RESEARCH ARTICLE

INSILICO INTERACTION ANALYSIS OF HERBAL BIOACTIVE MOLECULES WITH BIOFILM ASSOCIATED GENE rfaDIN *ESCHERICHIA COLI*

*Florida Tilton, Akshaya Puthanpurayil and Lakshmi Priya, K.

Biozone Research Technologies Pvt. Ltd., Chennai, India

ABSTRACT

Biofilm is acollection of microbes with a distinct architecture. Biofilms are generally responsible for clogging and corroding pipes, reservoirs, storage area, etc. The quality of household drinking water is also affected by biofilm formation. Biofilms not only cause industrial disasters but are also responsible for causing medical conditions by growing on the surfaces of catheters, heart valve replacements, contact lenses, pacemakers, artificial joints and other surgical implants. Biofilms affect millions of people in the world each year and as a consequence, many deaths occur. Standard antibiotic therapy is often inadequate and the only option may be to remove the contaminated implant. This study was mainly focused on finding novel lead molecules for drug discovery against biofilm associated gene rfaD. The structure of the protein rfaD was modeled using MODELLER. 53 amino acids were chosen in the active site of the protein rfaD using CASTp. Computer aided screening was performed against 54 active compounds from 9 medicinal plants using GOLD. This study provides an organized approach to screen active compounds for the identification of lead molecules for combating biofilm formation in bacteria.

INTRODUCTION

Quorum sensing involves the regulation of gene expression as are action to vacillations in cell-population density (Melissa and Bonnie, 2001). Some bacteria send out quorum sensing signals and form biofilm. As the name suggests, biofilms are biological film formed by microorganisms by adhering to the neighboring microbes. Biofilm forming bacteria secretes polymers to form biofilm. A biofilm could be formed between either same species (homogeneous biofilm) or different species (heterogeneous biofilms) (Wimpenny *et al.*, 2000). Research has revealed that, the bacteria involved in biofilm production are complex and diverse. Study of physiological and structural nature of biofilms has led to the notion that they are coordinated and cooperative groups, similar to multicellular organisms (Nadell *et al.*, 2009).The bacteria growing in a biofilm are nearly 1,000 times more resistant to antibiotics when compared to the same bacteria not growing in a biofilm (Rasmussen and Givskov, 2006). Biofilms could be formed on external or internal surfaces. External biofilm i.e. outside the body, like chronic wounds and dental plaque, may be removed manually. But in case of internal biofilm, they are more difficult to eradicate due of their inaccessibility and heightened resistance to antibiotic combinations and dosages. Researchers have estimated that 65% of all microbial infections are caused by biofilms (internal biofilms) such

---

Corresponding author: Florida Tilton, Biozone Research Technologies Pvt. Ltd., Chennai, India.

as catheter infections (caused by Staphylococcus aureus) (Rodney, 2001), urinary tract infections (caused by E. coli and other pathogens) (Nicolle, 2005), child middle-ear infections (caused by Haemophilusinfluenzae) (Hall-Stoodley *et al*, 2006), common dental plaque formation, and gingivitis (Offenbacher *et al.,* 2007). External biofilms causes wide range of problems in industrial environments like biofilm developed on the interiors of pipes could lead to corrosion and clogging. In the food preparation area, biofilms on floors and counters can render the area unhygienic. They not only affect the quality of household drinking water, but also have immense adverse impact on a number of industries, including petroleum, specialty chemicals, mining and utilities.

Ever since biofilms have led to clogged watersheds, pipes, storage space, contaminated reservoirs and contaminated food products, large scale industries which are negatively impacted by their presence have taken immense interest in supporting biofilm research, particularly research that specifies how biofilms can be eliminated (Amy Proal, 2008). Common pathogens found in biofilm are *E. coli, Legionella pneumophila, Pseudomonas, Flavobacterium, Arthrobacter, Acinetobacter, Sarcina, Micrococcus, Porteus, Bacillus, Klebsiella and Enterobacter. Escherichia coli* have over 100 genes associated with biofilm formation, of which the generfaD that is involved in LPS biosynthesis plays a vital role. Upon disruption of the gene, there was significant reduction in biofilm production. Hence, this study focuses on finding potential novel lead compound against gene rfaD in *E.*

coli by performing molecular docking studies against 54 active compounds from 9 medicinal plants.

MATERIALS AND METHODS

Structure Prediction

The structure of the protein rfaD was modeled using MODELLER 9.12 (http://salilab.org/modeller/modeller.html). Template for modeling was obtained from RCSB Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) with PDB ID 2X6T.

Active site

The active sites of the protein rfaD were selected using CASTp (http://sts.bioengr.uic.edu/castp/). The active site residue locations were used for performing molecular docking.

Active compounds from Medicinal Plants

Natural compounds were searched from variety of literatures. The compounds from different medicinal plants belonged to flavonoids, flavons, glycosides, lactones, lignans, quinines, terpinoids and saponins. The resulting structures of 53 compounds from 10 different plants were drawn using ACD Chemsketch. They were geometry optimized and saved in mol file format. These were then used as ligands for performing docking studies.

Docking studies using GOLD

Automated docking was performed using the genetic algorithm GOLD (Version 3.2 CCDC, Cambridge, UK) (Jones *et al.,* 1997) with the selected 53 active compounds against receptor rfaD protein. The algorithm used in this study had been previously validated and successfully tested on a data set of over 300 complexes extracted from the PDB (Selvaraj *et al.,* 2008). Genetic algorithm (GA) used by the GOLD allows to explore the full range of lig and rotational and conformational flexibility of selected receptor hydrogen. Grid was prepared for the protein with the center and the size of the bounding box set on 10 Å. The coordinates of the enclosing box (x = 121 Å; y = 87 Å; z = 45 Å) were defined starting from the set of active site residues. During docking process, a maximum of 10 different conformations was considered for the drug. The conformer with highest binding score was used for further analysis (Nissink *et al.,* 2002).

RESULTS AND DISCUSSION

The 3D structure of the protein rfaD was predicted (Fig 1) and subjected to validation using PROCHEK server. From the Ramachandran plot (Fig 2), it was illustrated that 91.9% of residues were in most favored region, 8.1% in additionally allowed region and 0% residues in generously allowed region as well as in the disallowed region. 0% of amino acid in the disallowed region reveals that the predicted protein structure has stable conformation. CASTp server that was used to determine the active sites of the receptor rfaD indicated the

presence of 53 amino acids in the active sites (Fig 3). The 53 amino acid residues provide a cavity for the ligands to interact with the receptor rfaD. Using the GOLD fitness score generated by GOLD software, the inhibitory effect of the compounds was evaluated. The fitness scores were generated based on the binding compatibility i.e. Docked energy in kcal/mol (fitness) (Nissink *et al.,* 2002).
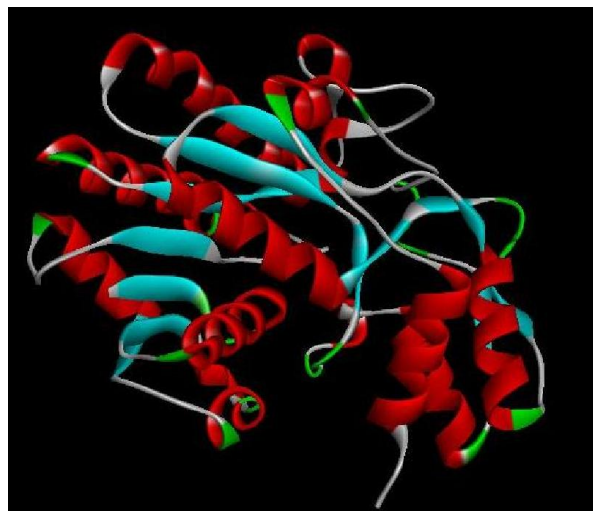

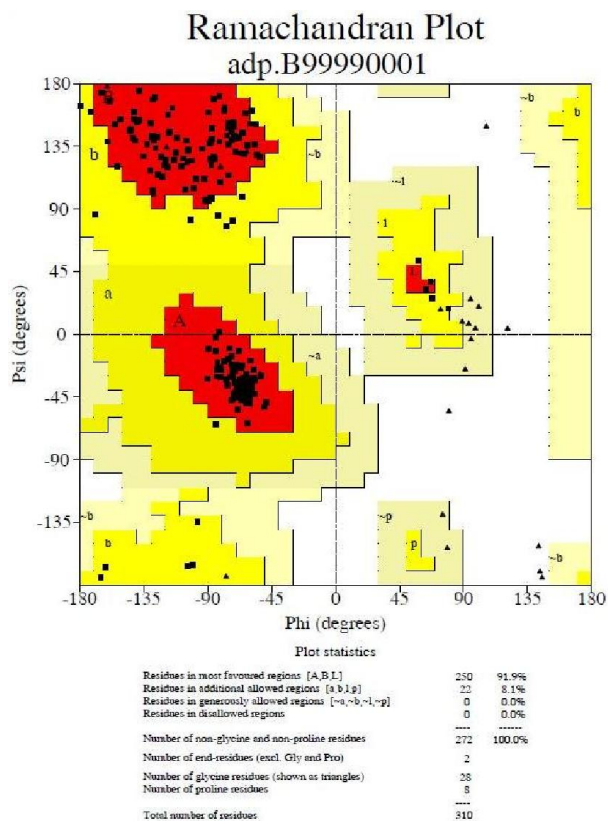**Fig. 1. Model generated by using MODELLER 9.12 for the rfaD protein**

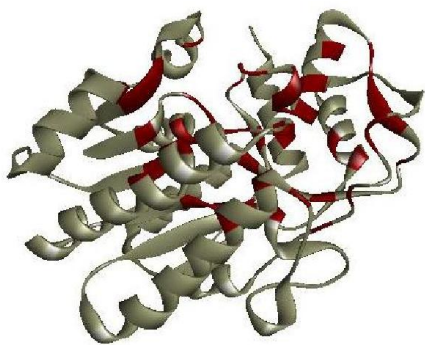**Fig. 2. Validation of the generated model using Ramachandran plot**

Fig. 3. Red color representations the active site of the protein rfaD

Table 1. List of medicinal plants and their active compounds

S.No.	Plant Name	Active Compound	GOLD Fitness Score
1.	*Anethumgraveolens*	Alpha-phellandrene	27.46
		Alpha-pinene	8.54
		Alpha-terpineol	24.46
		Anethole	34.62
		Apigenin	43.55
		Ascorbic-acid	39.10
2	*Azadirachtaindica*	Carpaine	2.53
		Caryophyllene	32.04
3	*Trigonellafoenum-graecum*	Chlorogenic-acid	55.33
		Chrysophanol	38.56
		Cinnamic-acid	28.94
		Ellagic-acid	40.97
4	*Terminaliachebula*	Emodin	41.31
		Esculetin	34.43
5	*Cassia alata*	Ferulic-acid	33.91
6	*Phyllanthusemblica*	Flavone	38.87
		Gallic-acid	34.68
		Gentianine	31.38
		Geraniol	25.43
		Guaiacol	31.01
		Hyperoside	42.58
		Isorhamnetin	49.54
7	*Cuminumcyminum*	Jasmone	39.28
		Kaempferol	44.36
		Lauric-acid	48.74
		Limonene	28.95
		Luteolin	48.48
		Menthol	32.26
		Menthone	34.65
		Myrcene	31.92
8	*Mint*	O-coumaric-acid	33.53
		Oleanolic-acid	-114.45
		P-coumaric-acid	32.14
		P-cymene	33.54
		Perillyl-alcohol	35.37
		Perillaldehyde	33.62
		Pectin	32.25
		Phenethyl-alcohol	33.69
		Protocatechuic-acid	33.26
		Pulegone	30.34
		Quercetin	46.79
		Quercitrin	46.43
		Rhein	40.43
		Rosmarinic-acid	48.13
		Rutin	33.69
		Sabinene	20.07
		Safrole	41.05
9	*Curcuma longa*	Scopoletin	36.17
		Terpinen-4-ol	21.07
		Terpineol	31.17
		Thymol	32.58
		Umbelliferone	35.07
		Ursolic-acid	-55.34
		Vanillic-acid	34.00

The active compound Chlorogeneic acid from *Anethumgraveolens* binds with the protein rfaD with the highest GOLD Score of 53.33 (Table 1). Relatively the active compounds Isorhamnetin, Luteolin, Rosmarinic acid, Quercetin, Quercitrin, Kaempferol, Apigenin and Hyperosidebinds with the receptor with the GOLD scores of 49.54, 48.48, 48.13, 46.79, 46.43, 44.36, 43.55 and 42.58 respectively. Further compounds like Emodin, Ellagic acid and Rheinalso shows significant binding affinity with GOLD scores of 41.31, 40.97 and 40.43 respectively. The analysis of the H-bond formations between the receptor and the active compounds revealed that Quercetin and Rosmarinic acid formed nine and eight H-bonds (Table 2).
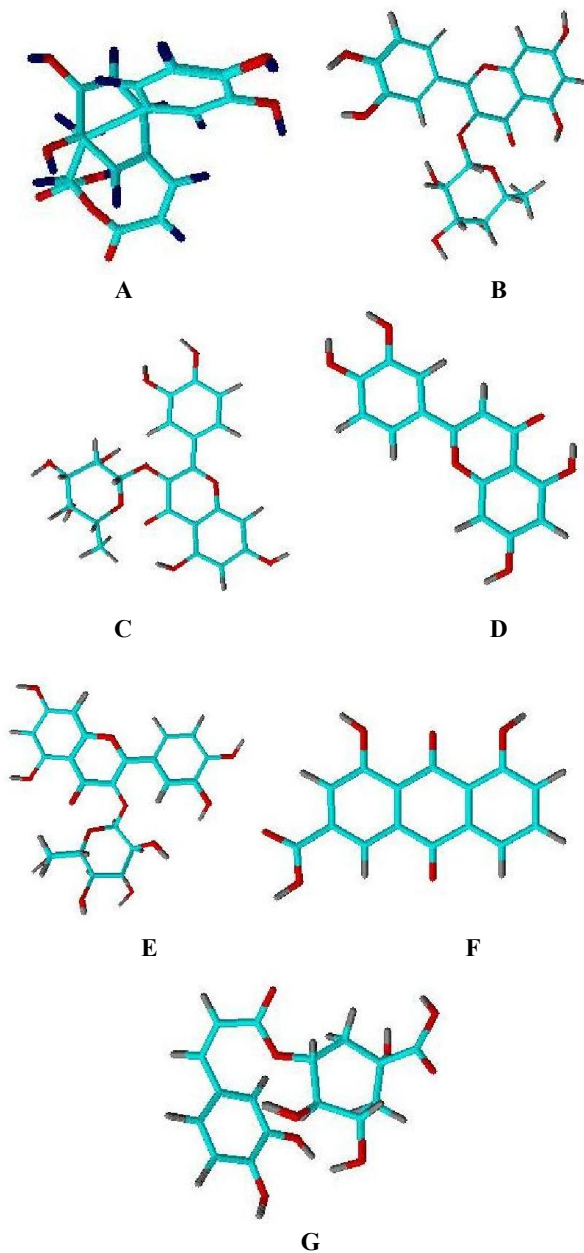
(A)Rosmarinic acid, (B)Quercetin, (C) Quercitin,(D)Luteolin, (E)Hyperoside, (F)Rhein and (G)Chlorogenic acid

Fig. 4. Active compounds from medicinal plants

Table 2. List of top 21 active compounds with their GOLD fitness scores

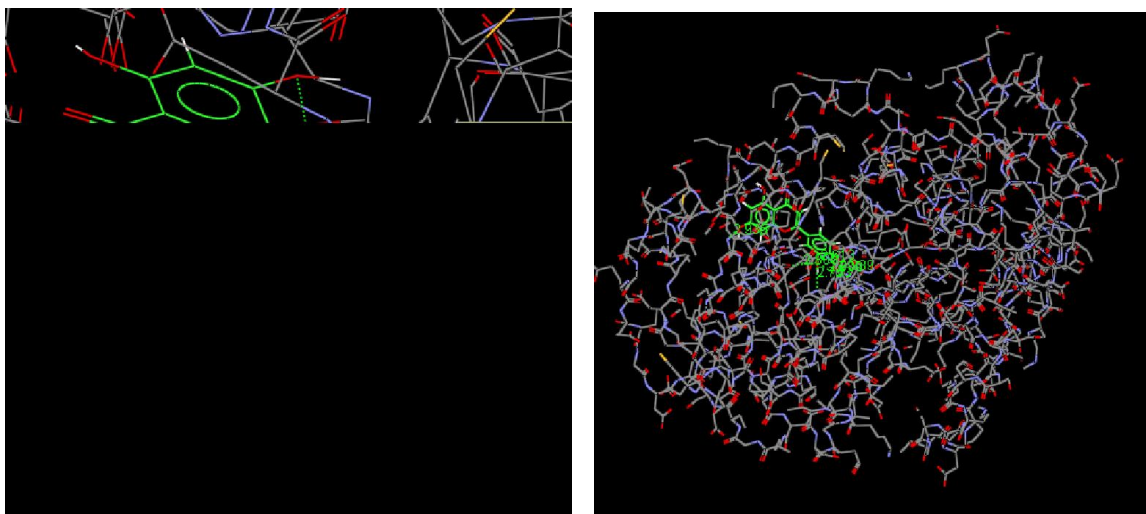| S.No.	Ligand Name	Atom in Protein	Atom in ligand	Distance of H-bond	Fitness Score
1	Chlorogenic acid	O10	GLU175:O	2.614	53.33
		O9	GLY6:O	2.763	
			GLY12:N	2.965	
		O23	GLY76:O	2.961	
		O24	HIS177:NE2	2.555	
		O25	HIS177:NE2	2.866	
2	Isorhamnetin	O22	ALA77:O	2.358	49.54
		O21	ASN92:OD1	2.537	
		O23	HIS177:NE2	2.982	
			PHE10:N	2.852	
3	Lauric acid	O14	PHE10:N	2.767	48.74
4	Luteolin	O18	ASP31:N	2.935	48.48
		O20	ILE11:N	2.809	
			GLY12:N	2.753	
			GLY6:O	2.809	
		O21	ILE11:N	2.939	
			PHE10:N	2.613	
5	Rosmarinic acid	O26	GLY6:O	2.959	48.13
			PHE10:N	2.943	
			ILE11:N	2.764	
			GLY12:N	2.679	
		O25	PHE10:N	2.569	
		O11	SER115:O	2.204	
		O24	ALA117:N	2.749	
			ALA118:N	2.761	
		O23	GLN273:OE1	3.01	
6	Quercetin	O19	ALA77:N	2.752	46.79
		O20	GLY6:N	2.45	
		O32	PHE10:N	3.058	
			ILE11:N	2.556	
		O31	GLY12:N	2.983	
			PHE10:N	2.78	
			GLY6:O	2.534	
		O28	SER79:OG	2.965	
		O29	SER79:OG	2.803	
7	Quercitin	O20	GLY6:N	3.019	46.43
			ASP31:OD1	2.853	
		O31	GLY12:N	3.064	
			ILE11:N	2.427	
			PHE10:N	2.691	
		O30	PHE10:N	2.819	
8	Kaempferol	O18	ILE11:N	2.89	44.36
		O19	HIS177:NE2	2.683	
		O21	TYR96:OH	2.742	
		O20	ASP31:OD1	2.732	
9	Apigenin	O19	HIS177:NE2	2.761	43.55
		O18	ILE11:N	2.932	
10	Hyperoside	O19	GLU175:OE1	2.9	42.58
		O28	PHE10:N	2.851	
			HIS177:NE2	2.962	
		O33	SER115:O	2.301	
		O30	GLY12:N	2.939	
			GLY6:O	2.801	
11	Ascorbic acid	O10	PHE10:N	3.055	39.10
		O12	ILE11:N	2.985	
		O11	PHE10:N	2.878	
			GLY12:N	2.751	
			GLY6:O		
12	Emodin	O18	HIS177:NE2	2.545	41.31
		O19	ILE11:N	2.767	
13	Safrole				41.05
14	Ellagic acid	O20	HIS177:NE2	2.456	40.97
		O7	ILE11:N	2.45	
15	Rhein	O21	SER79:OG	2.381	40.43
		O19	PHE10:N	2.461	
			ILE11:N	2.621	
		O20	GLY6:O	2.402	
			GLY21:N	3.056	
		O17	ALA77:N	2.69	
16	Jasmone	O11	ALA77:N	2.989	39.1
17	Flavone	O17	VAL170:N	2.689	38.87
18	Chrysophanol	O18	GLY6:N	2.344	38.56
19	Scopoletin	O14	HIS74:ND1	2.905	36.17
20	Perillylalcohol	O11	HIS74:NE2	2.637	35.37
21	Umbelliferone	O10	ASP148:OD1	2.885	35.07
		O12	GLY76:O	2.844	
			HIS74:O	2.333	

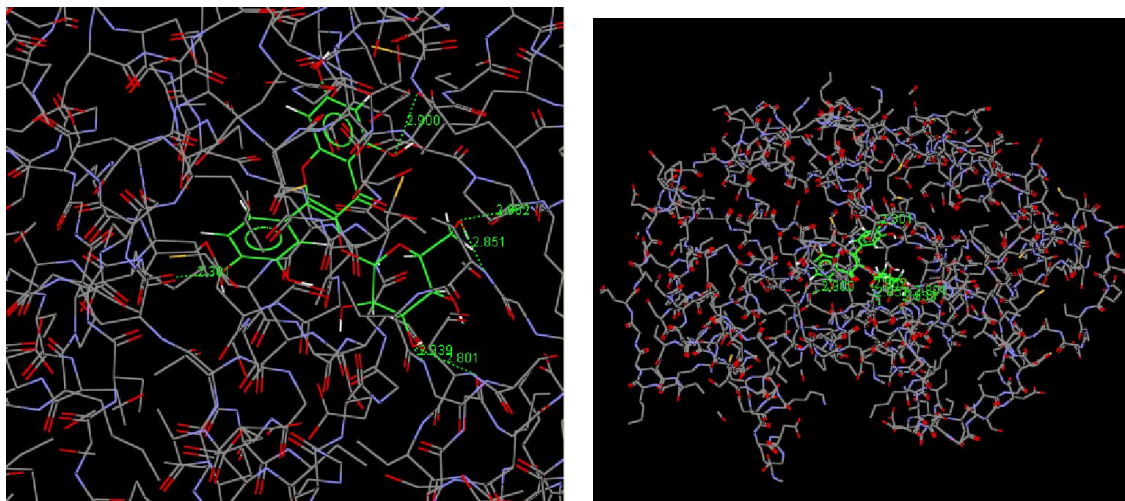**Fig. 5. Interaction between protein rfaD and Luteolin**

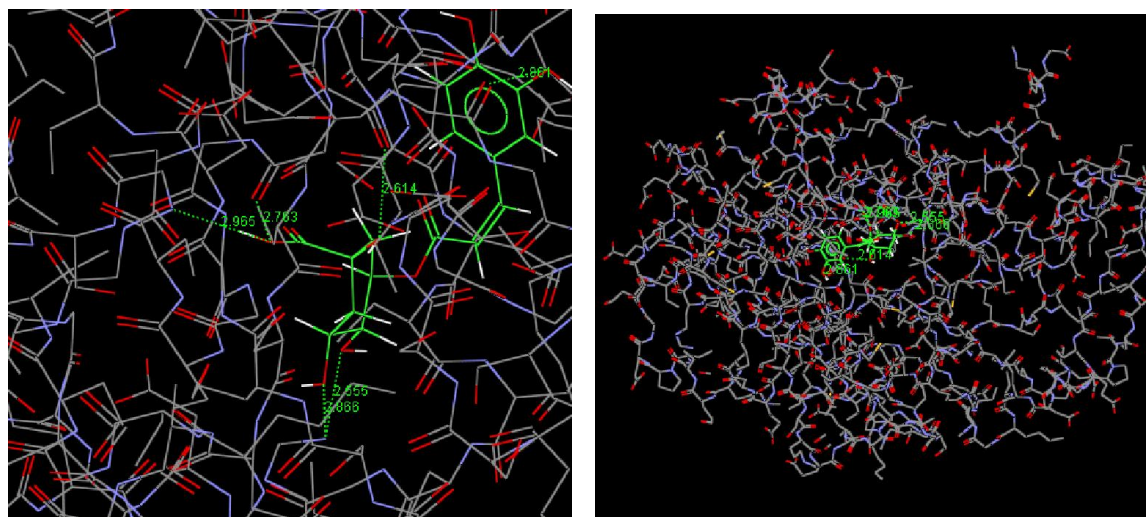**Fig. 6. Interaction between protein rfaD and Hyperoside**

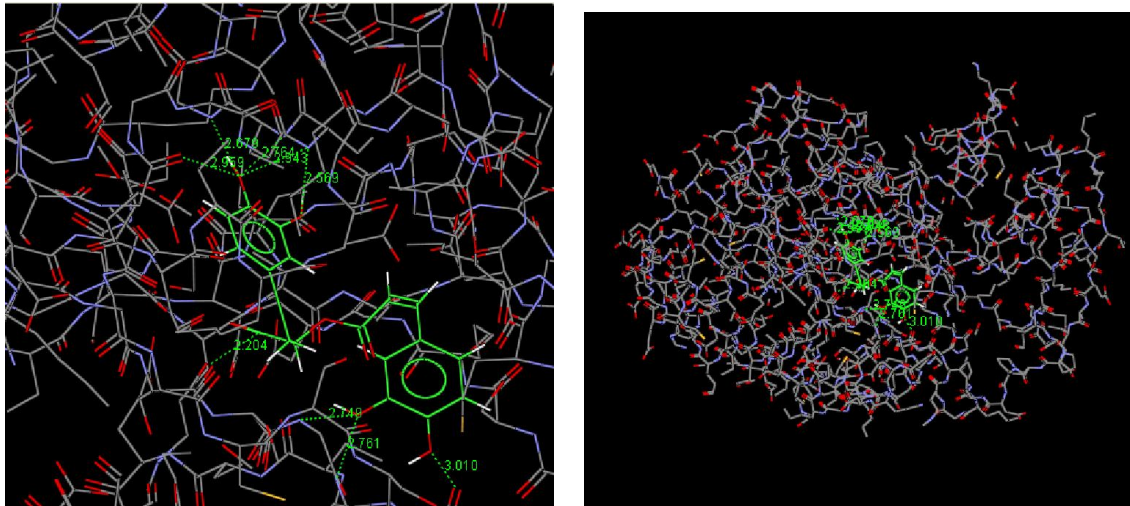**Fig. 7. Interaction between protein rfaD and Chlorogenic acid**

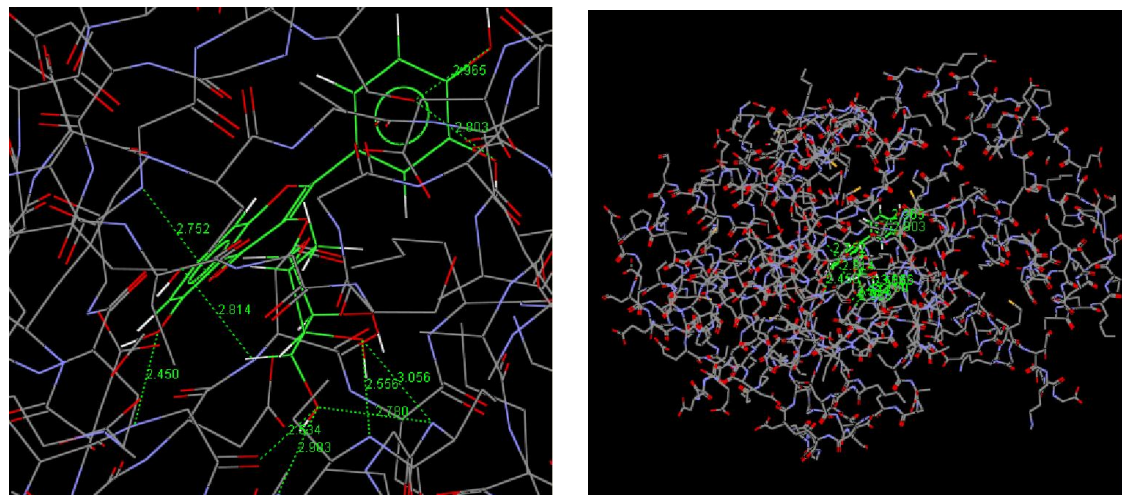**Fig. 8. Interaction between protein rfaD and Rosmarinic acid**

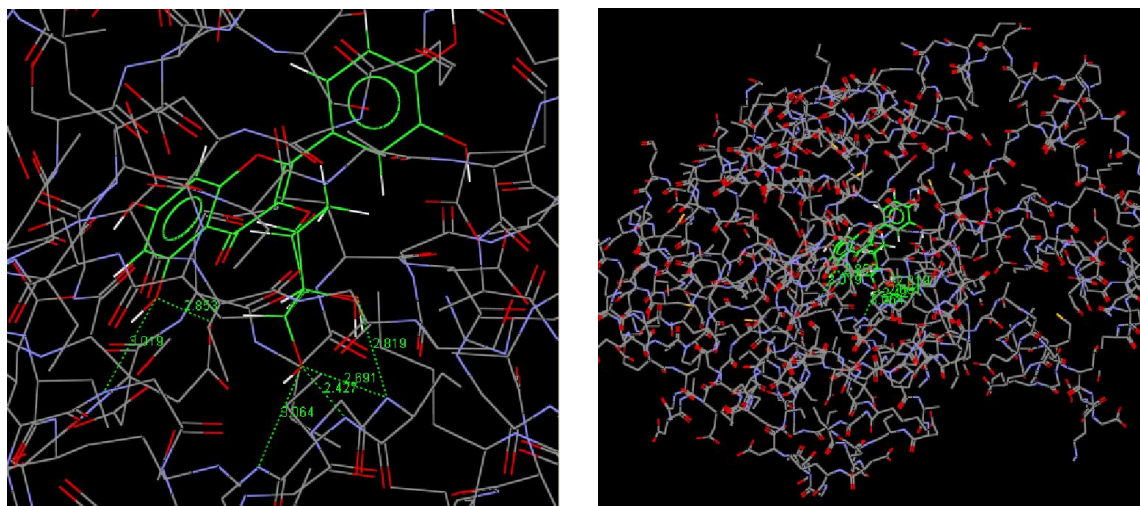**Fig. 9. Interaction between protein rfaD and Quercetin**

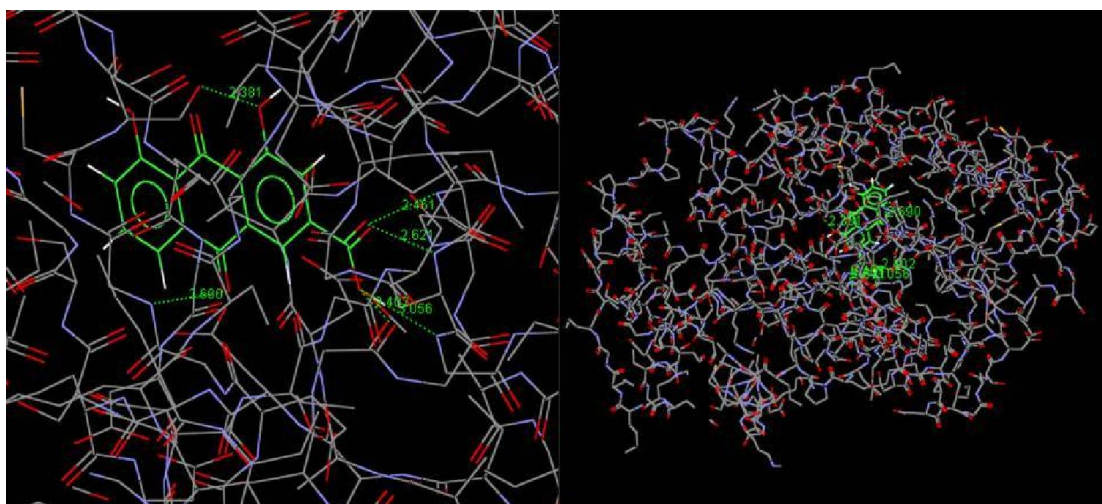**Fig. 10. Interaction between protein rfaD and Quercitin**

Fig. 11. Interaction between protein rfaD and Rhein

Further, active compounds Chlorogenic acid, Luteolin, Quercitrin, Rhein and hyperoxide formed six H-bonds with the receptor rfaD. Although Lauric acid from *Anethumgraveolens* has agood GOLD score of 48.74, it formed the least number of H-bonds i.e. one H-bond with the receptor. Figures 4 - 11shows interaction of the active compounds with protein rfaD.

From the experiments it is evident that Chlorogenic aciddemonstrates a better anti-biofilm potential when paralleled to the other active plant compounds. Similarly, the four active compounds Luteolin, Quercitrin, Rhein and hyperoxide could also be novel drug against biofilm forming *Eschericia coli*. Although Ascorbic acid has comparatively lower GOLD score, it could also be a potentlig and against *E. coli* involved in biofilm production since it formed five H-bonds with the receptor rfaD.

Conclusion

Molecular docking methods have become more popular and are being used widely. This method is economical and much faster in the process of discovering novel molecules than the traditional trial and error method. From the current study, molecular docking has proved to be an efficient method for the identification of novel lead compounds against biofilm producing *Escherichia coli* from a broad spectrum of plant compounds. Chlorogenic acid showed high binding affinity with the protein rfaD. Hence, chlorogenic acid may be a potent lead molecule against biofilm formation. Conclusively, this approach of using computer aided screening of library of active compounds can be useful for industrial sectors to minimize the complexities of identification and isolation of novel ligands. Also the study brings a good insight to structure based drug designing by demonstrating the different interactions of the receptor and the active compounds.

REFERENCES

Amy Proal. 2008. Understanding Biofilms. (http://bacteriality. com/2008/05/26/biofilm/). Last accessed on: 26.04.2014 at 6:30 pm IST.

Hall Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J and Forbes M. 2006. Direct detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. *The Journal of the American Medical Association.* 296(2):202-11.

Jones G, Willett P, Glen RC, Leach AR, Taylor R. 1997. Development and validation of a genetic algorithm for flexible docking. *J MolBiol* 267: 727-748.

Melissa B. Miller and Bonnie L. Bassler. 2001. Quorum Sensing In Bacteria. *Annual Review of Microbiology.* Vol. 55: 165-199.

Nadell CD, Xavier JB and Foster KR. 2009. The sociobiology of biofilms. *FEMS Microbiology Reviews.* 33:206-24.

Nicolle LE. 2005. Catheter-related urinary tract infection. Drugs Aging. 22(8):627-39.

Nissink JW, Murray C, Hartshorn M, Verdonk ML, Cole JC *et al.* 2002. A new test set for validating predictions of protein-ligand interaction. Proteins. 4: 457-471.

Offenbacher S, Barros SP, Singer RE, Moss K, Williams RC and Beck JD 2007. Periodontal disease at the biofilm-gingival interface. *J Periodontol.* 78(10):1911-25.

Rasmussen TB and Givskov M. 2006. Quorum-sensing inhibitors as anti-pathogenic drugs. *Int J Med Microbiol.* 296(2-3):149-61.

Rodney M. Donlan. 2001. Biofilms and Device-Associated Infections. *Emerging Infectious Diseases.* Vol. 7, No. 2: 277-281

Selvaraj M and Malik BK. 2008. Modeling of human CCR5 as target for HIV-I and virtual screening with therapeutic compounds. *Bioinformation,* 3(2): 89-94.

Wimpenny J, Manz W and Szewzyk U. 2000. Heterogeneity in biofilms. *FEMS Microbiology Review.* 24(5):661-71.

****