl-...	57.073	4.4	4.05	76.7	65.4	2.196	-0.63	0	1.373	445.515	2	-4.151

Table 8
Docked Ligand properties (Pose 7-9)

	POSE	Structure	Name	DOCK_SCO	LigScore1	LigScore2	-PLP1	-PLP2	ALogP	Jain	ADME	LIG_INTE	Molecu	Num	ADMET_Solubility
10	1		3-Amino-...	56.676	0.39	-1.4	51.07	70.45	4.92	3.67	2	-1.443	487.597	2	-7.797
11	3		5-(4-Buto...	55.296	3.82	4.95	90.57	77.37	4.48	2.07	0	5.846	400.473	1	-5.553
12	1		2-(6-Fluo...	54.611	5.17	4.21	42.67	33.91	4.809	-0.28	1	0.787	464.469	1	-6.205

Table 9
Docked Ligand properties (Pose 10-12)

8.SUMMARY AND CONCLUSION

Conformer results & analysis shows that the number of output poses in results.sd are 1098. Some of the output poses show better dock score than reference ligand. Further study of the poses having a good score show better fit in active site of the receptor & can be considered as potential ligands. Parallel computing is a better option for efficient screening of a large & varied dataset. Cheminformatics is evolving into an important facilitator that accelerates as well as improvises the efficiency of the drug discovery processes. With improvements in computing power as well as technology, cheminformatics is all set to play a vital role in the drug discovery processes.

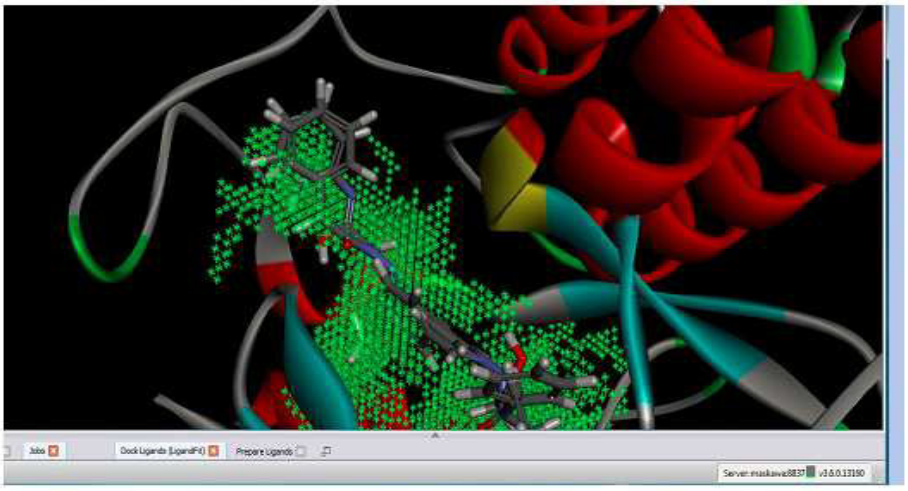


Figure16
Final Docked Ligand

ACKNOWLEDGEMENT

Sreenivasa Reddy P E, thank the professors, Dr. K. Hussain Reddy, Dr. M.C.S. Subha, Dr. R. Srinivasulu, Dr. G. Narayana Swamy, Dr. L.K. Ravindranath, Dr. V. Krishna Reddy, Dr. J. Sreeramulu, Dr. P. Ravindra Reddy and Dr. K. Sudhakar Babu, Department of chemistry, Sri Krishna Devaraya University, for their extended support. my sincere thanks due to Dr. Prasanth Naik, Head Informatics Division, Jubilant Biosys, Bangalore, Sri. Nafees Ahmed, Sr.Manager, Cheminformatics Division, Accelrys, Bangalore, Dr. K Nagabhusan Reddy and Dr. G Saayi Krushna. Last but not least my thanks to my father Pappireddy Reddenna, my mother P. Rani Prameela Devi, my wife P. Divya Reddy, daughter Reddy Praharshitha Pappireddy and my brothers P. Vishnuvardhana Reddy, P.Pavankumar Reddy, In-Laws Sri. P. Nageswara Reddy and Smt. Tejavathi my uncle Sri. V.S Prabhakara Reddy, V. Venugopal Reddy, V. Nagi Reddy, V. Madhusudhan Reddy and to my friends for their continuous support at all the time.

REFERENCES

1. Accelrys Software Inc., *Discovery Studio Modeling Environment, Release 1.5*, San Diego: Accelrys Software Inc., 2007.
2. Stierand, K., Maaß, P., Rarey, M. (2006) Molecular Complexes at a Glance: Automated Generation of two-dimensional Complex Diagrams. Bioinformatics, 22, 1710-1716.
3. Fricker, P., Gastreich, M., and Rarey, M. (2004) Automated Generation of Structural Molecular Formulas under Constraints. Journal of Chemical Information and Computer Sciences, 44, 1065-1078.
4. Stierand, K., Rarey, M. (2007) From Modeling to Medicinal Chemistry:

Automatic Generation of Two-Dimensional Complex Diagrams. ChemMedChem 2 (6), 853-860.
5. Stierand, K., Rarey, M. (2010) Drawing the PDB: Protein−Ligand Complexes in Two Dimensions. Medicinal Chemistry Letters, 1 (9), 540-545.
6. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (March 2001). "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings". *Adv. Drug Deliv. Rev.* 46 (1-3): 3–26
7. Richardson, C.M., Williamson, D.S., Parratt, M.J., Borgognoni, J., Cansfield, A.D., Dokurno, P., Francis, G.L., Howes, R., Moore, J.D., Murray, J.B., Robertson, A., Surgenor, A.E., Torrance, C.J. "CRYSTAL STRUCTURE OF THE HUMAN CDK2 COMPLEXED WITH THE TRIAZOLOPYRIMIDINE INHIBITOR". (2006) Bioorg.Med.Chem.Lett. 16: 1353. Accessed on "20 March 2006. http://www.rcsb.org/pdb/explore.do?structureId=2c69
8. Bajorath, J. Integration of virtual and high-throughput screening. Nature Rev. Drug Discov. 1, 882–894 (2002).
9. Walters, W. P., Stahl, M. T. & Murcko, M. A. Virtual screening — an overview. Drug Discov. Today 3, 160–178 (1998).
10. Schneider G, Fechner U (August 2005). "Computer-based de novo design of drug-like molecules". Nat Rev Drug Discov 4 (8): 649–63. doi:10.1038/nrd1799. PMID 16056391.
11. Sreenivasa Reddy P E, T. Sreenivasulu Reddy, G Saayi Krushna. Current Research in Drug Targeting 2011, Universal Research Publications; 1(2): 6-17
12. Lipinski CA (December 2004). "Lead- and drug-like compounds: the rule-of-five revolution". Drug Discovery Today: Technologies 1 (4): 337–341.
13. Sreenivasa Reddy P E, Rahul Sagajkar, Darshan Mirasdar and Sreenivasulu Reddy. International Journal of Pharmaceutical and Clinical Science 2011; Universal Research Publications; 1 (1): 9
14. Pierce B., Weng Z. A Combination of Rescoring and Refinement Significantly Improves Protein Docking Performance. Proteins, 2008, 72(1), 270-279.
15. Chen R., Weng Z. ZDOCK: An Initial-stage Protein-Docking Algorithm. Proteins 2003, 52, 80-87.
16. Pierce B, Weng Z. ZRANK: Reranking Protein Docking Predictions with an Optimized Energy Function. Proteins, 2007, 67(4), 1078-1086.
17. Wikipedia contributors. "Docking (molecular)". Wikipedia, The Free Encyclopedia Accessed on "20 March 2014. http://en.wikipedia.org/w/index.php?title=Docking_(molecular)&oldid=617372802
18. Molinspiration Cheminformatics Software. Calculation of Molecular Properties and Bioactivity Score "Cheminformatics on the Web" Accessed on 10 Jun 2010. http://www.molinspiration.com/cgi-bin/properties
19. Jurgen Suhnel Biocomputing Group. "3D Structures of Biological Macromolecules" Leibniz Institute for Age Research - Fritz Lipmann Institute. Accessed on 10 Jun 2010. http://www.imb-jena.de/www_bioc/3D/
20. Douglas B. Kitchen, Helene Decornez, John R. Furr and Jurgen Bajorath. "DOCKING AND SCORING IN VIRTUAL SCREENING FOR DRUG DISCOVERY: METHODS AND APPLICATIONS". Nature Reviews Drug Discovery 3, 935-949 (November 2004). Accessed on 10 Jun 2010. https://sakai.rutgers.edu/access/content/user/kparis/biomaps_513_references/10_B_01_NatRevDrugDisc3_935.pdf.