Characterization of the structure of Anopheles gambiae Odorant Receptor 1 (AgOR1) and its binding affinity with novel inhibitors and phytochemicals: A Molecular Dynamics and Docking Approach

Akash Deep Biswas
Molecular Modelling and Simulation Lab, Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, Assam-784028

Sujata Dev
Molecular Modelling and Simulation Lab, Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, Assam-784028

Mattaparthi Venkata Satish Kumar
Molecular Modelling and Simulation Lab, Department of Molecular Biology and Biotechnology, Tezpur University, Tezpur, Assam-784028

Abstract- Olfaction has a major impact on behavioural entomology. The malaria vector mosquito *Anopheles gambiae* sensu stricto identifies the host by using its olfactory system. The odorant receptors (OR) are known for its critical role on the olfactory signaling cascade. Hence many studies are now focused on designing inhibitors to block the functional aspect of the OR or to altogether inactivate the OR. There are nearly 80 *Anopheles gambiae* Odorant Receptor (AgOR) genes that have been identified in *Anopheles gambiae*. But the structural details of these receptors are less studied. In this work, we have modeled the structure of AgOR1 protein by submitting its amino acid sequence in the online I-TASSER server. The salient structural characterization of the model structure of AgOR1 was then studied using Molecular Dynamics Simulation. The equilibrated structure of AgOR1 was then docked with the 14 odorant molecules (five commercially used repellents, one newly developed antagonist and rest eight molecules are phytochemicals derived from *Ocimum tenuiflorum)*. MD Simulation was again done on the docked structures and Binding free energy calculation was done using MM/PBSA-GBSA complex structures. From the study, we observed that VUAA1, Ursolic acid and Rosmarinic acid from *Ocimum tenuiflorum* to show strong affinity with the AgOR1 along with the established repellent D-trans allethrin. In AgOR1, Asn79, Lys69, Ala78, Lys72 and Glu122 found to be the interacting residues with the inhibitors. In these precise steps we characterized the salient structural and docking features of AgOR1. From the present study, we propose that the phytochemicals from *Ocimum tenuiflorum* may play an important role as mosquito repellent and should be experimentally studied. Our findings in this study may aid in designing new and more effective repellents to control mosquito problem.

Keywords- *Anopheles gambiae,* Odorant Receptor, Ursolic acid, Rosmarinic acid, D-trans allethrin, I-TASSER, VUAA1, Molecular Docking, Molecular Dynamics Simulation, MM-PBSA/GBSA, *Ocimum tenuiflorum,* phytochemicals

List of Abbreviations-

MD: Molecular Dynamics

RMSD: Root Mean Square Deviation

AgOR: *Anopheles gambiae* Odorant Receptor

VUAA1: Vanderbilt University Allosteric Agonist 1

MM: Molecular Mechanics

PBSA: Poisson Boltzmann Surface Area

GBSA: Generalized Born Surface Area

I. INTRODUCTION

In biological world every organism has its own niche and ecological importance. But whereas in one hand while we can't dispose them off, we can't ignore the fact that many species are antagonistic to each other not only directly but also by providing a vehicle to carry other harmful species. This lead to the most alarming of diseases affecting a large part of the human population all over the world.

In case of mosquito-borne diseases, the World Health Organization (WHO) has submitted a report which sprouts harrowing facts. Around 3 million people are affected by some or the other mosquito-borne diseases. Amongst this major chunk of population, the cream of the icing goes to malaria which claims about 3.2 billion peoples' health that is almost half of the world's population are at risk of malaria. In the recent years, between 2000 and 2015, the frequency and incidence of malaria cases fell globally by 37%. Along with that, the death rates caused by malaria also fell by 60% among all age groups and the most commendable observation was that the rate went further down to 65% among children under the age of 5. The Sub-Saharan Africa carries a disproportionately high share of the global malaria burden. In 2015, the region was home to 89% of malaria cases and 91% of malaria deaths. So, the need to the hour is to develop new repellents.

The mosquito *Anopheles gambiae* (*A.gambiae*) is a well-known vector for *Plasmodium falciparum* (*P.falciparum*), a primary causative agent for malaria. It is spread to people through the bites of infected female Anopheles mosquitoes. These female mosquitoes are specially known as "malaria vectors". There are 5 known parasite species that cause malaria in humans and amongst these two species are the biggest threat. These two species are none other than *P.falciparum* and *P*lasmodium vivax (*P.vivax*). *P.falciparum* mostly prevalent on the African continent. It is the main culprit behind most of the malaria-related deaths all over the world. *P.vivax* is more spread-out than *P.falciparum* and is well-reported in countries other than Africa.

A.gambiae uses its sense of smell for its anthropophilic blood-meal host direction cues. The odorant cues (typically present in the sweat of humans) interact with the many specialised kinds of odorant receptor. Let's cite the case of *A.gambiae*. The gene *A*nopheles gambiae Odorant Receptor 1 (AgOR1) is expressed in the olfactory tissue of only female mosquitoes and respond strongly to 4-methylphenol present in human sweat which kick-starts an electrophysiological response in the antenna of female mosquito by binding to the odorant receptors. This response is downregulated after a blood-meal. [1] The odorant organs are covered by fine hair called sensilla and they have different sub-units. The olfactory sensilla are compartmentalised units, which are specifically honed for accepting and reacting to only one odorant molecule or maybe subset of odorants. Each sensillum encases two or three olfactory neurons. These neurons in turn each represent a single odorant receptor. Accessory factors and also proteins like Odorants binding proteins (OBPs) and odorant degrading enzymes are secreted into the bathing lymph fluid by supporting cells found in the base of each sensilla. A special note to OBPs should be made at this point. These specialised OBPs proteins solubilize the hydrophobic odorants in the lymph, protect them from degradation and help to transport them to the olfactory receptors. [2] The binding of odorants with the receptor sub-units activates the signalling cascade and prompts the signal to reach the brain. The commercial repellents are designed to mimic these natural odorants and bind to the receptor competitively.

The researchers have started to exploit the mosquito's inherent sense of smell and method blood-meal source detection against them. They have tried and tested a host of synthetic compounds like N, N-Diethyl-meta-toluamide (DEET), Vanderbilt University Allosteric Agonist 1 (VUAA1), Dimethyl phthalate, Dimethyl carbate, Ethyl anthranilate, Inadole, Permetherin, and Picardin etc. Out of all these DEET was found to be the most effective and established insect repellent in the market from the last 60 decades, in the meantime many repellents came to the market such as D-Trans Allethrin, Prallethrin, Rebemide, Transfluthrin, etc. till the advent of VUAA1, which is thousand times more potent than DEET. The main problem with VUAA1 is that it is not volatile, so it has not been commercialized yet, but scientists are still trying to shoot this problem and find a solution for it. DEET's high participation rate to bio magnification and toxic build-up in the ecosystem, has motivated scientists worldwide to come up with more eco-friendly alternatives. Hence the search for phytochemicals which could effectively block the mosquito's sense of smell by binding to the odorant receptors started with neem, catnip, citronella, eucalyptus etc. Most phytochemicals that has been studied so far are primarily from the Indian sub-continent. But these works need to be furthered to establish the phytochemicals into the mainstream commerce.[3]

So, to provide an impetus to the search of phytochemicals we have chosen *Ocimum tenuiflorum* (*O.tenuiflorum*) also known as the holy basil or tulasi. It is an Indian household plant and is known to be of great medicinal value. In the current study, various compounds from the *O.tenuiflorum* extract namely Oleanolic Acid, Ursolic Acid, Rosmarinic Acid, Eugenol, Carvacrol, Linalool, B-Caryophellene, B-Elemene along with the 5 established repellents namely DEET, D-Trans Allethrin, Prallethrin, Rebemide, Transfluthrin and 1 newly discovered repellent VUAA1 are docked into the olfactory receptor AgOR1 of *A.gambiae* and the results of the

phytochemicals has been compared to that of the repellents using Molecular Dynamics (MD) Simulation study. The main aim of this study is to establish a relation between the repellents and the phytochemicals and to check whether the listed phytochemicals can act as a good repellent.

In this work reported here, an in depth look has been made at the conformation changes when the AgOR1 underwent MD Simulation with and without ligands. We have done a total of 72 ns of MD Simulation on AgOR1, and AgOR1 with best docked ligands. The simulation presented in this article shows that different ligands docked with AgOR1 show conformational changes as well as shows stability in the AgOR1 + Ligands systems as compared to the AgOR1 system. To support the docking results and to cross check the stability of the complexes (AgOR1 + Ligands) we have done the Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) and Generalized Born Surface Area (GBSA) study. [4] The results suggests that the newly discovered repellent VUAA1 is having the best binding free energy among all the ligands as expected, but along with it interestingly some of the phytochemicals showed better results as compared to the 5 established repellents. From which it can be inferred that these phytochemicals or their derivatives can be experimentally studied for using them as repellents.

II. MATERIALS AND METHODS

A. Data Retrieval

The receptor of AgOR1 was retrieved from UniProt with ID: Q8WTE7. The FASTA sequence was retrieved and used for Modelling. The 3D Structures of 14 Compounds used for this study was retrieved from PubChem Database.[5] The Compounds are (1) DEET (Cid_4284) (2) D-Trans Allethrin (Cid_15558638) (3) Prallethrin (Cid_71306899) (4) Rebemide (Cid_15542) (5) Transfluthrin (Cid_656612) (6) VUAA1 (CID_1319135), the phytochemical compounds of *O.tenuiflorum* are (7) A-Elemene (Cid_80048) (8) Carvacrol (Cid_10364) (9) Caryophellene (Cid_5281515) (10) Eugenol (Cid_3314) (11) Linalool (Cid_6549) (12) Oleanolic Acid (Cid_10494) (13) Rosmarinic Acid (Cid_5315615) (14) Ursolic Acid (Cid_64945).

B. Sequence based Screening

The FASTA sequence of AgOR1 was used for primary sequence using BLAST against pdb database and later with non-redundant database. The similar sequences were retrieved and Multiple Sequence Alignment was done using Clustal Omega. The Aligned files were later analysed using ESPript 3.0. [6]

C. Structure Modelling and Structure Validation

The FASTA sequence of AgOR1 retrieved from UniProt was submitted in I-TASSER server. [7, 8] I-TASSER used 10 PDB ID's 1JB0, 1ST6, 1TR2, 2J69, 2O01, 2O8E, 2YEV, 3WDO, 4IGG, 4P7H as templates for threading and resulted with 5 models. Based on the best C-Score of -1.69 and Cluster Density of 0.1923 we screened the Model 2 for structure validation. The structure thus considered was further validated using PDBsum and the Ramachandran Map was plotted for it before the MD Simulation and after the MD Simulation. The percentage of structural transitions is calculated using YASARA software package. [9]

D. Molecular Dynamics Simulation

MD simulation was carried out on the 3D modelled AgOR1 protein structure. The initial coordinate and the topology files were achieved using the LEaP module of AMBER packages. [10] This system was then subjected to MD Simulation using sander module of AMBER packages [11] using ff99SB force field. [12] The whole system was then solvated in a cubic water box of 10Å using TIP3P water model [13] and neutralized using four Na+ ions. Using sander module the system was then relaxed by 2000 steps of energy minimization following steepest descent algorithm with Harmonic constraints with a force constant of 30 kcal/mol/Å^2 followed by 2000 steps of conjugate gradient minimization by which several overlapped water molecules and bad contacts on the proteins are removed. To carry out the MD Simulation and slowly relax the system, the temperature of the system was gradually increased from 0-300K using 30 kcal /mol/Å^2 of harmonic constraints on the solute to its starting structure. The temperature and the pressure were maintained at 300K and 1 bar by using modified Berendsen thermostat velocity rescaling and Parrinello-Rahman pressure coupling respectively.[14] The system were then subjected to equilibration under a NVT (canonical or isothermal-isochoric) ensemble followed by NPT (isothermal-isobaric) ensemble with a reference temperature of 300K and reference pressure of 1 bar. The non-bonded interactions such as long-range interactions were taken care by computing it with means of Particle Mesh Ewald (PME) method, whereas for short-range van der Waals interactions a cutoff of 1.0nm was maintained while computing. The SHAKE algorithm [15] was subjected to constrain the bond lengths, which includes covalent bonds and hydrogen bonds, along with it the equations of motion was integrated with a time-scale of 2fs. Coordinates for NVT, NPT and Productions were saved for

every 1ps, 1ps and 1ps respectively. The MD analyses were performed using ptraj module and locally developed shell scripts, whereas the graphs were prepared using the program Xmgrace.

E. Molecular Docking

To elucidate the study of AgOR1 protein and the drug delivery onto the active site of the receptor protein, we carried out Molecular Dynamics Simulation and Molecular Docking studies. The aim of our study is to see the dynamic changes in the AgOR1, and make it stable to determine few parameters viz.: (1) Binding affinity of AgOR1 (2) Binding sites (3) Structural changes during the whole process (4) Residues involved in ligand – receptor sites (5) Determining the binding pockets for the molecules. To investigate the plausible binding mode we have considered certain docking scoring functions. For docking we considered the equilibrated structure obtained from MD Simulation study and converted it to Protein data Bank (PDB) format. We performed the docking studies using Molegro Virtual Docker (MVD).[16] A cavity prediction algorithm and the differential evolution optimization technique combines to form an algorithm called glided differential evolution which is again called as MolDock, a new hybrid algorithm. Taking molecular surface as expanded *van* der Waals and max number of cavity volume: 5, maximum cavity volume: 10000, probe size: 1.20, maximum number of ray checks: 16, minimum number of ray hits: 12 and using grid resolution of 0.80 we got 5 cavities, the volumes as mentioned earlier (Fig. 3A). Docking was done taking the simulated AgOR1 protein and taking the 3D structures of the 14 ligands mentioned in the materials and methods with the 5 cavities.

We used all the 5 active sites of AgOR1 and docked with compounds with a grid resolution of 0.30, where maximum population size is 50 and the pose of energy threshold generation is 100. After all the above mentioned steps were taken care of, we used simplest evolutionary step 300 and maximum distance factors to be 1.00. In the next step the maximum number of poses returned is taken to be 10. For each cavity we set up side chain flexibility and a Tolerance of 1.10 and Strength of 1.00. After the docking, the script is generated; the log file and the poses were saved as Mol2 file format. The candidate's poses were minimized and incurrent energy used a grid-base method for evaluating protein ligand interaction energy was saved in Mol2. The H-bonding and electrostatic terms are identified by Piecewise Linear Potential (PLP), the extension of PLP is Docking Score function.[17] And further the docking accuracy is determined by the re-rank score. The best docked conformation can be selected on the basis of MolDock and re-rank score values, but we considered MolDock Score and Docking Score as because we were much interested to show the H-bond interaction between the ligand and the receptor. The visualization studies were done using Molegro Virtual Viewer [16], UCSF Chimera [18] and VMD software.[19] We performed docking simulation of 14 molecules (Table S1) with the five detected cavities of AgOR1 protein (Fig. 3A) respectively. Finally, the scores of Glided differential evolution or MolDock Score function and the Docking Score function based on Tripos force-field were studied and the best one was considered for searching the binding orientation and conformation of each candidate molecule. The selected poses interacting with AgOR1 was further analysed using LigPlot+. [20]

F. MM-GBSA/PBSA Calculations of AgOR1-Repellent Complexes

To validate the docking results we carried out the relative binding free energy of AgOR1 and 14 repellents complexes we used the MM-GBSA/PBSA method. We extracted the snapshots of the complex system (without water and ions) for MM-PBSA calculation from the system which has undergone the Molecular dynamics production run. In the next step we performed GBSA/PBSA analysis on the 14 different components of each of the complex systems: the AgOR1, 14 ligands and the complexes respectively. For each of these components, the individual interaction energy and solvation free energy were calculated and the average was considered to estimate the binding free energy. The binding free energy of the AgOR1 with Repellents was calculated by the difference

$\Delta G_{bind} = G_{tot}$ (Complex) – G_{tot} (Protein) –G_{tot} (Ligand)

Where,

Complex = Protein + Ligand

Protein = AgOR1

Ligand = VUAA1, DEET, D-Trans Allethrin, Prallethrin, Rebemide, Transfluthrin, A-Elemene, Carvacrol, Caryophellene, Eugenol, Linalool, Oleanolic Acid, Rosmarinic Acid, Ursolic Acid.

In MM-GBSA calculations, we have used the pairwise GB model [21] with parameters described by Tsui and Case.[22] For all other settings we have used the default values. While in the MM-PBSA calculations, we set the internal dielectric and external one as 1.0 and 80.0 respectively.

III. RESULTS AND DISCUSSIONS

A. Anopheles gambiae Odorant Receptor 1

A.1 Screening of AgOR1 performing BlastP and predicting the conserved regions

The Fasta sequence of AgOR1 retrieved from UniProt database with the accession number Q8WTE7 was used for performing protein sequences blast for searching similar sequences having similar functional properties. The sequence's accession numbers that shows close proximity are KFB49096, AAR01130, XP_001842656 and XP_001842654 (Fig. 1). The result suggested that it has close proximity with the odorant receptor of *Anopheles sinensis* and *Culex quinquefasciatus*, which are also the vectors for spreading malaria. So, the findings in these work may also apply for both the vectors. To check the proximity we further studied the Multiple Sequence Alignment (MSA) by using clustal omega and found that there are 49 conserved residues across the five sequences regions are shown in red and shares maximum of the residues having same properties shown in yellow (Fig. 1).

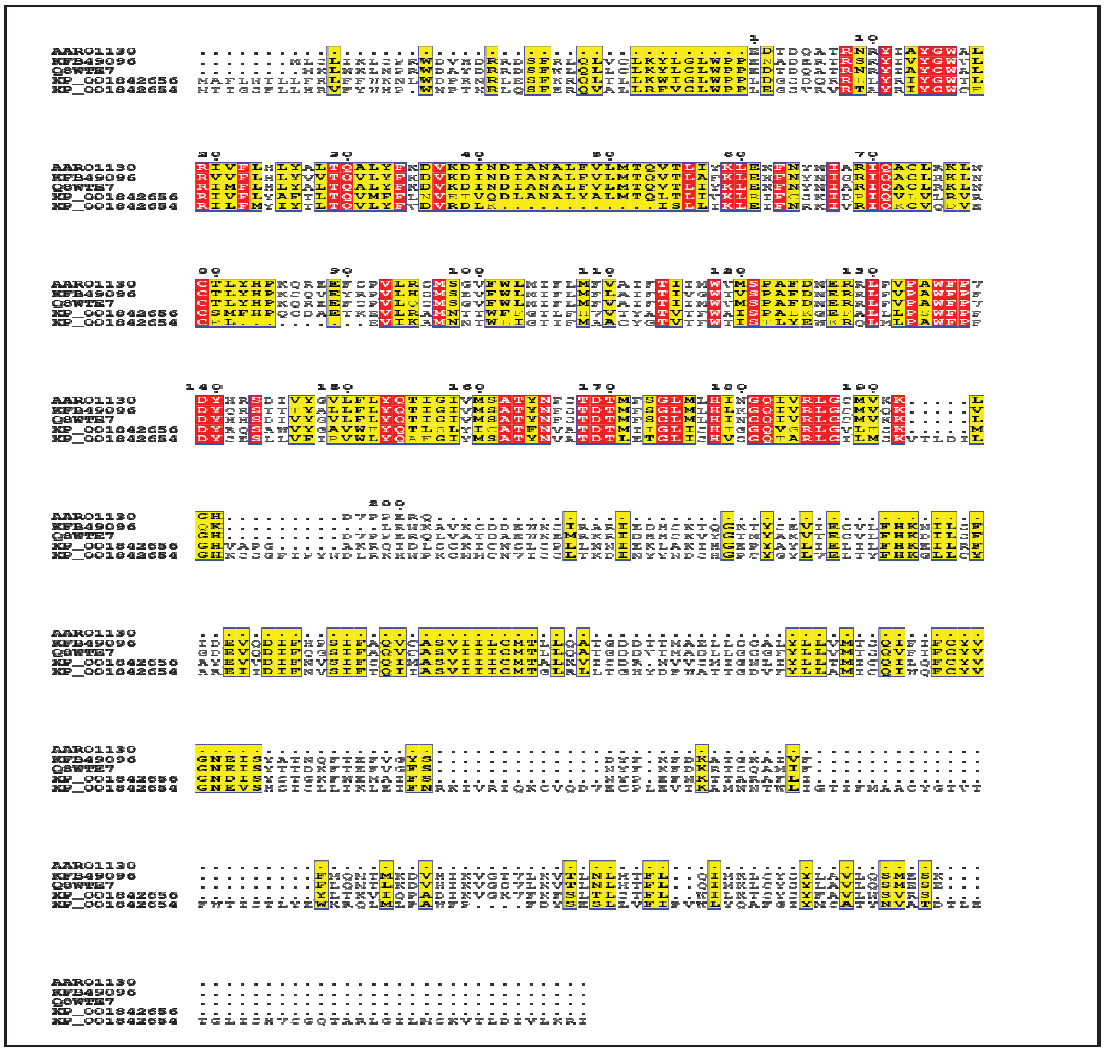


Figure 1: Multiple sequence alignment of AgOR1 protein with homologous sequences reported in the UniProt database. The conserved residues are shown in red background.

A.2 Structure Prediction and Active Site Prediction

The FASTA sequence of 8QWTE7 retrieved from UniProt database was subjected for BLAST search against the PDB database proteins, the query sequence was found to have a very low query coverage and identities. Hence the results could not be used for modelling the structure by using Homology modelling. It proved out to be a major problem in our study. So, to shoot out this problem we chose I-TASSER [7, 8] server for

modelling the structure. I-TASSER determined 5 models of the same sequences that were provided for structure prediction. The PDB ID of top 10 threading templates used as base for determining the structure are 1JB0, 1ST6, 1TR2, 2J69, 2O01, 2O8E, 2YEV, 3WDO, 4IGG, 4P7H. Based on the confidence score (C-Score) of -1.69 and cluster density of 0.1923 we considered model 2 and further proceeded for the structure validation. The structure validation was done using PDBsum, [23] the Ramachandran plot analysis of model2 suggested 79.8% of residues in most favoured region, 17.1% of residues in additional allowed region, 1.8% of residues are in generously allowed region, and 1.3% of residues in the disallowed region (Fig. 2A and Table 1). From the validation it can be concluded that the structure can be considered for further studies, which hankered us for understanding the dynamics and characteristics of the AgOR1 protein. To perform docking with all the 14 compounds, we need to have active sites of the protein. So, we used MolDock [16] algorithm embedded in Molegro Virtual Docker Software and predicted 5 cavities. The cavities are spread across the protein, and covered almost all the pockets available for suitable binding of compounds with AgOR1. The 5 cavities and the volume are Cavity1 [Vol=150.528Å^3], Cavity2 [Vol=68.608Å^3], Cavity3 [Vol=62.976Å^3], Cavity4 [Vol=62.976Å^3] and Cavity5 [Vol=48.128Å^3] (Fig. 3A).

Table 1: The structure validation was done by plotting Ramachandran Plot. It is framed in two batches, one before the MD Simulation and second plot after MD Simulation.

Results before MD Simulation	No. of residues	Percentage of residues	Results after MD Simulation	No. of residues	Percentage of residues
Most Favoured Regions	308	79.8%	**Most Favoured Regions**	313	81.1%
Additional allowed regions	66	17.1%	**Additional allowed regions**	70	18.1%
Generously allowed regions	7	1.8%	**Generously allowed regions**	1	0.3%
Disallowed regions	5	1.3%	**Disallowed regions**	2	0.5%
Total (Non-glycine and non-proline residues)	386	100%	**Total (Non-glycine and non-proline residues)**	386	100%
End residues (excluding Gly and Pro)	2		**End residues (excluding Gly and Pro)**	2	
Glycine residues	18		**Glycine residues**	18	
Proline residues	11		**Proline residues**	11	
Total number of residues	417		**Total number of residues**	417	

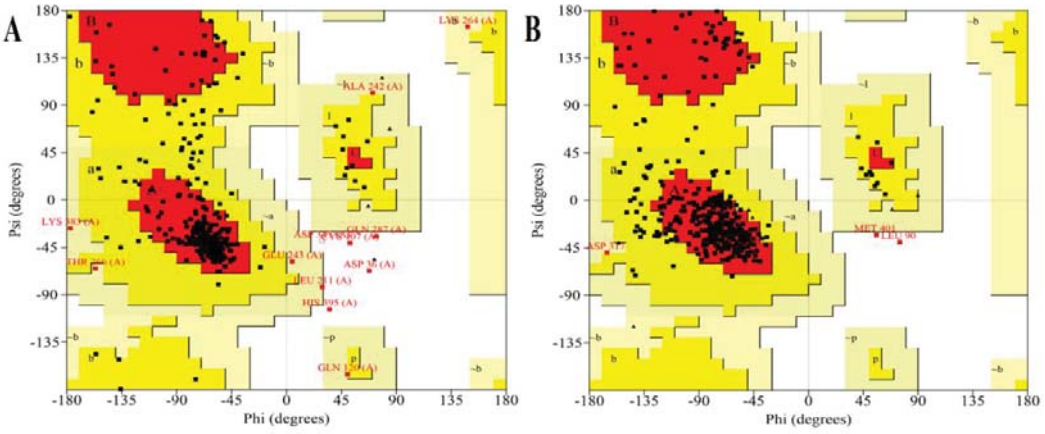


Figure 2: Ramachandran plot showing allowed dihedral (ψ and ϕ) angles having colour codings: A) The Ramachandran plot for the modelled AgOR1 protein before the MD Simulation and B) The Ramachandran plot for the modelled AgOR1 protein after the MD Simulation. The colour coding can be understood by following i) Red most favoured core regions (A,B,L) ii) Yellow allowed regions or additionally favoured (a,b,l,p) iii) Green generously allowed regions (~a,~b,~l,~p) iv) Grey disallowed regions v) Squares represent allowed residues vi) Triangles represent Glycine residue and vii) X represent disallowed residues. From these plot it is clearly understood that a structural transition has occurred during the MD Simulation and the structure has attained its stability.

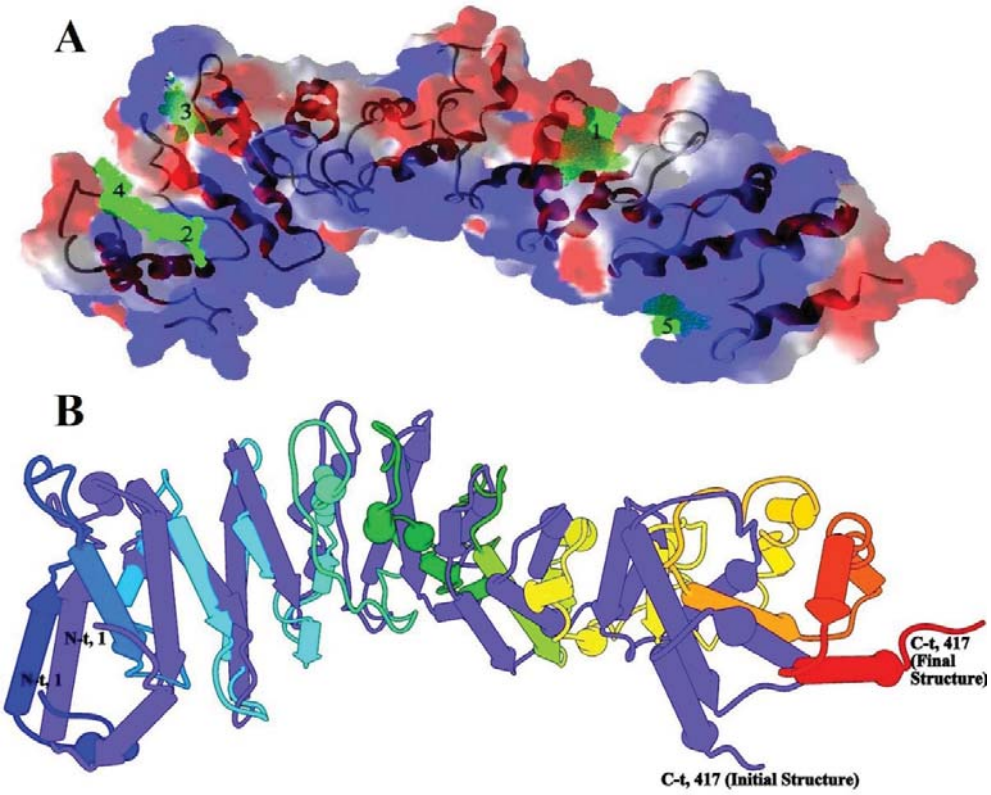


Figure 3: A) 3 D structure of Anopheles gambiae odorant receptor (AgOR1) protein. The five cavities that were detected using Molegro virtual docking software tool were shown in green colored mesh. B) Superimposed Structure between initially built and final equilibrated structure of the AgOR1 protein. The purple colour represents the initially built model structure of the protein while the rainbow colour represents the final equilibrated structure of the protein. These structures are obtained using pipes and planks option of UCSF Chimera.

A.3 Molecular Dynamics and Characterization of AgOR1

The degree of conformational changes that has occurred during the simulation of the initially built AgOR1 system is easily monitored by the C-α root mean square deviation. The backbone root mean square deviation (RMSD) of the AgOR1 with respect to the initial structure during the 12 ns MD simulation as a function of time was plotted as shown (Fig. 4). We can see that the structure of AgOR1 system has undergone vast change in conformation and then finally reached the equilibration during the course of simulation period and found that the RMSD values settle around 7.0 Å. The structural changes that have been observed between the initial and the final equilibrated structure of AgOR1 (Table 2), the result is obtained using YASARA [9] We can see that there are many secondary structural transitions across the chain of the AgOR1 system resulting in the reduction of helical structure and subsequent increase in the loop regions. Because of increase in the percentage of turns and coils, the RMSD value settles at a higher value that is around 7.0 Å. The superimposed image of initial and final equilibrated structure of AgOR1 obtained using UCSF Chimera [18] (Fig. 3B). The conformational changes can be understood by analysing the Ramachandran plot which has been obtained from the structure after the MD Simulation. This analysis indicates that the structure has attained its stability as the Simulated structure attains 81.1% of residues in most favoured region, 18.1% of residues in additional allowed region, 0.3% of residues are in generously allowed region, and 0.5% of residues in the disallowed region (Fig. 2B)

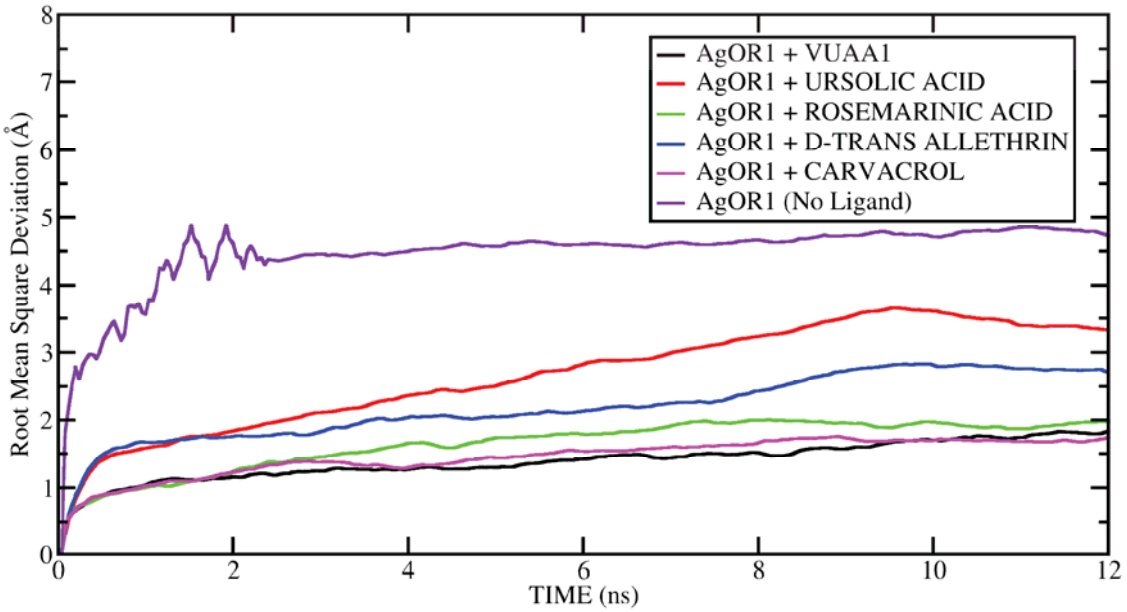


Figure 4: C-α RMSD plot for AgOR1 protein is shown as a function of time in nanoseconds. The RMSD of AgOR1-VUAA1 complex is represented in black, AgOR1-Ursolic acid complex in red, AgOR1-Rosmarinic acid complex in green, AgOR1-D Trans allethrin complex in blue, AgOR1-Carvacrol in pink and AgOR1 with no ligand in purple.

Table 2: Secondary Structure Analysis (%). This data has been established by comparing the initial (before MD Simulation) AgOR1 and the final (after MD Simulation) AgOR1 Protein Structure using YASARA bio-informatics software.

Name	Helix %	Sheet %	Turn %	Coil %	(3_{10}) Helix %
AgOR1 Initial Structure	72.60	0.0	5.80	20.40	1.20
AgOR1 Final Structure	34.80	0.0	27.8	31.2	6.2

In order to obtain the effective size and compactness of the AgOR1 protein during the simulation, the radius of gyration (Rg) was analysed. Information regarding the overall shape of the molecule can be gleaned from Rg. The time course of the radius of gyration for AgOR1 protein (Fig. 5). We can infer from the RMSD (Fig. 4) that although there is a structural transition in the structure (Table 2) there is increase in size of the system till 2 ns of simulation and after that the value of Rg shows little fluctuation and in comparison with the complex systems the structure is compact. This again confirms that the AgOR1 system has undergone rapid conformational changes since the initial time of simulation. To confirm the conformational changes we rolled a probe over the surface of the protein and studied the Solvent Accessible Surface Area (SASA) for the protein and can be better understood from (Fig. 7). We observed that the accessible surface area of AgOR1 changes rapidly during the course of simulation period. This may be due to the conformational changes that has occurred around the hydrophobic regions.

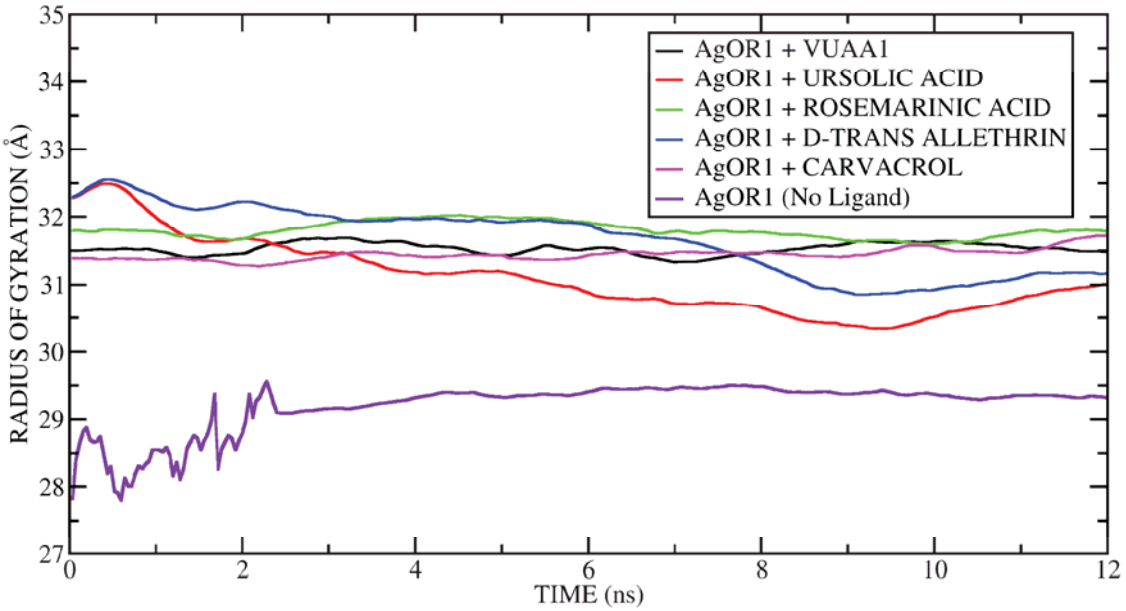


Figure 5: Radius of gyration of AgOR1 protein as a function of time course simulation in nanoseconds. The Rg of AgOR1-VUAA1 complex is represented in black, AgOR1-Ursolic acid complex in red, AgOR1-Rosmarinic acid complex in green, AgOR1-D Trans allethrin complex in blue, AgOR1-Carvacrol in pink and AgOR1 with no ligand in purple.

In order to understand the local deformations and dynamic behaviour of the residues in the AgOR1 protein chain, B-Factor value was calculated from the MD simulation. From the plot (Fig. 6), we can see that residues having index from 30 – 45, 82 – 100, 188 – 209, 244 – 257, 322 – 330, 343 – 346, 358 – 365, 381 – 387, 398 – 403, and 414 – 417 are more flexible than the other regions in the protein. From the secondary structure analysis (Table 2) we can predict that the above mentioned regions are generally loop regions.

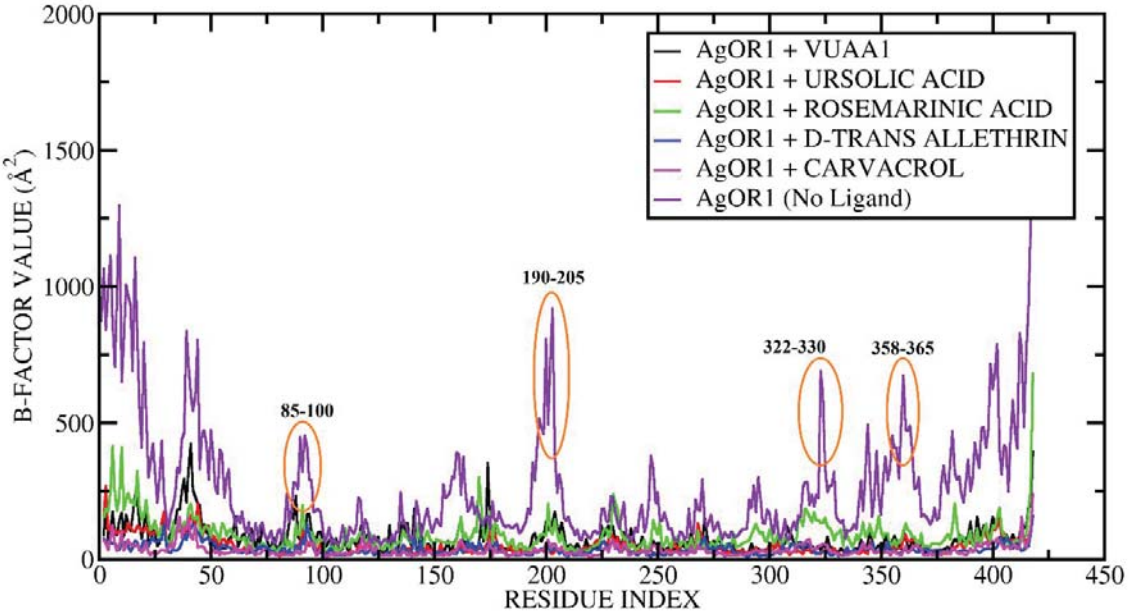


Figure 6: B-factor values for amino acids in AgOR1 protein obtained from 10 nanoseconds molecular dynamics simulation. The residues having higher B-factor values are encircled. The B-Factor of AgOR1-VUAA1 complex is represented in black, AgOR1-Ursolic acid complex in red, AgOR1-Rosemarinic acid complex in green, AgOR1-D Trans allethrin complex in blue, AgOR1-Carvacrol in pink and AgOR1 with no ligand in purple.

In order to analyse the change in secondary structure patterns in AgOR1, we applied the software tool Database of Secondary Structure in Proteins (DSSP) by Kabsch and Sander. [24] This tool particularly employs H-bonding patterns and various other geometrical features to assign secondary structure labels to the amino acid

residues of a protein. We have plotted the secondary structure patterns of AgOR1 as shown (Fig. 8). The analysis of the plot shows the structural variation of each residue during the time course of simulation period. We observe rapid conformational drift from one secondary structure to another secondary structure in the residues having index: 240 - 255, 190 - 210, 384 - 402, 88 - 102, 340 - 347, 323 - 330, 56 - 61, and in 29 - 42. These transitions are observed to be mainly from helix to turns and coils leading to flexible loop regions (Table 2). We can also get the probable secondary structure information of AgOR1 structure during the course of 12 ns MD Simulation (Fig. 9).

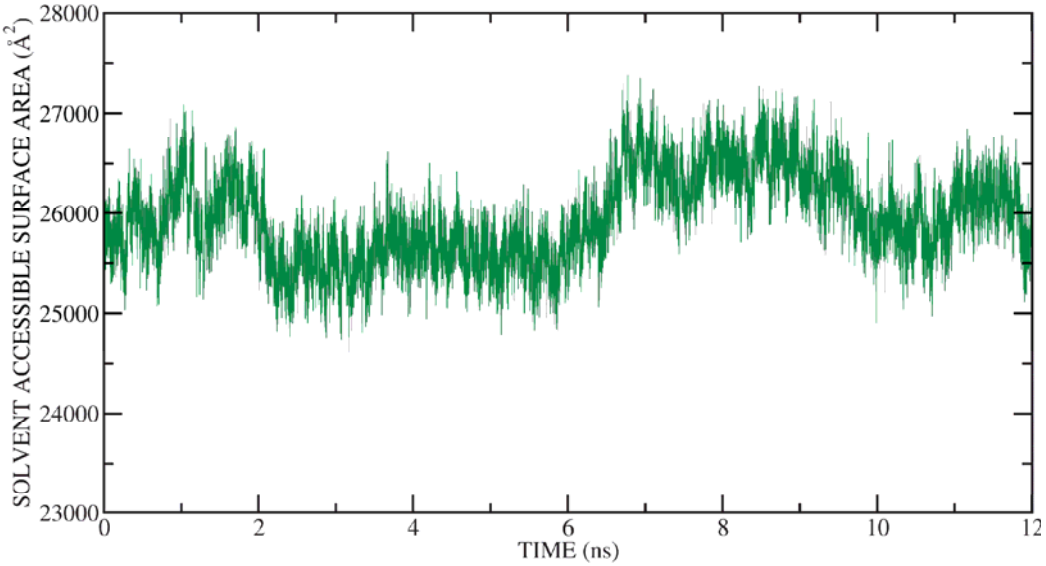


Figure 7: Solvent Accessible Surface Area of AgOR1 protein as a function of simulation time period in nanoseconds.

From the Probability score as shown (Fig. 9) we can predict that most of the time we observe the Para type of secondary structure to be predominant in the residue index 231 and 266. Anti-region is predominant in Residue index 132 - 133, and 157 - 164. 3 10 Helix is predominant in the residue index of 53 - 71, 129 - 138, 150 - 157, 205 - 211, 348 - 355, 382 - 388, and 394 - 403. The Alpha region shows predominance in the residue index of 18 - 29, 45 - 54, 104 - 112, 125 - 131, 125 – 131, 212 - 218, 256 - 267, 279 - 291, 298 - 307, 366 - 375, and 407 - 412. The pi region although has very low probability score but can also be seen (Fig. 9) that, it can be predominant in Residue index of 204 - 212. The turns are found to have high probability to be predominant in residue index of 36 - 41, 133 - 138, 183 - 185, 224 - 229, 239 - 244, 292 - 298, and 339 - 342.

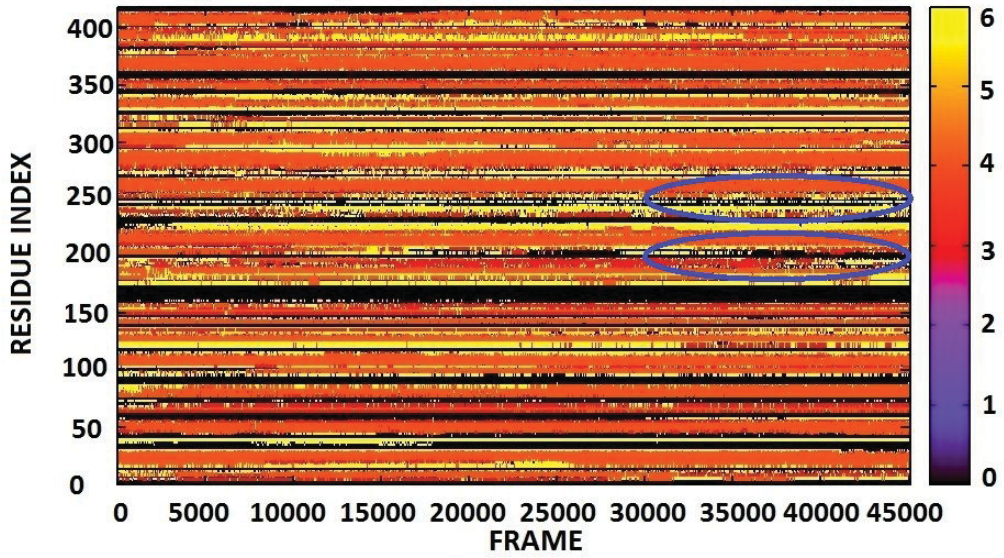


Figure 8: Secondary Structure Analysis of AgOR1 during the time course of MD Simulation. The encircled portion of the plot represents the region of the protein having highest fluctuation.

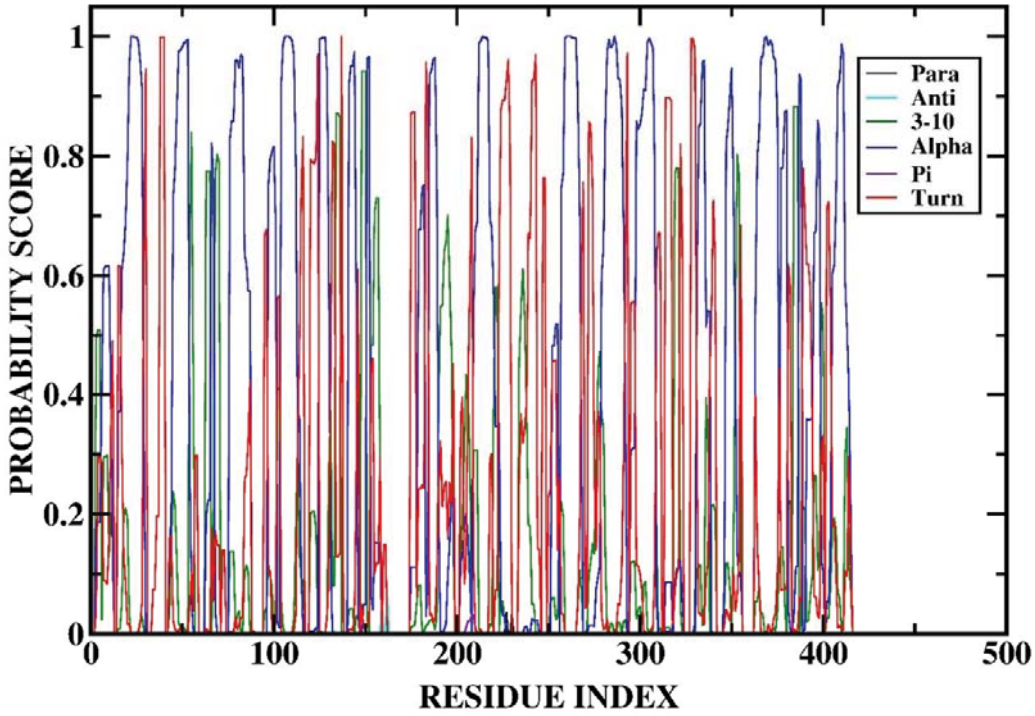


Figure 9: Probability score for the individual secondary structure content vs. Residue Index plot for AgOR1 protein.

B. VUAA1 shows best binding with cavity 3 and best binding affinity towards AgOR1

The AgOR1 structure is docked with VUAA1 at five different sites (cavities) spread across the protein and detailed (Table S1). The VUAA1 molecules shows the strongest interaction with Lys72, Asn79 and Glu22 with the docking score of -182.696 and shows the best interaction among all the 14 molecules (Fig. 10A). To better understand the stability of the complex we performed MD Simulation of the docked structure of AgOR1 and VUAA1. The RMSD analysis shows that the lowest deviation of 1.8Å as compared to AgOR1 alone and also among all the competitive molecules which infers that the AgOR1 structure is not allowed to change its conformation as the binding site is occupied by VUAA1, which indicates that the structure is stable and the OR is incapable of accommodating any other odor, thus deviating itself from the prey (Fig. 4). The compactness of the system is also seen by plotting the Rg, (Fig. 5) it settles at the range of 31-31.5 Å which is again less as compared to AgOR1 system alone, along with it the B-factor or the Temperature factor show the deformation of the complex system in three different regions with the residual index of 35-42, 85-98 and 170-179 (Fig. 6). The Binding Energy Calculation (BEC) has been done to check the reliability of the docking results and it has been shown (Table 3), the Poisson Boltzmann Total (PBTOT) and Generalized Born Total (GBTOT) indicates the total binding free energy of VUAA1 in AgOR1 protein using MM-PBSA and GBSA. The result of BEC of VUAA1 has found to be the best scoring among all the other compounds and the ΔG_{bind} –PBTOT and ΔG_{bind}– GBTOT are -30.83 ±3.23 and -30.03±3.91 (Table 3).

Table 3: MM-GBSA/PBSA Binding free energy results for AgOR1 interacting with the ligands VUAA1, Ursolic Acid, Rosmarinic Acid, D-Trans Allethrin and Carvacrol.

Parameters	ΔG_{bind} (AgOR1-VUAA1)		ΔG_{bind} (AgOR1-Ursolic Acid)		ΔG_{bind} (AgOR1-Rosmarinic Acid)		ΔG_{bind} (AgOR1-D-Trnas Allethrin)		ΔG_{bind} (AgOR1-Carvacrol)	
	MEAN	**STD**	**MEAN**	**STD**	**MEAN**	**STD**	**MEAN**	**STD**	**MEAN**	**STD**
ELE	-56.73	5.45	-10.08	3.62	-31.66	4.96	-18.11	3.83	-16.55	3.17
VDW	-45.94	2.71	-47.95	3.48	-47.00	2.96	-42.16	3.15	-25.45	2.15
INT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GAS	-102.67	4.53	-58.03	3.75	-78.65	5.40	-60.27	5.29	-42.01	2.64
PBSUR	-3.99	0.08	-3.93	0.12	-4.05	0.11	-3.79	0.09	-2.58	0.06
PBCAL	75.84	4.54	31.93	3.82	59.23	6.18	42.67	4.65	31.34	2.38
PBSOL	71.85	4.52	28.00	3.81	55.18	6.16	38.87	4.61	28.76	2.38
PBELE	19.11	3.23	21.85	2.86	27.57	5.86	24.55	3.65	14.79	3.41
PBTOT	**-30.83**	**3.23**	**-30.03**	**3.91**	**-23.47**	**5.76**	**-21.40**	**5.33**	**-13.25**	**3.13**
GBSUR	-3.99	0.08	-3.93	0.12	-4.05	0.11	-3.79	0.09	-2.58	0.06
GB	68.48	3.69	24.30	2.91	44.29	4.40	3.89	3.69	21.14	1.67

GBSOL	64.48	3.68	20.36	2.92	40.24	4.38	28.09	3.68	18.56	1.68
GBELE	11.74	3.19	14.21	1.53	12.63	2.20	13.77	1.81	4.59	2.28
GBTOT	**-38.19**	**3.23**	**-37.67**	**3.34**	**-38.41**	**3.46**	**-32.18**	**4.23**	**-23.45**	**2.35**

ELE - Non-bonded electrostatic energy + 1, 4-electrostatic energy
VDW - Non-bonded van der Waals energy + 1, 4-van der Waals energy
INT - Bond, angle, dihedral energies
GAS - ELE + VDW + INT
PBSUR - Hydrophobic contribution to solvation free energy for PB calculations
PBCAL - Reaction field energy calculated by PB
PBSOL - PBSUR + PBCAL
PBELE - PBCAL + ELE
PBTOT - PBSOL + GAS
GBSUR - Hydrophobic contribution to solvation free energy for GB calculations
GB - Reaction field energy calculated by GB
GBSOL - GBSUR + GB
GBELE - GB + ELE
GBTOT - GBSOL + GAS

C. Established repellents showing interaction with AgOR1

The five repellents DEET, D-Trans Allethrin, Prallethrin, Rebemide, and Transfluthrin were docked into the five cavities and it has been screened that all the molecules shows better binding affinity towards cavity 3 with the docking score of -72.3972, -152.02, -156.445, -92.7751 and -142.168 respectively. To minimize the computational cost we screened the molecules based on their protein interaction score for further studies (Table S1). The interactions of the five molecules with AgOR1 can be seen (Fig. 10B & Fig. S1). The docked structure of D-trans allethrin and AgOR1 has been screened which has protein interaction score of -109.762 having interaction with Lys69 and Asn79 with H-bond interaction of 2.28Å and 2.54Å respectively. All the other repellents that is DEET, Prallethrin, Rebemide and Transfluthrin show internal inter bonds and hydrophobic bonds with the AgOR1 protein with protein interactions of 179.925, 401.824, 152.639 and -17.967 respectively. Further into the study we have considered the D-trans allethrin and AgOR1 complex for MD Simulation and later using the snapshots as explained in Materials and Methods calculated the binding free energy. The BEC of the D-trans allethrin accommodated in the AgOR1 protein ΔG_{bind}–PBTOT and ΔG_{bind}–GBTOT are -21.40±5.33 and -32.18±4.23 respectively. These values are less than AgOR1-Ursolic Acid and AgOR1-Rosmarinic Acid (Table 3) from which we can again come to a conclusion that some of the phytochemical can tightly bound to the AgOR1 inhibiting its odor sensing property. The MD Simulation results tell us that the RMSD value approximately settles around 3Å as compared to the RMSD of AgOR1. The fluctuation is quite less, but more than VUAA1, phytochemicals such as Rosmarinic acid and Carvacrol, which suggest that these phytochemicals may have good interactions and can better inhibit the AgOR1 (Fig. 4). The Rg and B-Factor value shows contraction and less fluctuation in the residues respectively, which again suggest that the repellent retains its inhibitory properties.

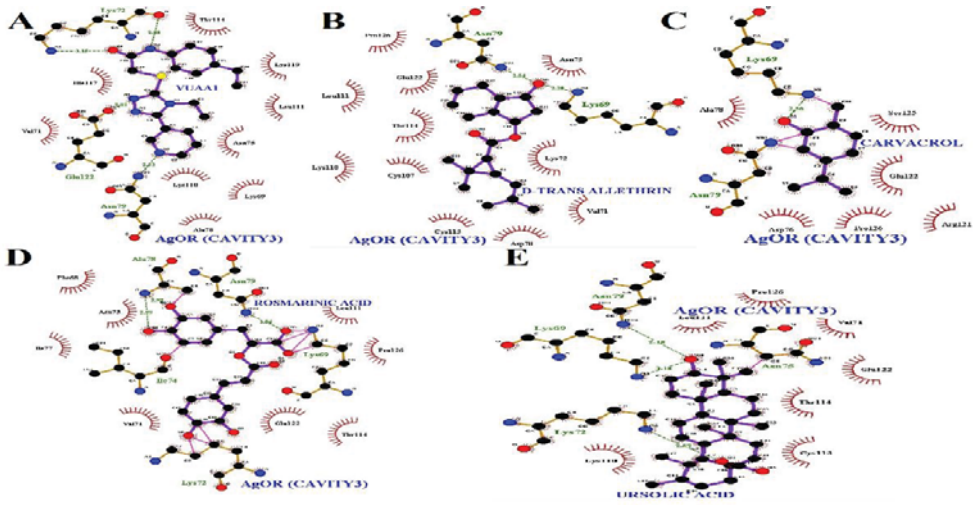


Figure 10: Interaction of Odorant molecules (inhibitors) with *Anopheles gambiae* protein. These figures are made using LigPlot+: A) The interaction between Anopheles gambiae Odorant Receptor (AgOR1) and VUAA1 is shown. Here the H-Bond interaction with Lys72, Asn79, and Glu122 can be seen. The residues with red spikes are shown as the Hydrophobic Bonds. B) Lys69 and Asn79 shows H-Bond interactions, no external bonds are formed but some residues are found to have hydrophobic interactions with D-Trans allethrin. C) Lys69 alone shows H-Bond interaction and along with Asn79 shows external bond with Carvacrol. Few residues also show hydrophobic bonds. D)

Here the H-Bond interacting residues are Ala78 and Asn79. The residues Lys69, Lys72, and Ile74 shows external bonding and all other residues shows hydrophobic bonds with the Rosemarinic Acid. E) Lys69, Lys72, and Asn79 shows H-Bond Interaction, Asn75 shows external bonds as well as other depicted residues shows hydrophobic interaction with the Ursolic Acid. The H-bond interactions are shown in green dotted lines between the residues. The bonds represented as pink colour are internal bonds between the molecules and the AgOR1 protein residues. The red spikes are represented as hydrophobic interactions between the AgOR1 protein residues and the molecules.

D. Phytochemicals show relative results as the repellents

The phytochemicals from the *Ocimum tenuiflorum* namely A-Elemene, Carvacrol, Caryophellene, Eugenol, Linalool, Oleanolic acid, Rosmarinic acid and Ursolic acid are docked with the modelled AgOR1 structures at five different sites. From the docking results we can infer that the phytochemicals has consistent binding affinity towards cavity 3 and analysing relatively based on their docking score, protein interaction score and hydrogen bond formation (Fig. 10C-10D-10E, Fig. S1 and Table S1) we considered Ursolic acid and Rosmarinic acid for further studies. We have also chosen Carvacrol which has low docking score and protein interaction score than some of the repellents such as Transfluthrin and Rebemide to make a comparative analysis with D-trans allethrin which is already in use as a repellent. MD Simulation was done on three complexes AgOR1-Ursolic acid, AgOR1-Rosmarinic acid and AgOR1-carvacrol respectively. The snapshots of the respective complexes are used for BEC. The ΔG_{bind}-PBTOT and ΔG_{bind}-GBTOT values computed for AgOR1-Ursolic acid complex are -30.03±3.91 and -37.67±3.34, the ΔG_{bind}-PBTOT and ΔG_{bind}-GBTOT values computed for AgOR1-Rosmarinic acid complex are -23.47±5.76 and -38.41±3.46, similarly the ΔG_{bind}-PBTOT and ΔG_{bind}-GBTOT values computed for AgOR1-Carvacrol complex are -13.25±3.13 and -23.45±2.35 (Table 3). The MD Simulation trajectory analysis of the complexes reveal that B-Factor graphs of Carvacrol and Ursolic is following the same profile as D-trans allethrin but Rosmarinic acid is showing deformation in the residual range of 85-98 and 170-179 similar to the deformation profile of AgOR1-VUAA1 complex. The RMSD results suggests that AgOR1-Ursolic acid complex is showing a deviation of 3.4Å which is highest among the other complexes whereas Carvacrol and Rosmarinic acid are having same profile as VUAA1. The radius of gyration tells us that the Ursolic acid has undergone relatively much more contraction than the other complexes. From these results we can infer that the phytochemicals are sharing more or less similar results as repellents. Although our results might prove out to be correct but the experimental evidences are required to support our work.

IV. CONCLUSIONS

In this study we have obtained the salient structural features of the modeled structure of the AgOR1 and its binding interaction with some of the commonly available mosquito repellants and phytochemicals from *O.tenuiflorum* by carrying out Molecular Docking, MD Simulation and Binding Free Energy Calculation. From the docking and binding free energy calculations, we observed that VUAA1 (repellent), D-trans allethrin (repellent), Ursolic acid, Rosmarinic acid and Carvacrol from *O.tenuiflorum* to show strong affinity with the AgOR1 among the 14 selected chemicals. In AgOR1, Asn79, Lys69, Ala78, Lys72 and Glu122 found to be the interacting residues with the inhibitors. It is also observed that the residues Lys69 and Asn79 are involved in H-bond interaction with most of the inhibitors under study. We found the inhibition order of the inhibitors under this study with AgOR1 to be VUAA1 > Ursolic acid > Rosmarinic acid > D-Trans allethrin > Carvacrol > Linalool > Prallethrin > Oleanolic Acid >Transfluthrin > DEET > Rebemide > Eugenol >Caryophellene > A-Elemene based on Docking Score, and Protein – Ligand Interaction, MM-GBSA/PBSA studies. Although VUAA1 shows strong inhibition action on AgOR1, it has the problem of being non-volatile in nature. So, this study suggests that we have to look for an inhibitor having similar structural characteristics as that of VUAA1 and at the same time it should be volatile. This study also suggests that the phytochemicals of *O.tenuiflorum* plant such as Ursolic acid, Rosmarinic acid and Carvacrol can also play a critical role in inhibiting the Odorant Receptor and thus controlling the spread of disease. As a whole we see that our findings in this study may aid in designing new and more effective repellents to control mosquito related health problem. This study also suggests that the phytochemicals can be experimentally studied so that they can be used as repellents. Their derivatives (pharmacophores) can also be studied for designing potential repellents.

ACKNOWLEDGEMENTS:

We would like to thank the DBT funded Bioinformatics Infrastructure facility in the Department of Molecular Biology and Biotechnology at Tezpur University for providing us computational facility to carry out this research work.

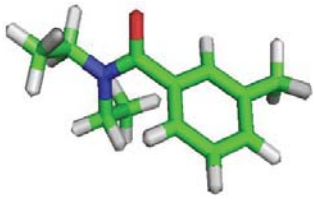
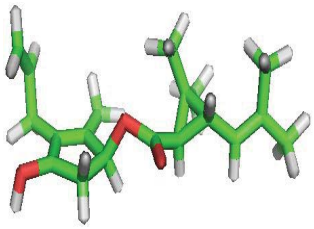
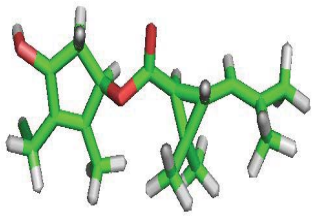
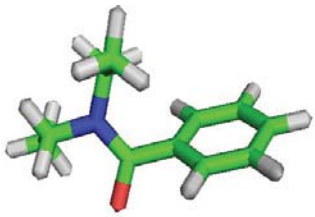
REFERENCES

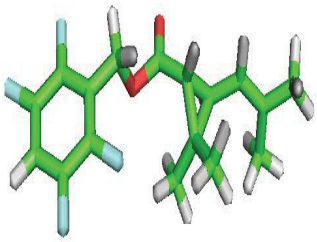
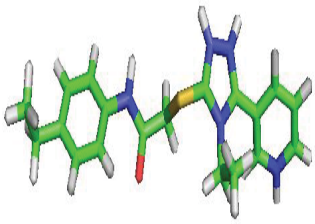
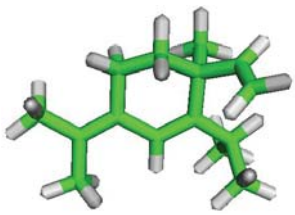
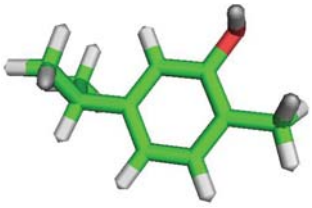
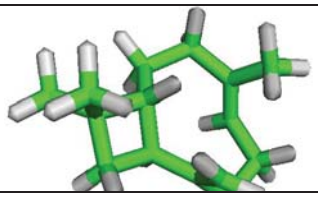
[1] Hallem EA, Fox AN, Zwiebel LJ, Carlson JR. Olfaction: mosquito receptor for human-sweat odorant. Nature. 2004 Jan 15;427(6971):212-3

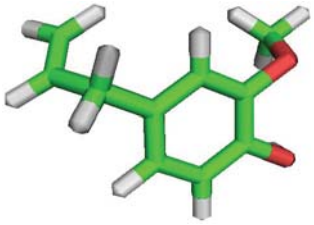
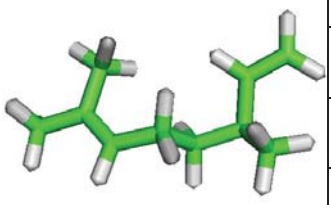
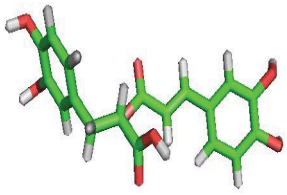
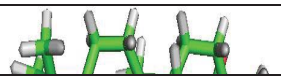
[2] Ziemba BP, Murphy EJ, Edlin HT, Jones DN. A novel mechanism of ligand binding and release in the odorant binding protein 20 from the malaria mosquito Anopheles gambiae. Protein Science. 2013 Jan 1;22(1):11-21

[3] Rane R, Ravichandran R, Panda PK, Joshi S. Computer Aided Pharmacophore Analysis Of Inhibitory Action Of Phytochemicals Against Mosquito Odor Receptors In Comparison To Chemical Drugs. World Journal Of Pharmacy And Pharmaceutical Sciences. 2015 May 16;4(7):2012-21
[4] Homeyer N, Gohlke H. Free energy calculations by the molecular mechanics Poisson− Boltzmann surface area method. Molecular Informatics. 2012 Feb 1;31(2):114-22
[5] Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic acids research. 2009 Jun 4:gkp456
[6] Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic acids research. 2014 Jul 1;42(W1):W320-4
[7] Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols. 2010 Apr 1;5(4):725-38
[8] Zhang Y. I-TASSER server for protein 3D structure prediction. BMC bioinformatics. 2008 Jan 23;9(1):40.
[9] Krieger E, Vriend G. YASARA View—molecular graphics for all devices—from smartphones to workstations. Bioinformatics. 2014 Oct 15;30(20):2981-2
[10] Schafmeister CE, Ross WS, Romanovski V. LEaP, University of California, San Francisco. There is no corresponding record for this reference. 1995
[11] Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, Ferguson D, Seibel G, Kollman P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications. 1995 Sep 2;91(1):1-41
[12] Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics. 2006 Nov 15;65(3):712-25
[13] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics. 1983 Jul 15;79(2):926-35
[14] Berendsen HJ, Postma JV, van Gunsteren WF, DiNola AR, Haak JR. Molecular dynamics with coupling to an external bath. The Journal of chemical physics. 1984 Oct 15;81(8):3684-90
[15] Ryckaert JP, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics. 1977 Mar 31;23(3):327-41
[16] Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. Journal of medicinal chemistry. 2006 Jun 1;49(11):3315-21
[17] Yang JM, Chen CC. GEMDOCK: a generic evolutionary method for molecular docking. Proteins: Structure, Function, and Bioinformatics. 2004 May 1;55(2):288-304
[18] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004 Oct 1;25(13):1605-12
[19] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of molecular graphics. 1996 Feb 29;14(1):33-8
[20] Laskowski RA, Swindells MB. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. Journal of chemical information and modeling. 2011 Oct 5;51(10):2778-86
[21] Hawkins GD, Cramer CJ, Truhlar DG. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. The Journal of Physical Chemistry. 1996 Dec 19;100(51):19824-39
[22] Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers. 2000 Jan 1;56(4):275-91
[23] de Beer TA, Berka K, Thornton JM, Laskowski RA. PDBsum additions. Nucleic acids research. 2014 Jan 1;42(D1):D292-6
[24] Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983 Dec 1;22(12):2577-637

Table S1: Here in the table we have mentioned the molecules that were considered for docking with the equilibrated structure of AgOR1 to check the best inhibitors out of the 14 structures. The 3D structures of the inhibitors are mentioned below with the cavity with which it is docking, mentioning the results of the other parameters: The MolDock Score, H-Bond Distance, Docking Score, The Protein Interaction Score, and the Interacting Residues of AgOR1 of cavity 3 along with virtual screening are considered to rank the chemicals.

Sl. No	MOLECULE NAME AND 3D STRUCTURE	CAVITY NO.	DOCKING SCORE	PROTEIN INTERACTION	INTERACTING RESIDUES
1.	 Deet CID 4284	Cavity 1	-92.7882	310.423	Val257, Tyr258, Gly259, Thr260, Met261, Phe286, Ser297, Val298, Ile300, Cys302, Thr304.
		Cavity 2	-81.3533	192.959	Gly49, Leu52, Arg53, Ile54, Leu57, Trp60, Ala61, Leu81, Leu84, Met85.
		Cavity 3	-102.129	179.925	Asp76, Asn79, Lys110, Leu111, Thr114, Arg121, Glu122, Ser125, Pro126, Gln129.
		Cavity 4	-72.3972	37.7865	Ile77, Asp76, Asn79, Ala80, Leu81, Arg121, Ser125, Leu128.
		Cavity 5	-84.3202	-5.37625	Gln367, Phe397, Ile400, Lys402.
2.	 D- Trans Allethrin CID 15558638	Cavity 1	1121.64	114.779	Tyr258, Tyr262, His272, Leu276, Phe278, Gly279, Asp280, Gln283, Ala308, Asp312, Thr314, Leu318.
		Cavity 2	-137.375	258.121	Leu29, Leu52, Arg53, Tyr60, Ala61, Ala65, Leu81, Phe82, Leu84, Met85, Thr86.
		Cavity 3	-152.02	-109.762	Lys69, Asp70, Val71, Lys72, Asn75, Asn79, Cys107, Lys110, Cys113, Thr114, Glu122, Pro126.
		Cavity 4	-119.725	-23.8739	Glu35, Arg42, Ile74, Asp78, Lys119.
		Cavity 5	-130.262	60.8327	Phe359, Lys363, Gln367, Ile400, Met401, Lys402, Leu403.
3.	 Prallethrin CID 71306899	Cavity 1	-143.46	324.311	Tyr262, Thr266, His272, Lys273, Leu276, Gln283, Ala308, Thr309, Asp312, Leu318.
		Cavity 2	-132.355	30.9269	Leu29, Tyr45, Ile46, Gly49, Leu52, Arg53, Met55, Phe56, Ala61, Leu81.
		Cavity 3	-156.445	401.824	Lys69, Asp70, Val71, Asp73, Ile74, Asn75, Lys110, Leu111, Asn112, Thr114, Glu122, Pro126.
		Cavity 4	-118.908	-10.0648	Lys27, Tyr28, Leu29, Gly30, Leu31, Pro33, Pro34, Glu35, Asp36, Thr37, Asp38, Arg44, Tyr45.
		Cavity 5	-123.719	-57.3019	Phe359, Lys363, Ser366, Ile370, Phe397, Ile400, Lys402, Leu403, Ser404.
4.	 Rebemide CID 15542	Cavity 1	-81.8749	107.135	Tyr262, Thr266, Leu276, Ala308, Thr309, Asp312.
		Cavity 2	-74.2111	238.092	Gly49, Trp50, Leu52, Arg53, Ile54, Ala61, Leu81, Phe82, Leu84, Thr86.
		Cavity 3	-92.7751	152.639	Leu66, Phe68, Asp70, Arg103, Cys107.
		Cavity 4	-64.4979	56.0848	Gly30, Pro33, Pro34, Glu35, Asp36.
		Cavity 5	-82.1501	121.226	Phe397, Gln399, Ile400, Met401, Lys402, Leu403, Ser404.

5.	 Transfluthrin CID 656612	Cavity 1	-138.886	40.8735	Gly259, Tyr262, Ala263, Thr266, Leu276, Gly279, Val282, Ala308, Asp312, Leu318.
		Cavity 2	-124.726	552.168	Gly49, Arg53, Ala65, Ala80, Leu81, Phe82, Leu84, Thr86, Gln87, Val88, Thr89.
		Cavity 3	-142.168	-17.967	Lys69, Val71, Lys72, Asn75, Asn79, Lys110, Cys113, Thr114, Tyr116, Glu122.
		Cavity 4	-111.116	394.481	Pro33, Pro34, Asp36, Asp38, Tyr45.
		Cavity 5	-127.27	42.5034	Phe359, Lys363, Ser366, Gln367, Tyr45, Ile400, Lys402, Leu403.
6.	 VUAA1 CID 1319135	Cavity 1	-168.336	99.6885	Tyr262, Leu276, Gln283, Asp312, Asp317, Leu318, Cys321.
		Cavity 2	-159.626	786.237	Leu31, Tyr45, Ile46, Tyr48, Gly49, Trp50, Ala51, Leu52, Ile54, Met55, Phe56, Leu57, Leu81.
		Cavity 3	-182.696	-120.321	Lys69, Val71, Lys72, Asn75, Ala78, Asn79, Lys110, Leu111, Thr114, Lys119, Hie117, Glu122.
		Cavity 4	-139.791	119.275	Leu29, Gly30, Pro34, Pro35, Asp36, Thr37, Asp38, Arg44, Tyr45, Ile74, Ile77.
		Cavity 5	-146.044	94.5297	Gln367, Ile400, Met401, Lys402, Leu403, Ser404.
7.	 A-Elemene CID 80048	Cavity 1	-94.0628	86.2447	Tyr262, Leu276, Asp312, Asp317, Leu318.
		Cavity 2	-90.4247	507.655	Gly49, Trp50, Leu52, Ile54, Met55, Phe56, Leu57, Leu84, Val88.
		Cavity 3	-104.875	225.146	Asp73, Ile74, Asn75, Lys110, Leu111, Cys113, Glu122, Val171.
		Cavity 4	-75.053	455.344	Asn75, Asp76, Ile77, Ala78, Asn79, Phe82, Val83, Leu84, Met85, Arg121 Ser125.
		Cavity 5	-80.4981	-23.669	Lys363, Arg364, Ser366, Gln367, Ile370, Phe397, Ile400, Leu403.
8.	 Carvacrol CID 10364	Cavity 1	-80.8665	12.9488	Gln283, Ala308, Asp312, Thr314, Leu318.
		Cavity 2	-71.0615	153.167	Tyr48, Gly49, Trp50, Ala51, Leu52, Arg53, Ile54, Ala61, Leu84, Met85.
		Cavity 3	-87.3938	167.525	Lys69, Asp76, Ala78, Asn79, Arg121, Glu122, Ser125, Pro126.
		Cavity 4	-65.9111	67.0677	Pro33, Pro34, Glu35, Asp36, Thr37, Asp38, Tyr45.
		Cavity 5	-74.9159	111.592	Ile400, Met401, Lys402, Leu403, Ser404.
9.		Cavity 1	-101.241	-82.0038	Tyr258, Gly259, Tyr262, Ala263, Gly279, Val282, Phe286, Thr304, Leu305, Ala308.
		Cavity 2	-79.572	500.988	Gly49, Trp50, Ala51, Leu52, Ile54, Met55, Phe56, Leu57, Leu84, Val88.

	Caryophyllene CID 5281515	Cavity 3	-104.812	238.501	Lys69, Val71, Asp73, Ile74, Asn75, Lys110, Thr114, Glu122.
		Cavity 4	-78.2611	48.7036	Pro33, Pro34, Glu35, Asp36, Asp38, Tyr45.
		Cavity 5	-97.2162	-76.3737	Phe359, Ser366, Gln367, Ile370, Phe397, Ile400, Lys402, Leu403, Ser404.
10.	 Eugenol CID 3314	Cavity 1	-89.3654	114.88	Lys256, Val257, Thr260, Lys264, Ile301, Thr304.
		Cavity 2	-88.6926	-68.3998	Gly49, Leu52, Arg53, Ala61, Leu81, Leu84, Met85.
		Cavity 3	-91.8741	131.232	Val71, Asp73, Ile74, Asn75, Lys110, Leu111, Thr114, Glu122.
		Cavity 4	-82.3275	42.2531	Gly30, Leu31, Pro33, Pro34, Glu35, Asp36, Thr37, Asp38, Tyr45.
		Cavity 5	-88.6686	147.704	Phe397, Ile400, Met401, Leu403, Lys402, Ser404.
11.	 Linalool CID 6549	Cavity 1	-85.1753	46.7338	Tyr262, Thr266, Thr308, Thr309, Asp312, Leu318.
		Cavity 2	-81.3423	56.0756	Leu 29, Gly49, Leu52, Arg53, Ala61, Leu81, Leu84.
		Cavity 3	-91.638	40.5134	Lys69, Val71, Asn75, Asn79, Glu122, Pro126, Val127, Gln129.
		Cavity 4	-80.293	227.972	Lys27, Tyr28, Pro33, Glu35, Asp36, Thr37, Asp38, Arg44, Tyr45.
		Cavity 5	-87.7807	466.375	Gln293, Phe361, Asp362, Thr365, Ser366, Gln367.
12.	 Oleanolic Acid CID 10494	Cavity 1	-146.998	445.943	Tyr262, Thr266, Phe271, His272, Asp275, Leu276, Gly279, Asp280, Gln283, Ala308, Asp312, Leu318.
		Cavity 2	-106.583	52.0688	Asp76, Ala80, Leu84, Arg121, Leu164.
		Cavity 3	-126.949	253.266	Asp70, Val71, Lys72, Asp73, Asn75, Lys110, Leu111, Thr114, Leu115, Glu122.
		Cavity 4	-101.909	-24.8072	Asp38, Arg42, Asp76, Ala80, Leu81, Arg121.
		Cavity 5	-112.697	-98.9877	Phe359, Lys363, Phe397, Ile400, Lys402, Leu403, Ser404.
13.	 Rosmarinic Acid CID 5315615	Cavity 1	-172.544	352.202	Tyr262, Ala263, Thr266, His272, Lys273, Ile275, Ala308, Asp311, Asp312.
		Cavity 2	-155.77	37.3538	Gly49, Arg53, Ala61, Asp76, Ile77, Ala80, Leu81, Leu84, Met85, Arg121.
		Cavity 3	-191.049	10.7543	Phe68, Lys69, Val71, Lys72, Ile74, Asn75, Ile77, Ala78, Asn79, Leu111, Thr114, Glu122, Pro126.
		Cavity 4	-143.166	63.939	Tyr28, Gly30, Leu31, Trp32, Pro33, Pro34, Glu35, Asp36, Thr37, Asp38, Arg44, Tyr45.
		Cavity 5	-177.587	161.777	Tyr358, Ser366, Ile370, Leu373, Phe397, Leu398, Ile400, Lys402, Leu403, Ser404, Tyr407.
		Cavity 1	-123.058	-7.41167	Tyr258, Tyr262, Leu276, Phe278, Gly279, Ala308, Asp312, Asp317, Leu318.

14.	Ursolic Acid CID 64945	Cavity 2	-110.383	905.843	Leu29, Ile46, Gly49, Trp50, Arg53, Leu81, Phe82, Val83, Thr86, Thr89, Lys96, Asn100.
		Cavity 3	-127.545	-107.923	Lys69, Val71, Lys72, Asn75, Asn79, Lys110, Leu111, Cys113, Thr114, Glu122, Pro126.
		Cavity 4	-105.496	6.09314	Arg42, Asp38, Asp76, Ile77, Ala80, Leu81, Arg121.
		Cavity 5	-117.263	-88.7998	Asp362, Lys363, Arg364, Ser366, Gln367, Ile370, Phe397, Lys402, Leu403.

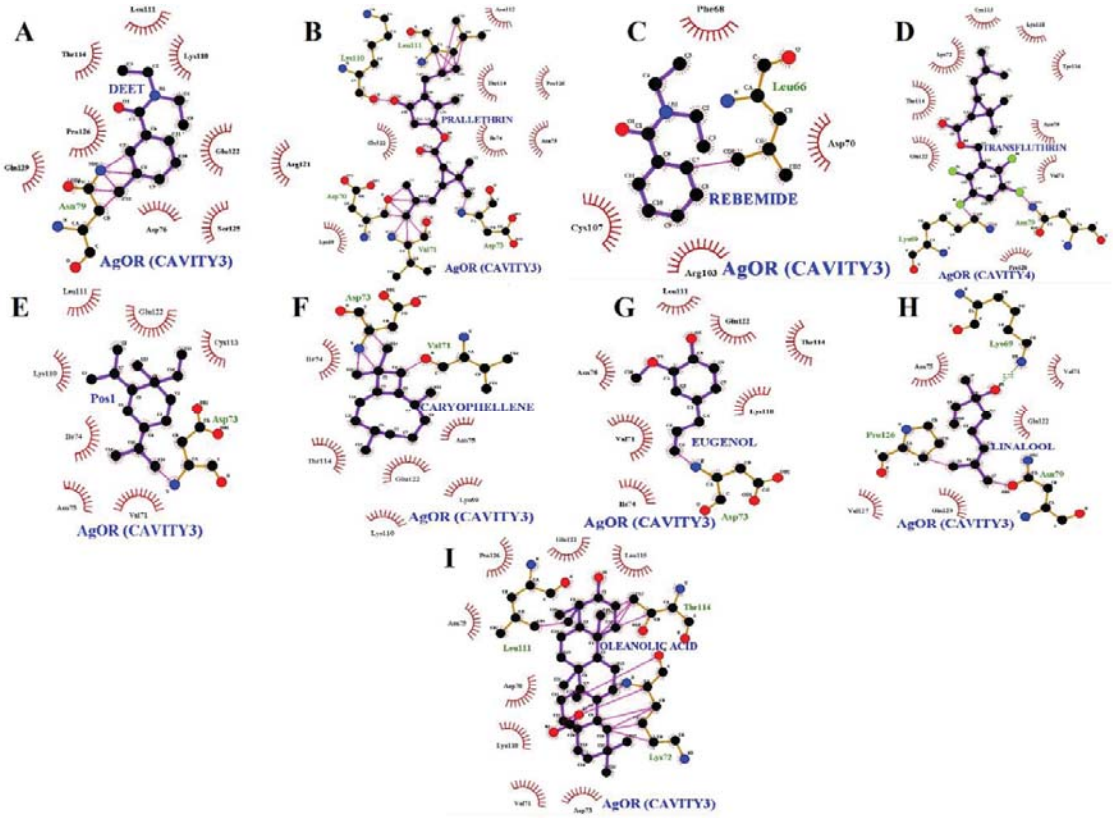


Figure S1: The H-bond interactions are shown in green dotted lines between the residues. The bonds represented as pink colour are internal bonds between the molecules and the AgOR1 protein residues. The red spikes are represented as hydrophobic interactions between the AgOR1 protein residues and the molecules.