

Windows Bioinformatics Platform: Approach for Molecular Docking in Drug Discovery

Ayad Ghany Ismaeel, Suhad Jihad Al-Dulaimy, Forat Falah Hasan

Abstract— There are many reasons may cause effort useless, impossible or incapacitating in traditional drug discovery, e.g. the traditional way of development a new drugs takes 10-15 years and costs 400-800 millions US dollars, and complex diseases like cancer are regulated by large, interconnected networks with many pathways affecting cell proliferation, invasion, and drug resistance reaching to deadlock [18]. The fact that most of bioinformatics tools based on linux environment when try to dock using open source the Autodock Vina (a next generation of Autodock 4) in drug discovery and avoided working under windows platform (which is common in use) because from view of practice it has some gaps/drawbacks. This research offers the development of high-throughput method in both biology and chemistry that can be used the bioinformatics tools in drug discovery, as insilico method in drug discovery based on molecular docking using Windows platform, and solving all technical problems which contained it. The experimental result reveals the suggested approach gives windows platform results matching to linux platform, i.e. experimental result shows solving all the conjugation bugs/gaps to the windows platform when implementing the docking procedure.

Index Terms— AutoDock Vina, Drug Discovery, Druggable gene, Insilico Method, Molecular Docking, Receptor, Target.

1 INTRODUCTION

THE human genome project was initiated on Oct 1, 1990, and the complete DNA sequence of the human genome was completed in 2001 (www.nhgri.gov). A large number of researches focus on prediction early the cancers via mutations in genes caused them [1], i.e. not focusing on therapy techniques or drug discovery of this diseases. Central to the drug discovery process is gene identification followed by determination of the expression of genes in a given disease and understanding of the function of the products. It is of interest that the identification in the early 1980s, of the gene believed to be responsible for cystic fibrosis took researchers approx. 9 Years to discover, whereas the gene responsible for Parkinson's disease was identified within a period of several weeks. This quantum leap in the ability to associate gene with a disease can be attributed primarily to the extraordinary progress that has been made in the areas of gene sequencing and information technologies. Selection and validation of novel molecular targets have become paramount in light of the abundance of new potential therapeutic drug targets that have emerged from human gene sequencing [2, 3]. The development of high-throughput methods in both biology & chemistry is therefore necessary. In addition, it has become increasingly challenging to translate successfully basic scientific discoveries into clinical experimental medicine and novel therapeutics, consequently, a new paradigm for drug discovery has

emerged. The integration of clinical, genetic, genomic, and molecular phenotype data partnered with cheminformatics is involved in this process. Central to this process, the data that are generated are managed, collated, and interpreted with the use of informatics [4].

Fig. 1 shows the difference between traditional drug discovery and drug discovery using bioinformatics tools.

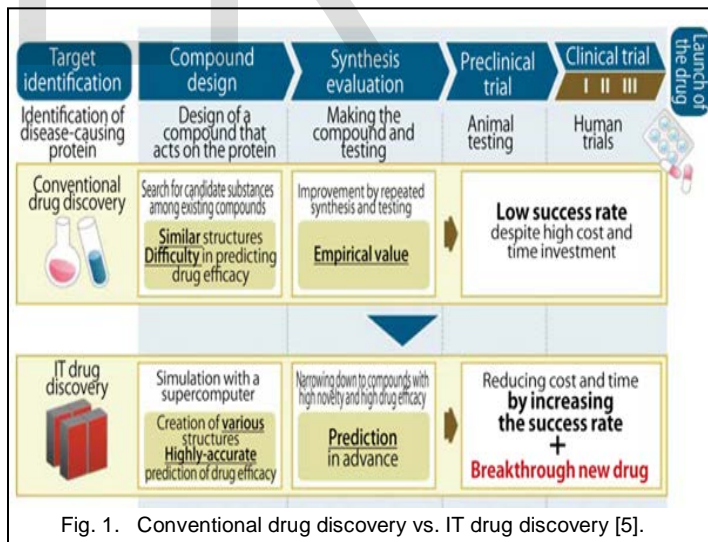


Fig. 1. Conventional drug discovery vs. IT drug discovery [5].

1.1 Drug-gene Relation

1. Explanation of Drug Work:

May be some one asks:

- How does a drug act anyway?
- How does Aspirin relieve headaches?
- Why do β -blockers lower blood pressure?
- Where does acalcium channel blocker act?
- How does cocaine work?
- How do sulfonamides prevent the proliferation of bacterial pathogens?

- Ayad Ghany Ismaeel is currently pursuing Professor, Computer Science PhD in department of Information System Engineering, Technical Eng. College- Erbil Polytechnic University (previous FTE- Erbil), Iraq. PH-009647703580299. E-mail: dr_a_gh_i@yahoo.com Alternative dr.ayad.ghany.ismaeel@gmail.com
- Suhad Jihad Al-Dulaimy is currently pursuing masters degree program in Information System Engineering in Technical Engineering College, Erbil Polytechnic University, Iraq. E-mail: suhadja2@yahoo.com
- Forat Falah Hasan is currently pursuing Head of Computer Technical Engineering Dept - University Al Kitab College, Kirkuk-Iraq. PH-009647701002648. E-mail: foratmis@yahoo.com

The answers of these questions related to purposefully design an active substance, this active substance must bind to a very special target molecule in the body to exert its pharmacological action. Usually this is a protein, but nucleic acids in the form of RNA and DNA can also be target structures for active molecules. An important prerequisite for the binding is that the active substance has the correct size and shape to fit into a cavity on the surface of the protein, a binding pocket, as well as possible. Furthermore, it is also necessary that the surface properties of ligand and protein fit together so that the specific interactions can form [6].

2. Terminologies Related:

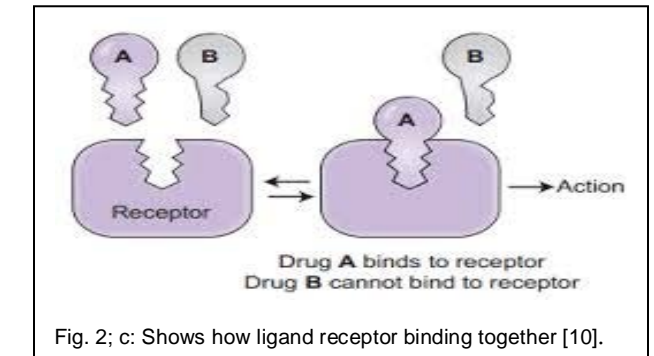
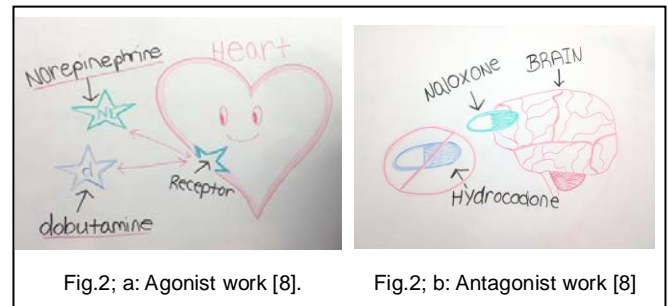
Target: Target-based drug discovery starts with an understanding of the disease mechanisms and the role of enzymes, receptors or proteins within the disease pathology. Therefore, must be first understanding pathways involved in DNA damage and repair. As an example: Most cancer cells have defects in their DNA damage response. If can identify aberrations in DNA Damage Response DDR pathway functions in tumours, and can use the deep understanding of these pathways to identify new targets, either in the aberrant pathway or in a second pathway with which it shows synthetic sensitivity to cell growth or genotoxic drugs. This will allow the development of a new generation of highly tumour-specific drugs and therapeutic strategies, e.g. Cancer Target Discovery CTD bridges the gap between the enormous volumes of data generated by genomic characterization studies and the ability to use these data for the development of human cancer therapeutics) <https://ocg.cancer.gov/programs/ctd2> [7].

Druggable gene: Subset of genes which express proteins capable of binding small drug like molecules.

Ligand, it is a (usually small) molecule that binds to a biological macromolecule [7].

Receptor: It is any functional macromolecule in a cell which a drug binds to produce its effect. So what exactly dose this mean, a receptor is like a light switch, it has two functions: ON and OFF, can say the switch has been turned ON by **Agonist** which is a drug that stimulates a cell receptor that would normally be stimulated by naturally occurring substances in a person's body. For example if a patient was being treated for congestive heart failure, meaning the patient's heart could not pump enough blood to the rest of his body, and would need to increase his cardiac output. In order to increase the patient's cardiac output could give the patient **Dobutamine**, a drug that directly mirrors the effects of **Norepinephrine**, a chemical made by the body, (Fig 2; a) shows this idea. While the switch has been trund OFF by **Antagonist**, which is a drug that interferes with the physiological action of another substance, especially by combining with and blocking its receptor. For example if a patient arrived in the emergency room with an opioid overdose of **Hydrocodone**, would need to prevent the medication from taking effect to prevent the **Hydrocodone** from taking effect would give him **Naloxone** to block the receptors in the body, thereby preventing the **Hydrocodone** from binding and causing an overdose, (Fig

2; b) shows this idea [8, 9]. Finally (Fig2; c) shows how ligand and receptor binding together.



1.2 Insilico Methods in Drug Discovery

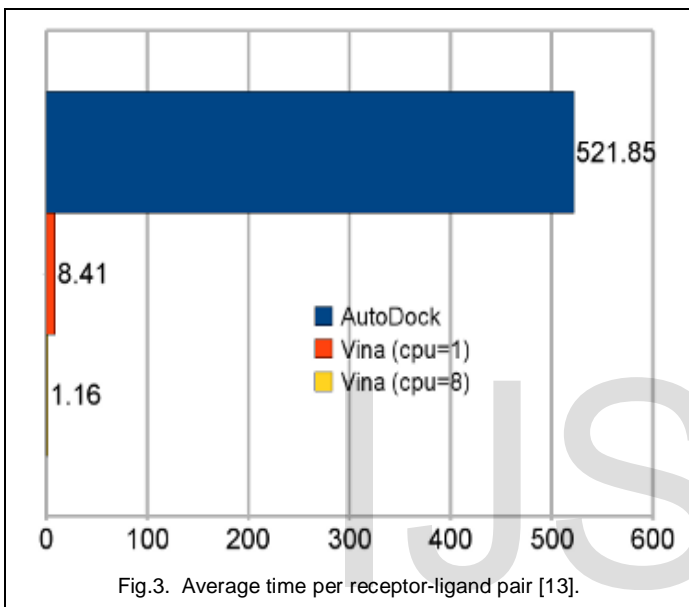
Using bioinformatics companies can generate more and more drugs in a short period of time with low risk. Drugs are usually only developed when the particular drug target for those drugs' actions have been identified and studied. In silico methods can help in identifying drug targets via bioinformatics tools, 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.

The use of computers and computational methods permeates all aspects of drug discovery today and forms the core of structure-based drug design. High-performance computing, data management software and internet are facilitating the access of huge amount of data generated and transforming the massive complex biological data into workable knowledge in modern day drug discovery process. The use of complementary experimental and informatics techniques increases the chance of success in many stages of the discovery process, from the identification of novel targets and elucidation of their functions to the discovery and development of lead compounds with desired properties [11, 12]. There are many Insilico methods in drug discovery in this research will use Molecular docking method.

2 RLEATED WORK

Oleg Trott (<http://vina.scripps.edu/index.html>) was designed and implemented AutoDock Vina as an open-source program for doing molecular docking in the Molecular Graphics Lab at the

Scripps Research Institute. AutoDock Vina significantly improves the average accuracy of the binding mode predictions compared to AutoDock 4. AutoDock Vina has been tested against a virtual screening benchmark called the [Directory of Useful Decoys](#) by the Watowich group, and was found to be "a strong competitor against the other programs and at the top of the pack in many cases". It should be noted that all of the other docking programs are distributed commercially. AutoDock Vina tends to be faster than AutoDock 4; Vina can take advantage of multiple CPUs or CPU cores on the system to significantly shorten its running time Fig. 3 shows how autodock vina with multiple CPU has shortest running time [13].



The motivation of previous technique implemented using linux platform, but for ease use will implemented the above properties of autodock vina using windows platform which is commonly used, although facing some obstacles will solve them as shown in the experimental results section later.

3 PROPOSED OF MOLECULAR DOCKING METHOD

The general tasks of molecular docking method for drug discovery can summarized as flowchart shown in Fig. 4.

4 EXPERIMENTAL RESULTS

4.1 Configuration of Molecular docking

Docking or Computer aided drug designing: can be broadly classified as "Receptor based methods" which make use of the structure of the target protein and "Ligand based methods" which is based on the known inhibitors Fig. 5, shows molecular docking. It is the computational determination of binding affinity between molecules (protein structure and ligand). Given a protein and a ligand find out the binding free energy of the complex formed by docking them. Fig. 6 shows

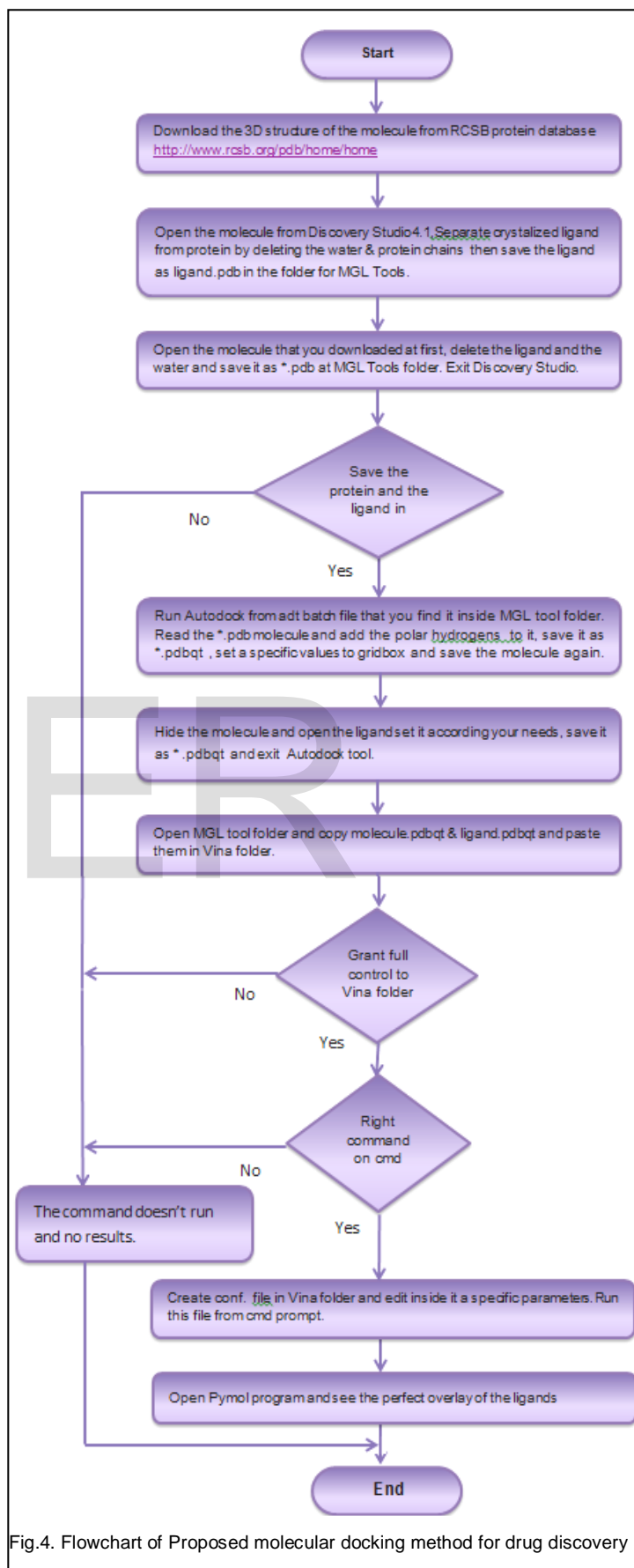


Fig.4. Flowchart of Proposed molecular docking method for drug discovery the stages of high throughput docking for protein-ligand

complex binding.

What are the programs installation needed to start docking?

1. Discovery Studio 4.1 Client
2. AutoDockTools Vina
3. AutoDockTools-1.5.6
4. Python 3.4
5. PyMol Molecular graphic system.

And if needed to feel that works under Linux instead of windows recommend the install:

- o Cygwin Terminal <https://www.cygwin.com/> : a large collection of GNU and Open Source tools which provide functionality similar to a Linux distribution on Windows(Linux feeling - on Windows) and this is optional can work under windows as shown in the next article .

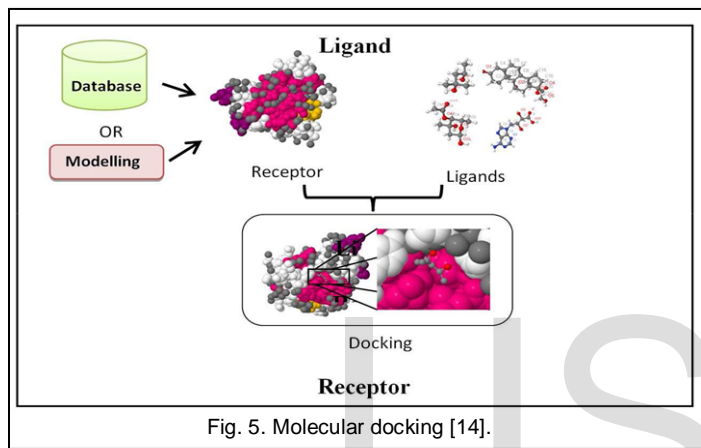


Fig. 5. Molecular docking [14].

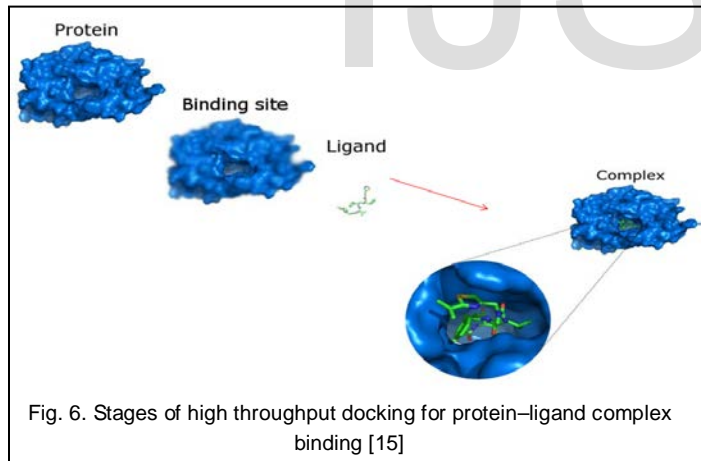


Fig. 6. Stages of high throughput docking for protein–ligand complex binding [15]

4.2 Implementation of Molecular docking

Therapy takes multiple techniques some of them related with replacement the mutations in genes caused diseases [16], some other techniques focus on molecular docking in drug discovery, etc. The HIV (human immunodeficiency virus)-1 protease is a member of the aspartyl protease family and is essential for the life-cycle of HIV. Inhibition of this protease prevents maturation of HIV particles and has therefore, been the focus for the design and development of many drugs that inhibit this enzyme.

1hsg is a crystal structure at 1.9 angstroms resolution human immunodeficiency virus (HIV).

Now start the docking:

- Download [1_HSG](http://www.rcsb.org/pdb/home/home.do) from RCSB protein database <http://www.rcsb.org/pdb/home/home.do> Which is saved as 3D structure.
- Open this file from Discovery Studio 4.1. Fig.7. shows this file.

Separate crystalized ligand from protein:

- Go to Scripts-Ligand Interaction-Show ligand Binding SiteAtoms to show label residues.
- Go to: Scripts-Selection-Select Water Molecules, Edit -Delete, Scripts-Selection-Select protein chains and Edit -Delete.
- Save the ligand as ligand.pdb in the folder for MGL Tools.

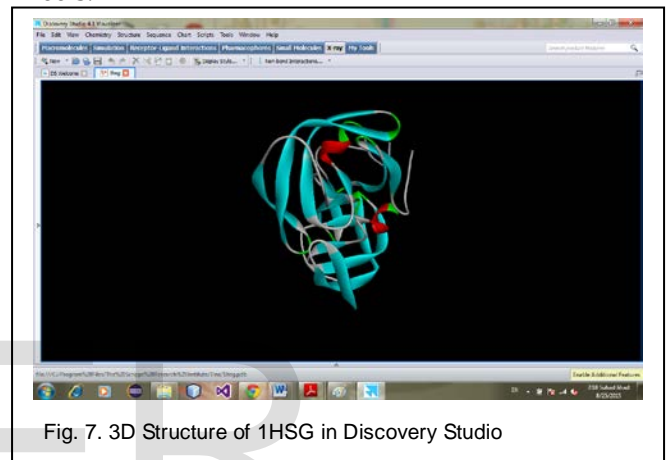


Fig. 7. 3D Structure of 1HSG in Discovery Studio

- Open 1hsg.
- Go to: Script -Selection-Ligand, Edit-delete, Script-Select-Select Water Molecules and Edit-delete.
- Save as 1hsg.pdb at MGL Tool folder.
*Note: if don't save them in the MGL Tool folder the result doesn't appear.
- Exit Discovery Studio.
- Now run autodock from MGLTool folder but must be ensure the two files are inside MGLTool and run autodock from adt batch file only, then find this file in MGL Tool folder also.
- In Autodock Tools program go to file-read molecule and open 1hsg.
- Go to: Edit-Add-hydrogens-only polar.
Grid-macromolecule-choose-1hsg-select molecule
- Then the molecule will initialized save it as 1hsg.pdbqt
- Go to: Grid-gridbox-then set the values as:
No. of points in x-direction=26,
No. of point in y-direction=26,
No. of points in z-direction=26,
Spacing =1.000, X-center=16.072,
Y-center=26.5007 and Z-center= 3.7748.
- Now in this grid option dialog go to File-close current

saving

- Hide this molecule.
- Open the ligand by Go to:
Ligand -input- open and choose a ligand file
Ligand - TorsionTree -Set Number of Torsions
Put No. 6 and press dismiss
Ligand-Output-Save as .pdbqt and File -Exit
- Open MGLTool folder and copy (1hsg.pdbqt & lig-and.pdbqt) and paste them in Vina folder in the computer.
- In vina folder create text document file name it as conf. and write in it :
receptor=1hsg.pdbqt, ligand=ligand.pdbqt,
out=out.pdbqt, center_x=16.072,
center_y=26.5007, center_z=3.7748,
size_x=26, size_y=26, size_z=26 and
exhaustiveness=8.
- Save and close conf. file.
- Run cmd prompt :
- Change the directory to vina folder in the computer
C:\ProgramFiles\TheScrippsResearchInstitute\ vina.
- Write on this folder Vina - -config conf.txt - -log
log.txt but be careful on every wrong space
because this prevents the execution of the command.

The execution of this command shown in Fig. 8.

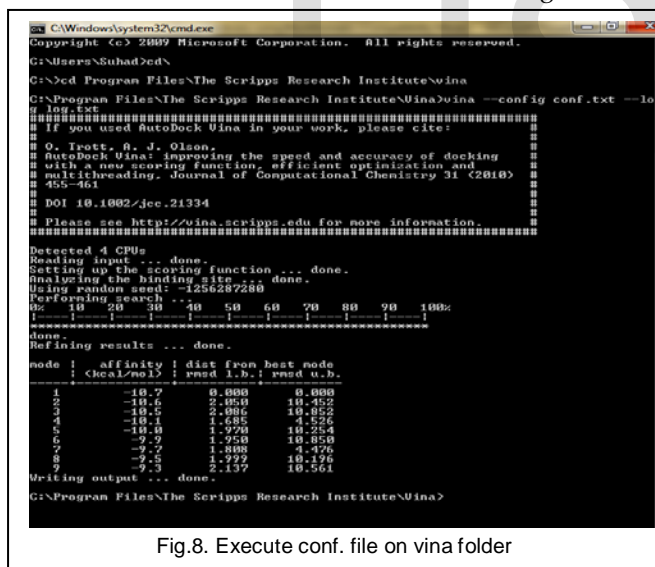


Fig.8. Execute conf. file on vina folder

*Useful Hint : now may get a message that tell windows couldn't read the file conf.txt to solve this problem go to the folder Vina on the computer and right click on it, choose Security, click Users and check Full controll then Apply. Fig.9 shows exactly what needed to do.

- After the execution, write vina_split --input out.pdbqt the output will look like Fig.10.
- Now open vina folder and see the nine out_ligands files. Then can check each of them. Fig.11 shows the nine ligands which has been produced on folder vina.
- At the last stage open Pymol and open 1hsg that download from RCSB and ligands that produced to

check the best docking parameters, and can open the ligands by click on the letter A on the right side then a menu appear select from it preset which open another menu select from it ligands. Fig.12 shows Pymol program.

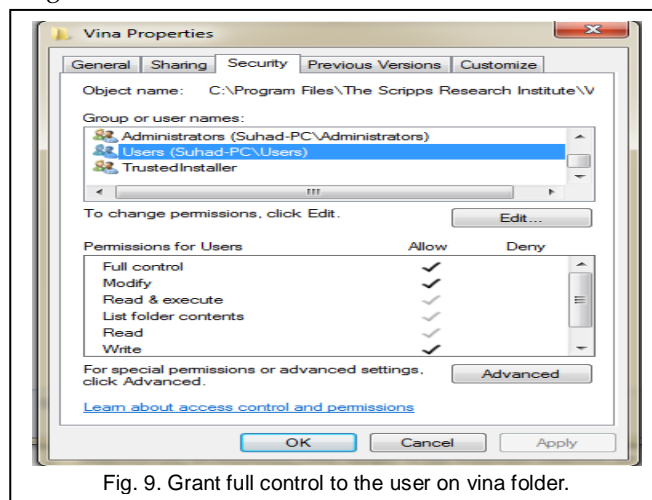


Fig. 9. Grant full control to the user on vina folder.

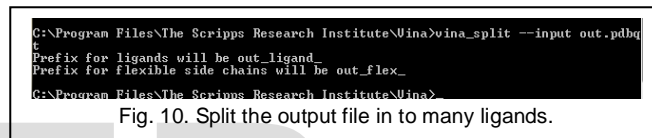


Fig. 10. Split the output file in to many ligands.

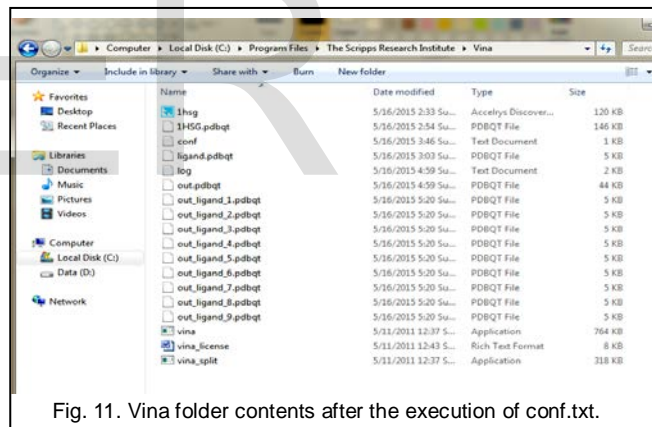


Fig. 11. Vina folder contents after the execution of conf.txt.

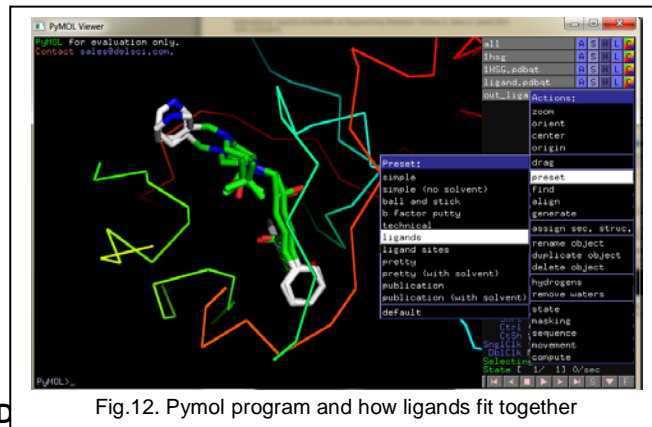


Fig.12. Pymol program and how ligands fit together

4.3 D

The results of proposed molecular docking method in experimental results section (4) are same to which obtained from method of Oleg, and Table 1 shows comparing the pro-

posed method with Oleg.

5 CONCLUSIONS AND FUTURE WORK

Every insilico drug discovery researcher know that performing docking in Windows PC using Autodock tools is a tedious

TABLE 1
REVEALS COMPARISON OF PROPOSED METHOD WITH OTHER RELATED METHOD

Proposed molecular docking method based on Windows platform	Molecular docking method of Oleg [13]
The configuration needed programs installation to start docking as referring to in subsection 4.1 (Discovery Studio 4.1 Client, AutoDockTools Vina, AutoDockTools-1.5.6, Python 3.4 and PyMol Molecular graphic system)	The configuration If working under Linux instead of windows recommended to install as referring to in subsection 4.1: Cygwin Terminal https://www.cygwin.com/
Solve the Cygwin of docking under windows platform	Need Cygwin which create the Linux or Windows environment
Solving all bugs/gaps that faced under windows platform.	Never referring to any bugs/gaps can face or how can solve them.
Using a AutoDock Vina as an open-source program for doing molecular docking	Using same program (AutoDock Vina) for doing molecular docking

job under windows but with autodock vina can enjoy this docking by simply solve some bugs/gaps that faced when docking such bugs and their solutions as follow:

1. In the run stage windows shows Error: could not open "conf.txt" for reading to solve this problem go to folder vi-na and grant it full control.
2. If pass (1) above and trying to write Vina - -config conf.txt - -log log.txt then the result does not appear, notice the spaces in the sentence and modify them till execute it.
3. The input files must be saved in MGL folder.
4. Run Autodock from MGLtool by double click adt batch file.

By solving the above problems will get the same results that Oleg got them with the docking as shown in section 4.

Future work of pharmacogenomics is the study of how genes affect a person's response to drugs. This relatively new field combines pharmacology (the science of drugs) and genomics (the study of genes and their functions) to develop effective, safe medications and doses that will be tailored to a person's genetic makeup. The field of pharmacogenomics is still in its infancy. Its use is currently quite limited, but new approaches are under study in clinical trials. In the future, pharmacogenomics will allow the development of tailored drugs to treat a wide range of health problems, including cardiovascular disease, Alzheimer disease, cancer, HIV/AIDS, and asthma [17]. Fig. 13 shows the hope of future medicine.



Fig.13. Pharmacogenomics.

References

- [1] Ayad Ghany Ismaeel, "New Approach for Prediction Pre-cancer via Detecting Mutated in Tumor Protein P53", International Journal of Scientific & Engineering Research, Volume 4, Issue 10, June-2013. Pages 463-468.
- [2] Role of bioinformatics and pharmacogenomics in drug discovery and development process – Springer Pramod Katara (1) Department of Bioscience and Biotechnology, Banasthali University, Banasthali, 304022, India (2) Present address: Center for Bioinformatics, IIDS, University of Allahabad, Allahabad, 211007, India Pramod Katara.Email: pmkatara@gmail.com Received: 28 February 2013.Revised: 27 July 2013.Accepted: 29 July 2013. Published online: 10 August 2013
- [3] Personalized Medicine - NIH News in Health, December 2013.html A monthly newsletter from the National Institutes of Health, part of the U.S. Department of Health and Human Services.
- [4] Drug Design Methodology, Concepts, and Mode-of-Action Springer 2013 Authors: Klebe, Gerhard.
- [5] <http://www.scls.riken.jp/eng/newsletter/Vol.7/special04.html>
- [6] AutoDock4.2.6_UserGuide.pdf. *Automated Docking of Flexible Ligands to Flexible Receptors*.Garrett M. Morris, David S. Goodsell, Michael E. Pique, William "Lindy" Lindstrom, Ruth Huey, Stefano, Forli, William E. Hart, Scott Halliday, Rik Belew and Arthur J. Olson. Modification date: July 28, 2014 15:30 D7/P7.
- [7] Target discovery using bioinformatics.html Threading algorithms and other state-of-the-art techniques allow researchers to identify the biochemical targets used in drug design. Mark Swindells,Richard Fagan. June 2001.Vol. 31, No. 6, pp 24-28.
- [8] <http://www.slideshare.net/carissa-matlock/pharm-receptors-ppt>.
- [9] *Computer-Aided Drug Design: An Innovative Tool for Modeling* Pranita P. Kore, Madhavi M. Mutha, Rishikesh V. Antre, Rajesh J. Oswal, Sandip S. Kshirsagar Medicinal Chemistry Research Laboratory, JSPM's Charak College of Pharmacy & Research, Pune, India Email: rishiantre@gmail.com Received October 1, 2012; revised November 5, 2012; accepted November 15, 2012.
- [10] <http://pharmacymagazine.blogspot.com/2013/11/how-do-drugs-work.html>

- [11] Turning Docking and Virtual Screening as simple as it can get....html
<http://www.fc.up.pt/pessoas/nscerque/vsLab/vLab/HomePage.html>
- [12] Molecular Docking for Drug Discovery: Machine-Learning Approaches for Native Pose Prediction of Protein-Ligand Complexes Hossam M. Ashtawy and Nihar R. Mahapatra Springer International Publishing Switzerland 2014 E. Formenti et al. (Eds.): CIBB 2013, LNBI 8452, pp. 15-32, 2014. DOI: 10.1007/978-3-319-09042-9_2
<http://vina.scripps.edu/index.html>
- [13] <http://www.intechopen.com/books/an-integrated-view-of-the-molecular-recognition-and-toxinology-from-analytical-procedures-to-biomedical-applications/computer-based-methods-of-inhibitor-prediction>
- [14] <http://www.slideshare.net/baoilleach/proteinligand-docking>
- [15] Ayad Ghany Ismaeel, " A Computer-Assisted Therapy: New Methodology of Predict Restricting Enzymes to Replace Mutations Gene Caused Cancers", International Journal of Scientific & Engineering Research, Volume 5, Issue 8, August-2014. Pages 437-441.
- [16] Genetic home references
<http://ghr.nlm.nih.gov/handbook/genomicresearch/pharmacogenomics>
- [17] Bioinformatics and Drug Discovery edited by Richard S. Larson October 1, 2005.
- [18]

IJSER

- [14] <http://www.intechopen.com/books/an-integrated-view-of-the-molecular-recognition-and-toxinology-from-analytical-procedures-to-biomedical-applications/computer-based-methods-of-inhibitor-prediction>
- [15] <http://www.slideshare.net/baoilleach/proteinligand-docking>
- [16] Ayad Ghany Ismaeel, " A Computer-Assisted Therapy: New Meth-