International Journal of Current Research and Academic Review

ISSN: 2347-3215 Special Issue-3 (August-2016) pp. 207-218

Journal home page: http://www.ijcrar.com

Molecular Docking Studies of Allylsulfur Compounds from *Allium sativum* against EGFR Receptor

R. Padmini[1,2*], Razia Muthusamy[2], V. Roopavahini[1] and S. Sai Aruna[1]

[1]Department of Biochemistry and Bioinformatics and Clinical Trial Management, Dr. MGR Janaki College of Arts and Science, Chennai, Tamil Nadu, India
[2]Department of Biotechnology, Mother Teresa Women's University, Kodaikanal, Tamil Nadu, India

Corresponding author

KEYWORDS

Lung cancer, *Allium sativum*, Allylsulfur compounds, EGFR, Molecular docking

A B S T R A C T

Lung cancer is an abnormal cell growth that starts off in the cell lining of one or both lungs. Two main types of lung cancer, they are Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). NSCLC is the most common type of lung cancer. About 85% of lung cancers are NSCLC. The sub types of NSCLC are squamous cell carcinoma, adenocarcinomal and large cell carcinoma. EGFR (Epidermal Growth Factor Receptor) is a receptor protein which belongs to Receptor Tyrosine Kinase family (RTKs). EGFR is the key paradigm of molecular targeted therapy in lung cancer. EGFR plays an important role in the development of the malignant phenotype of many cancers. Activated EGFR induces multiple downstream pathways which are involved in cell survival and cell proliferation. Garlic (*Allium sativum*) is a vegetable that belongs to *allium* family. Various research findings suggest the relationship between excess garlic intake and reduction in risks of various cancers. An Allylsulfur compound (Allicin, Alliin, s-Allylcysteine, s-Allylmercaptocysteine) of garlic shows anticancer activity. These compounds inhibit the gene which is responsible for lung cancer. Molecular docking studies using AutoDock provides the comprehensive overview of inhibition of EGFR through the binding of the above allylsulfur compounds.Docking studies of the bioactive compounds from garlic which has anticancer activity with EGFR showed that alliin and s-allylmercaptocystein (ligand) has good binding energies. From the docking result, alliin and s-allylmercaptocystein forms 2 hydrogen bonds. Based on docking energies and hydrogen bond interactions, alliin and s-allylmercaptocystein plays a key role in inhibiting the EGFR receptor protein.

Introduction

Cancer is a collection of heterogeneous genetic diseases united by common alterations in multiple cellular signaling pathways (Luo *et al*., 2009). This is the second most common disease after cardiovascular disorders for maximum deaths in the world (Jemal *et al*., 2007). The International Agency for Research on Cancer (IARC) estimates that there were 12.7 million new cancer cases in 2008 and that this number is expected to grow to 21.4 million by 2030(Ferlay *et al*., 2010).

Lung cancer is the second most common cancer in incidence but the leading cause of cancer deaths in men and women. Lung cancer forms in tissues of the lung (shown in Figure 1), usually in the cells lining air passages. The primary risk factor for lung cancer is smoking, which accounts for more than 85% of all lung correlated deaths (Doll *et al*., 1976).

Common symptoms of lung cancer include cough, dyspnea, weight loss, and chest pain; symptomatic patients are more likely to have chronic obstructive pulmonary disease. Lung cancer is categorized into Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC).

NSCLC is further classified into squamous cell carcinoma, adenocarcinoma, and large cell carcinoma (Knop *et al*., 2005).NSCLC represents 80% of all lung cancers, with adenocarcinoma accounting for 40% of all cases of lung cancer. Squamous cell carcinoma occurs most frequently in the central zone of the lung whereas adenocarcinoma tumors are peripheral in origin, arising from the alveolar surface epithelium or bronchial mucosal glands. Large cell carcinoma composes only 15% of all lung cancers and appears to be decreasing in incidence because of improved diagnostic techniques (Van Cleave *et al*., 2004).

Molecular signaling pathways of growth factor receptors, such as the Epidermal Growth Factor (EGF) Receptor (EGFR), stimulate cancer cell growth, survival and resistance to cytotoxic therapy. Binding of EGF activates EGFR and initiates signal transduction pathways, including the phosphatidylinositol 3-kinase (PI3K)-Akt and Ras-Erk, resulting in cell proliferation and survival (Mitsudomi and Yatabe *et al*., 2010).Increased EGFR signaling is linked to cancer and abnormal EGFR expression is found in approximately 80% of cases of NSCLC (Mitsudomi and Yatabe *et al*., 2010).

Garlic (*Allium sativum*), a member of the family Liliaceae, contains an abundance of chemical compounds that have been shown to possess beneficial effects to protect against several diseases, including cancer (shown in Figure 2). The name ‘‘*Allium sativum*’’ is derived from the Celtic word ‘‘*all*’’, meaning burning or stinging, and the Latin ‘‘*sativum*’’ meaning planted or cultivated (Mahady *et al*., 2001; Srivastava *et al*., 1995).A recent meta-analysis also showed that a high intake of garlic may be associated with decreased risks for stomach and colorectal cancer (Fleischauer *et al*., 2000).Garlic can strengthen immune system, which is vitally important for fighting cancer. Garlic contains several potentially important agents that possess anti tumor and anti carcinogenic properties. It contains allylsulfur compounds that slow or prevent the growth of tumor cells.

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 (Lengauer and Rarey *et al*., 1996). Docking is mostly used for finding the

binding between the ligand and the receptor. Hence in drug designing docking plays a vital role (Kitchen *et al*., 2004). Between two molecules, the binding affinities strength is predicted using the preferred orientation. For docking, 3D structure of the protein and ligands was required as the input, for which the bound conformation of the ligand with that of the protein active site is predicted (Rarey *et al*., 1995).

Materials and Methods

Preparation of protein

Uniprot

UniProt is a comprehensive, high-quality and free online database of protein sequence and functional information, mainly derived from genome sequencing projects. It contains a large amount of information about the biological function of proteins derived from the research literature. The UniProt/Swissprot Knowledgebase UniProtKB is the central access point for extensive curated protein information, including function, classification, and cross-reference (Uniprot C, "The Universal Protein Resource UniProt in 2010". Nucleic Acids Res, 38 Database issue: D142–D148, 2009; http://www.uniprot.org/).

PDB

The PDB is a repository for 3-D structural data of proteins and nucleic acids. These data, typically obtained by X-ray crystallography or NMR Spectroscopy, are submitted by biologists and biochemists from around the world, are released into the public domain, and can be accessed for free. The mission of the PDB is to maintain a single Protein Data Bank Archive of macromolecular structural data that is freely and publicly available to the global community (www.rcsb.org/pdb).

ACD Chem Sketch

Advanced chemistry development's ACD/chem sketch is a chemically intelligent drawing interface that allows you to draw almost any chemical structure including organics, organometallics, polymers, and markush structures. Use it to produce professional looking structures and diagrams for reports and publications (www.acdlabs.com/../chemsketch/).

Open Babel

Open Babel is a chemical toolbox designed to speak the many languages of chemical data. It's an open, collaborative project allowing anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas (http://openbabel.org/).

AutoDock

AutoDock is a suite of automated docking tools. The software is used for modelling flexible small molecules such as drug molecules and its binding to receptor proteins of known three dimensional structures. It uses Genetic Algorithms for the conformational search and is a suitable method for the docking studies. The technique combines simulated annealing for conformation searching with a rapid grid based method of energy evaluation. Auto Dock tools are used to prepare, run and analyze the docking simulations, in addition to modeling studies (http://autodock. scripps.edu/resources/tools).

PyMOL

PyMOL is an open-source tool to visualize

molecules available from (**z**. It runs on Windows, Linux and MacOS equally well. PyMOL has excellent capabilities in creating high-quality images from 3D structures, it has well developed functions for manipulating structures and some basic functions to analyze their chemical properties. The possibilities to write scripts and plugins as well as to incorporate PyMOL in custom software are vast and superior to most other programs. PyMOL has been written mostly in the Python language (www.python.org), while the time-critical parts of the system have been coded in C. This way, Python programs interact most easily with the PyMOL GU

Results and Discussion

Sequence retrieval: EGFR

The sequence of Epithelial growth factor receptor –tyrosine kinase is retrieved from UNIPROT database and sequence accession number is P00533 from *Homo sapiens* (Human).

Structure retrieval

The three dimensional structure (crystal structure) of EGFR is derived from PDB database. From the PDB databank (Berman *et al*., 2000; www.pdb.org), the PDB file was collected and the PDB ID was 4ZSE. The structure was visualized using RASMOL. The final stable structure of the EGFR protein obtained is shown in Figure 1**,** alpha helix is shown in pink color, beta sheet are shown in yellow color, white colors showing turns.

Preparation of ligands

The four allylsulfur compounds which shows anti cancer activity from *Allium sativum* is used as ligand and was retrieved from NCBI PubChem Compound database (Bolton et al., 2008; http://pubchem. ncbi.nlm.nih.gov/). For further docking analysis, the ligands were designed using ChemSketch and their 2d structure was converted to 3D structures . These data are saved as a molecular format file (MDL MOL format). The molecular format converter tool (Open Babel) is used to convert this file into the PDB format and is used during docking analysis. The structure and molecular formula of allylsulfur compounds of *Allium sativum* was shown in Table 1.

Molecular docking study of the allylsulfur compounds against EGFR

The 4 allylsulfur compounds (Allicin, Alliin, s-Allylcysteine, s-Allylmercaptocysteine) are docked against EGFR protein.

The Graphical User Interface program "AutoDock Tools" was used to prepare, run, and analyze the docking simulations. Koll man united atom charges, solvation parameters and polar hydrogens were added into the receptor PDB file for the preparation of protein in docking simulation.

Auto Dock (Good sell *et al*., 1996; Jones *et al*., 1997; Rarey *et al*., 1996) requires precalculated grid maps, one for each atom type present in the flexible molecules being docked and its stores the potential energy arising from the interaction with rigid macromolecules. This grid must surround the region of interest in the rigid macromolecule. The grid box size was set at 126, 126 and 126A° (*x, y*, and *z*) to include all the amino acid residues that present in rigid macromolecules.

Table.1 The Ligand (inhibitor) molecules used for docking studies

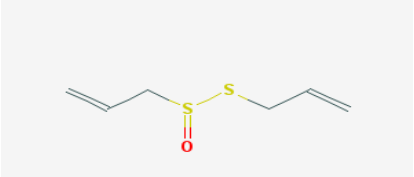
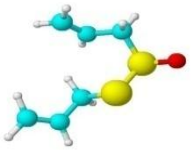
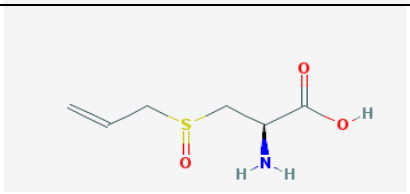
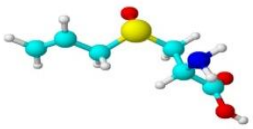
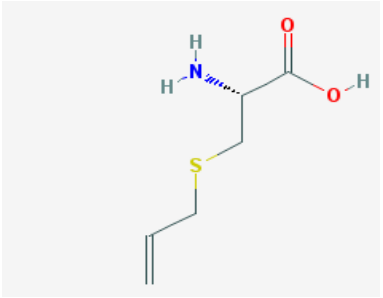
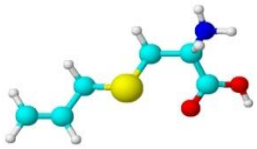
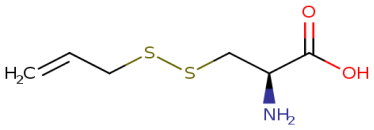
S.No	Compounds	Molecular Formula	2D Structure	3D Structure
1	ALLICIN	$C_6H_{10}OS_2$		
2	ALLIIN	$C_6H_{11}NO_3S$		
3	S-ALLYL CYSTEINE	$C_6H_{11}NO_2S$		
4	S-ALLYL MERCAPTO CYSTEINE	$C_6H_{11}NO_2SO_2$		

Table.2 Lipinski's properties of the compounds of *Allium sativum*

Ligand molecule	Molecular weight	Xlogp3 value (<=5)	H-bond donar	H-bond acceptor
Allicin	162.273g/mol	1.3	0	3
Alliin	177.22144g/mol	-3.5	2	5
s-Allylcysteine	161.22204g/mol	-2.1	2	4
s-Allylmercaptocysteine	193.287g/mol	-1.8	2	3

Table.3 Docking interaction between EGFR and Allicin

EGFR		Allicin (Atom)	Distance	Binding energy
Residue	Atom			
MET-793	N	O	2.9	-3.56

Table.4 Docking interaction between EGFR and alliin

EGFR		Alliin (Atom)	Distance	Binding energy
Residue	Atom			
ASP-770	ODI	H	2.2	-3.16
ASN-771	N	O	2.9	
ASN-771	O	H	2.3	
ASN-771	O	O	3.4	
VAL-774	O	O	3.0	
VAL-774	O	H	2.4	
LYS-852	NZ	O	2.8	

Table.5 Docking interaction between EGFR and s-Allylcysteine

EGFR		S-allylcysteine (Atom)	Distance	Binding energy
Residue	Atom			
LYS-949	NZ	O	2.9	-3.0
LYS-949	NZ	O	3.2	
GLU-967	OE1	H	2.2	
GLU-967	OE2	H	2.1	

Table.6 Docking interaction between EGFR and s-Allylmercaptocysteine

EGFR		S-allylmercaptocysteine (Atom)	Distance	Binding energy
Residue	Atom			
ARG-776	NE	O	3.1	-2.99
ASP-770	ODI	H	1.8	

Table.7 The key amino acid residues of EGFR involved in the hydrogen bonding interaction with various ligands studied during molecular docking.

Ligands	**Binding Energy (Kcal/mol)**	**Amino acid and atoms involved in Hydrogen bond interaction**	**No.of Hydrogen Bonds**
Allicin	-3.56	MET-793(N)	1
Alliin	-3.16	ASP-770(H),ASN-771(O),ASN-771(H), ASN-771(O),VAL-774(O),VAL-774(H), LYS-852(O)	2
s-Allylcysteine	-3.0	LYS949(NZ),LYS949(NZ),GLU967(OE), GLU-967(OE2)	1
s-Allylmercaptocysteine	-2.99	ARG-776(NE), ASP-770(ODI)	2

Fig.1 Lung Cancer Proliferation

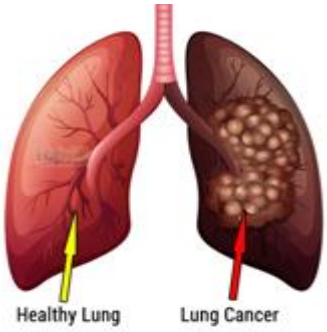


Fig.2 Plant of *Allium sativum*

Fig.3 Crystal structure of the EGFR

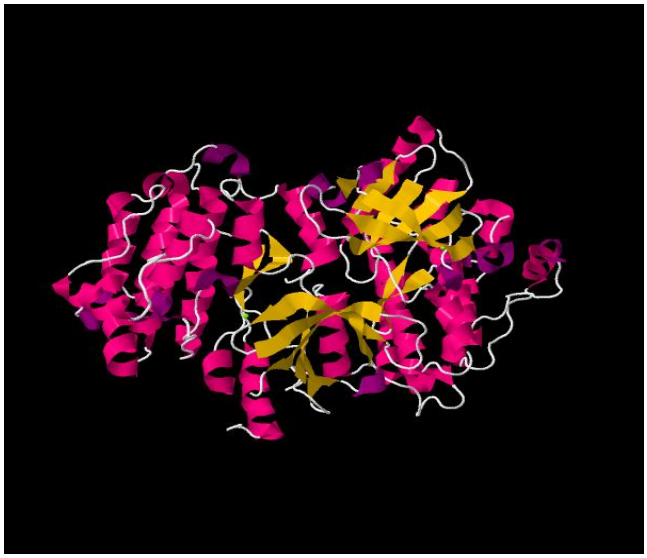
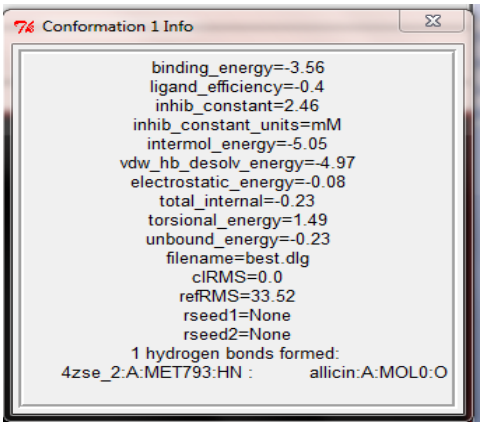
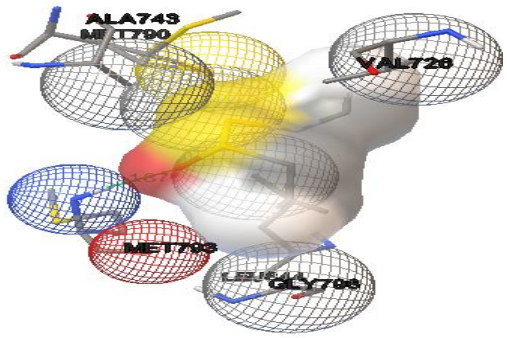
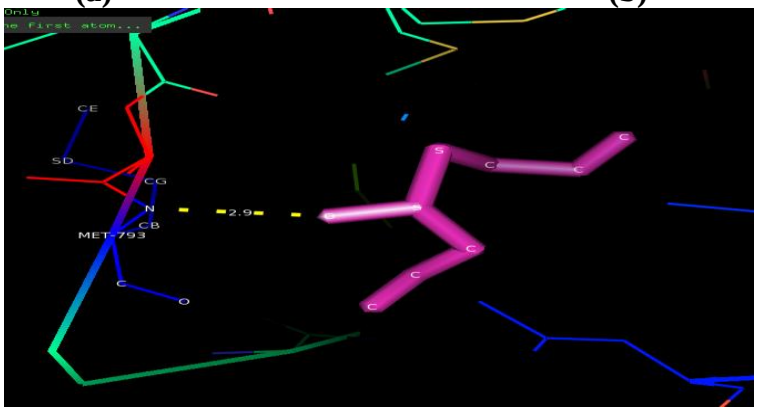


Fig.4 Docking of EGFR and Allicin (a)Binding energy score (b)Interaction between EGFR and Allicin was visualized using Autodock (c)Hydrogen bond interactions between EGFR and Allicin was visualized using Pymol
(ligand in pink color and protein in blue color)



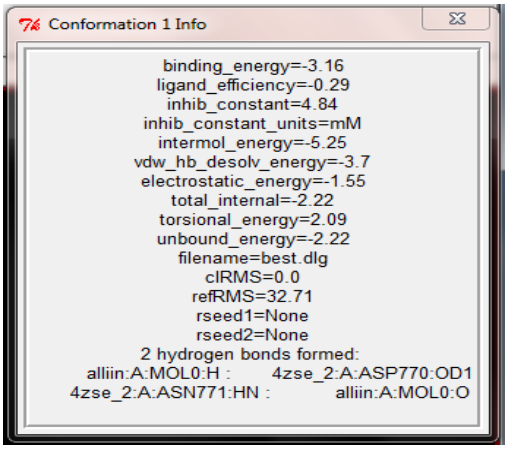


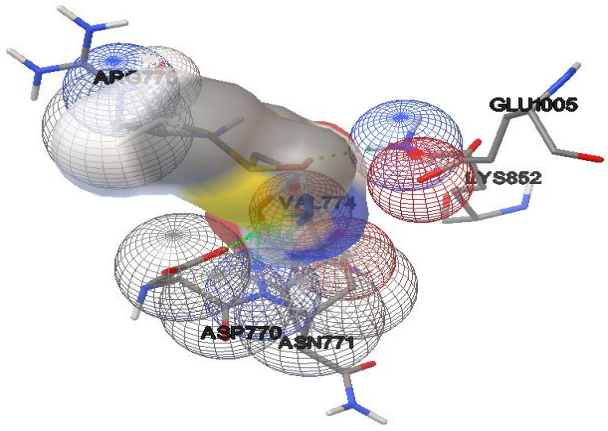
(a) **(b)**



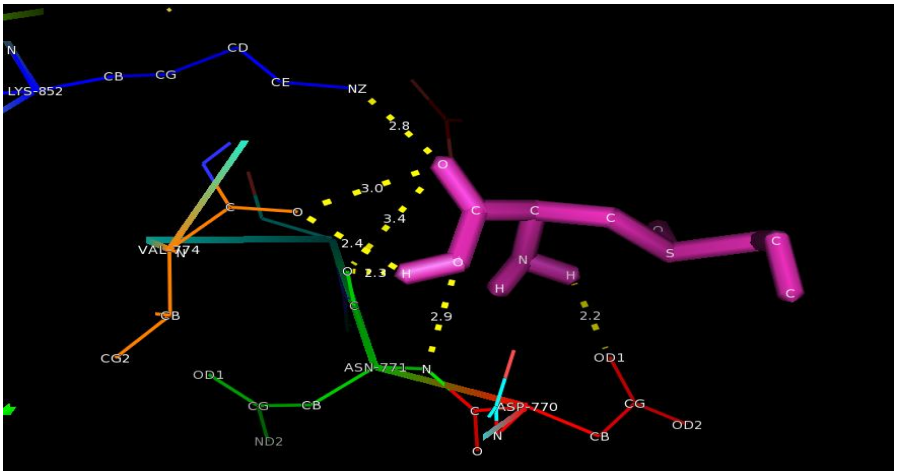
(c)

Fig.5 Docking of EGFR and Alliin (a) Binding energy score (b) Interaction between EGFR and Alliin was visualized using Autodock (c)Hydrogen bond interaction between EGFR and Alliin was visualized using Pymol
(ligand in pink color and protein in blue, green and red color)



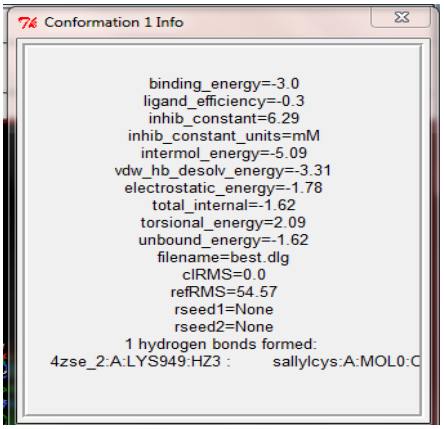


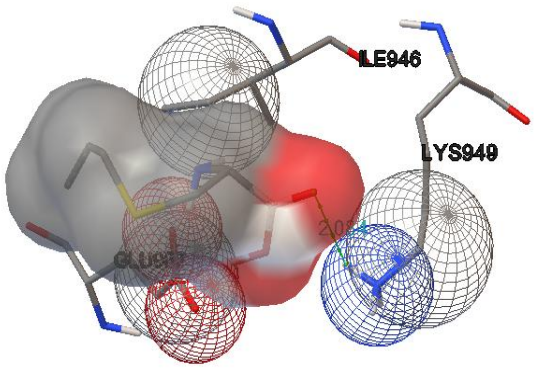
(a) **(b)**



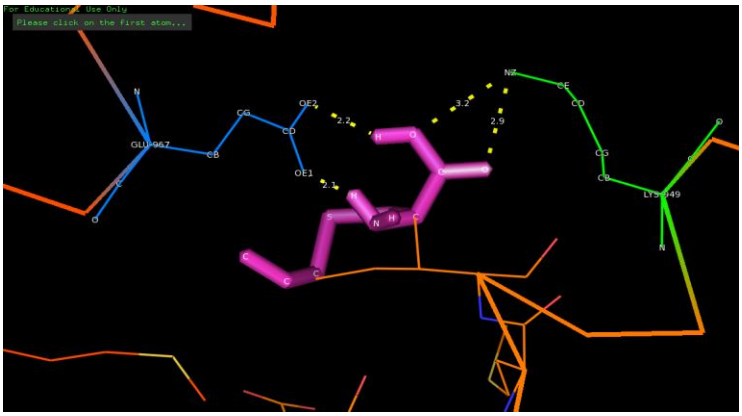
(c)

Fig.6 Docking of EGFR and s-Allylcysteine (a) Binding energy score (b) Interaction between EGFR and s-Allylcysteine was visualized using Autodock (c)Hydrogen bond interaction between EGFR and s-Allylcysteine was visualized using Pymol
(ligand in pink color and protein in blue and green color)



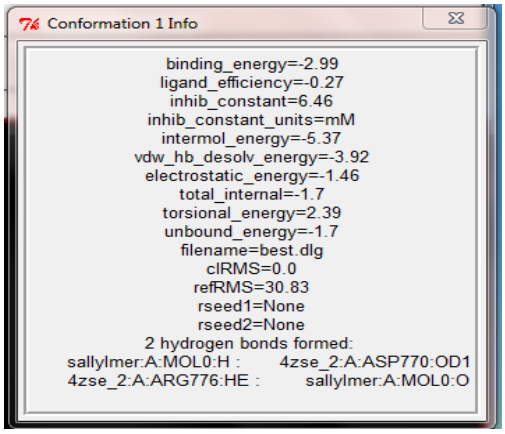


(a) **(b)**

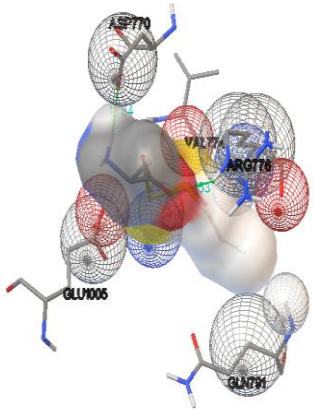


(c)

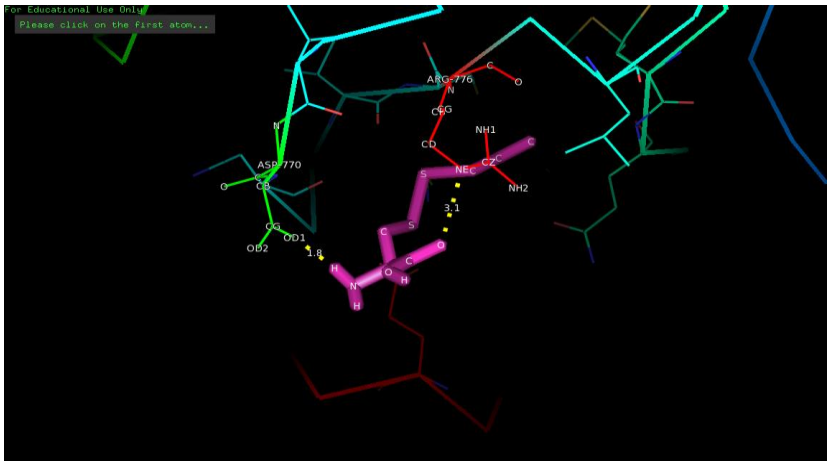
Fig.7 Docking of EGFR and s-Allylmercaptocysteine (a)Binding energy score (b) Interaction between EGFR and s-Allylmercaptocysteine was visualized using Auto dock (c) Hydrogen bond interaction between EGFR and
s-Allylmercaptocysteine was visualized using Pymol
(ligand in pink color and protein in green and red color)



(a)



(b)



(c)

AutoGrid 4.2 Program, supplied with AutoDock 4.2 was used to produce grid maps. The spacing between grid points was 0.375 angstroms. The Lamarckian Genetic Algorithm (LGA) (Morris *et al*., 1998) was chosen search for the best conformers. During the docking process, a maximum of 10 conformers was considered. The population size was set to 150 and the individuals were initialized randomly. Maximum number of energy evaluation was set to 25,00,000, maximum number of generations 27,000, maximum number of top individual that automatically survived set to 1, mutation rate of 0.02, crossover rate of 0.8, Step sizes were 0.2 A for translations, 5.0° for quaternions and 5.0° for torsions. Cluster tolerance 0.5Ao,

external grid energy 1,000.0, max initial energy 0.0, max number of retries 10,000 and 10 LGA runs was performed. The best ligand-receptor structure from the docked structures was chosen based on the lowest energy and minimal solvent accessibility of the ligand.

GFR the macromolecule and the ligands (Allicin, Alliin, s-Allylcysteine, s-Allyl mercaptocysteine) were docked using AutoDock 4.2.The binding energy of allylsulfur compounds with the active site of the EGFR during docking with ten conformations was generated. Hydrogen bond interactions and the distance between the donors and acceptors were measured for the best conformers (Archana *et al*., 2010). The Binding energy is correlated with the probability of affinity and stable bound between ligand and its receptor. Binding energy values may also predict the bioactivity value for a ligand to the corresponding receptor (Kartasasmita *et al*., 2009). Allicin (-3.56Kcal/Mol), Alliin (-3.16 Kcal/Mol), s-Allylcysteine (-3.0Kcal/Mol), s-Allylmercaptocysteine (-2.99Kcal/Mol). The binding energy of allylsulfur compounds (Allicin, Alliin, s-Allylcysteine, s-Allylmercaptocysteine) and EGFR are shown in figure 4a-7a. The interaction between protein and ligand is visualized using AutoDock is shown in figure 4b-7b. The hydrogen bond interaction between protein and ligand using pymol is shown in figure 4c-7c. Distance of hydrogen bond between donor and acceptor is shows in Table 3-6. The table 7 explains the mode of binding interactions of the ligands on the active site of EGFR in relation to its crystal structure.

In conclusion Computational biology paves the way to speed up the process of drug discovery by reducing the cost and time. It involves variety of methods to identify the novel compounds, one such method is docking. It aims to identify the correct conformation of ligands in the binding pocket of a protein and to predict the affinity between the ligand and the protein. In the present *in silico* docking analysis, out of 4 allylsulfur compounds, alliin and s-allylmercaptocystein forms 2 hydrogen bonds, good interactions and least binding energy. This analysis reveals that the allylsulfur compounds of *Allium sativum* shows binding affinity with EGFR protein and have attempted to interpret inhibition mechanism of small molecules for the target protein involved in lung cancer. This analysis could be further extended for advanced drug designing protocols and for identifying novel targets in future studies.

References

Archana, P., Sathishkumar, N., Bharathi, N. 2010. *In Silico* docking analysis of Curcumin—An inhibitor for obesity, *Int. J. Pharma. Bio. Sci.,* 1(4).

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. 2000. *The Protein Data Bank Nucleic Acids Res.,* 28(1): 235-242.

Bolton, E.E., Wang, Y., Thiessen, P.A., Bryant, S.H. 2008. PubChem: integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.,* 4: 217–241.

Doll, R., Peto, R. 1976. Mortality in relation to smoking: 20 years' observations on male British doctors. *Br. Med. J.,* 2: 1525–1536.

Ferlay, J., Parkin, D.M., and Steliarova-Foucher, E.2010. "Estimates of cancer incidence and mortality in Europe in 2008," *European J. Cancer*, 46, 765–781.

Fleischauer, A.T., Poole, C., Arab, L.2000. Garlic consumption and cancer

prevention: Meta analyses of colorectal and stomach cancers. *Am. J. Clin. Nutr.* 72, 1047–1052.

Goodsell, D.S., Morris, G.M., Olson, A.J. 1996. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recogn.,* 9(1), 1–5.

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., Thun, M.J. 2007. Cancer statistics. *CA Cancer J. Clin.,* 57, 43-66.

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R.1997. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 267: 727–748.

Kartasasmita, R.E., Herowati, R., Harmastuti, N., Gusdinar, T. 2009. Quercetin derivatives docking based on study of flavonoids interaction to cyclooxygenase-2. *Indones. J. Chem.*, 9: 297-302.

Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev/Drug Discov.,* 3, 11: 935–949.

Knop, C.S. 2005. Lung cancer. In Yarbro, C.H., Frogge, M.H., & Goodman.,(Eds.), Cancer nursing: Principles and practice, 6th ed., 1379-1413. Boston, MA., Jones & Bartlett.

Lengauer, T., Rarey, M.1996. Computational methods for bimolecular docking. *Curr. Opin. Struct. Biol.,* 6: 402-406.

Luo, J., Solimini, N.L., and Elledge, S.J. 2009. "Principles of cancer therapy: Oncogene and non-oncogene addiction," *Cell,* 136, 823–837.

Mahady, G.B., Fong, H.H.S., Farnsworth, N.R. 2001. Botanical Dietary Supplements: Quality, Safety and Efficacy. Swets & Zeitlinger, Lisse, the Netherland.

Mitsudomi, T., Yatabe, Y. 2009. epidermal growth factor receptor in relation to tumor development: EGFR gene and cancer. *FEBS J.*, 277(2): 301–308.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.,* 19(14): 1639-1662.

Rarey, M., Kramer, B., Lengauer, T., Klebe, G. 1996. A fast flexible docking method using an incremental construction algorithm, *J. Mol. Biol.*, 261: 470–489.

Rarey, M., Kramer, B., Lengauer, T.1995. Time-efficient docking of flexible ligands into active sites of proteins. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 3: 300-308.

Srivastava, K.C., Bordia, A., Verma, S.K. 1995. Garlic (*Allium sativum*) for disease prevention. *South African J. Sci.,* 91, 68–77.

Uniprot C, "The Universal Protein Resource UniProt in 2010". *Nucleic Acids Res.,* 38 Database issue: D142–D148, 2009.

Van Cleave, J.H., & Cooley, M.E. 2004. Lung cancer. In C.G. Varricchio (Ed.), A cancer sourcebook for nurses ,8th ed., 215-229 ,Sudbury, MA., Jones & Bartlett.

How to cite this article:

Padmini, R., Razia Muthusamy, V. Roopavahini and Sai Aruna, S. 2016. Molecular Docking Studies of Allylsulfur Compounds from *Allium sativum* against EGFR Receptor. *Int.J.Curr.Res.Aca.Rev.* Special Issue-3: 207-218.