Interaction of N-3-oxododecanoyl homoserine lactone with transcriptional regulator LasR of *Pseudomonas aeruginosa*: Insights from molecular docking and dynamics simulations

Hovakim Grabski*[1], Lernik Hunanyan[1], Susanna Tiratsuyan[1, 2], Hrachik Vardapetyan[1]

[1]Department of Medical Biochemistry and Biotechnology, Bioinformatics Group, Russian-Armenian University, Yerevan, Armenia
[2]Faculty of Biology, Yerevan State University, Yerevan, Armenia

* To whom correspondence should be addressed.
Email hovakim.grabski@rau.am
Phone: +374 98 73 02 94

E-mail addresses: hovakim.grabski@rau.am, lernik.hunanyan@rau.am, susanna.tiratsuyan@rau.am, hrachik.vardapetyan@rau.am

ABSTRACT

Background: In 2017 World Health Organization announced the list of the most dangerous superbugs and among them is *Pseudomonas aeruginosa*, which is an antibiotic resistant opportunistic human pathogen as well as one of the 'ESKAPE' pathogens. The central problem is that it affects patients suffering from AIDS, cystic fibrosis, cancer, burn victims etc. *P. aeruginosa* creates and inhabits surface-associated biofilms. Biofilms increase resistance to antibiotics and host immune responses, because of those current treatments are not effective. It is imperative to find new antibacterial treatment strategies against *P. aeruginosa*, but detailed molecular properties of the LasR protein are not clearly known to date. In the present study, we tried to analyse the molecular properties of the LasR protein as well as the mode of its interactions with autoinducer (AI) the N-3-oxododecanoyl homoserine lactone (3-O-C12-HSL).

Results: We performed docking and molecular dynamics (MD) simulations of the LasR protein of *P. aeruginosa* with the 3-O-C12-HSL ligand. We assessed the conformational changes of the interaction and analysed the molecular details of the binding of the 3-O-C12-HSL with LasR. A new interaction site of the 3-O-C12-HSL with LasR protein was found, which involves interaction with conservative residues from ligand binding domain (LBD), beta turns in the short linker region (SLR) and DNA binding domain (DBD). It will be referenced as the LBD-SLR-DBD bridge interaction or "the bridge". We have also performed LasR monomer protein docking and found a new form of dimerization.

Conclusions: This study may offer new insights for future experimental studies to detect the interaction of the autoinducer with "the bridge" of LasR protein and a new interaction site for drug design.

Keywords: Pseudomonas aeruginosa, transcriptional regulator, LasR, antibiotic resistant, AMR, ESKAPE, 3-O-C12-HSL, homology modelling, molecular docking, molecular dynamics.

List of abbreviations: PDB, Protein data bank; MD, Molecular Dynamics; PCA, Principal Component Analysis; PC, Principal Component; 3-O-C12-HSL, N-3-oxododecanoyl homoserine lactone; AI, autoinducer; SLR, Short Linker Region; BLAST, Basic local alignment search tool; DBI, David-Bouldin Index; pSF, pseudo-F statistic.

Background

Pseudomonas aeruginosa (*P. aeruginosa*) is a Gram-negative, monoflagellated, obligatory aerobic bacteria [1, 2]. This bacterium is one of the 'ESKAPE' pathogens [3]. It can be found in many diverse environments such as soil, plants, hospitals, etc. [1, 4]. *P. aeruginosa* is an opportunistic human pathogen, because it rarely infects healthy persons. The central problem is that it affects patients suffering from AIDS, cystic fibrosis, cancer and burn victims [2, 5]. Most of the deaths caused by cystic fibrosis are due to this pathogen [1, 6]. The pathogenicity of *P. aeruginosa* occurs due to the synthesis of virulence factors such as proteases, hemolysins, exotoxin A, production of antibiotic pyocyanin, Hydrogen Cyanide (HCN), secretion systems of Types 1 (T1SS), 2 (T2SS), 3 (T3SS), 4[7], 5 (T5SS), and 6 (T6SS) [8], ramnolipids and biofilm formation by this organism [7, 9].

Biofilm formation is characteristic to nearly all bacteria where cell-to-cell communication occurs as the population density increases in the human body during pathology. This system is called quorum sensing (QS) that uses hormone-like molecules called autoinducers (AI) that are accumulated in the extracellular matrix. When a threshold is reached, the AIs bind to its cognate receptor and then a response regulator modulates gene expression of QS genes virulence, which includes adaptation, colonization, antibiotic resistance, plasmid conjugation, etc. Among *P. aeruginosa* the AI type 1 system is represented by LuxI/LuxR-typed proteins. AI-1 diffuses freely through the bacterial membrane and binds to the transcriptional activator LuxR. The latter has two LuxI/LuxR-type systems, the first LasI that produces 3-O-C12-HSL and the second, RhlI that synthesizes C4-HSL; both AHLs regulate virulence and biofilm formation. Thus, in *P. aeruginosa* the operon called hcnABC is responsible for HCN biosynthesis through enzyme HCN synthase [10].

Exposure to HCN can lead to neuronal necrosis through the inhibition of cytochrome c oxidase, the terminal component of the aerobic respiratory chain [11, 12]. P. aeruginosa has four QS systems [13]. Three transcriptional regulators that perform in a cluster [10] control the transcription of hcnABC genes through LasR, ANR and RhlR [14]. The IQS system has been characterized very recently and the genes involved in IQS synthesis are non-ribosomal peptide synthase genes of the ambBCDE operon. The transcriptional regulator of IqsR is the IQS receptor [15].

Nonetheless it was proposed that LasR was the crucial activator of hcnABC genes through mutagenesis experiments [16]. The transcriptional activator protein LasR regulates the target gene expression by recognizing a conserved DNA sequence termed as lux box [16, 17]. LasR has two domains, 1) ligand binding domain at N-terminus (LBD) and 2) DNA binding domain at C-terminus (DBD) [18]. In DBD LasR has a DNA binding Helix-Turn-Helix (HTH) motif [19]. Binding of AI 3-O-C12HSL stabilizes LasR and promotes its dimerization, thereby contributing to the resulting LasR-AI homodimer complex to contact the promoter of the target DNA and activate gene transcription [18].

During colonization and invasion, both the pathogen and the host are exposed to molecules released by each other like bacterial AIs or hosts stress hormones and cytokines. The mechanisms and receptors that might be involved in cross-talk between microbial pathogens and their hosts are not well described to date. LuxR homologues studies have demonstrated that they are homodimers, and consist of two domains. These two functional domains are joined by a short linker region [20, 21].

There remains a need for understanding of the LasR monomer because till date there is no molecular detail information about LasR monomer behaviour. Hence, we analysed the molecular details of the interactions of the 3-O-C12-HSL with LasR protein. So far this is the first report that shows that the 3-O-C12-HSL can interact as well with conservative amino acid residues from ligand binding domain (LBD), beta turns in the short linker region (SLR) and DNA binding domain of LasR. To make easier to reference we refer it as LBD-SLR-DBD bridge or “the bridge”. Could this mean a more nuanced modulation of the transcriptional regulator LasR. This study may be utilized for the development of new therapeutics against *P. aeruginosa* by targeting both the LBD as well as the LBD-SLR-DBD bridge of LasR and inhibit activation of virulence genes.

Methods

Analysis of LasR protein sequence

The methodology was based on a previous study [1]. The Amino acid sequence of the LasR protein of *P. aeruginosa* was obtained from UniprotKB (Uniprot id: P25084). The crystal structure of the LBD of LasR protein (amino acid residues 7 to 169) from *P. aeruginosa* was

acquired from the Protein Data Bank [22] (PDB id: 3IX3). However, the entire structure of LasR protein was needed in order to have a better understanding of LasR protein properties.

For this reason, the amino acid sequence was used to search in the PDB databank using BLAST [23] to identify suitable templates for homology modelling. The crystal structure of QscR, bound to 3-O-C12-HSL was found to be the best fit (PDB id: 3SZT) [20]. The full list of similar structures obtained from BLAST search are shown in Table 1.

Table 1 List of similar structures to LasR obtained from BLAST search.

PDB ID	Sequence identity (%)	Query coverage (%)	E-value	Structure
3SZT	30	97	2×10^{-26}	Quorum Sensing Control Repressor, QscR, Bound to N-3-oxo-dodecanoyl-L-Homoserine Lactone
1FSE	37	28	2×10^{-4}	Crystal structure of Bascillus subtilis regulatory protein gene
4Y15	30	76	3×10^{-16}	Crystal structure of Sdia in complex with 3-oxo-c6-homoserine Lactone
3ULQ	32	25	0.09	Crystal structure of the Anti Activator Rapf complexed with the response regulator coma DNA binding domain

Reconstruction of LasR monomer

HHPred web server [24] server was used for the homology modelling of the full LasR protein. The templates used for the homology modelling of LasR are 3SZT, 3IX3, 1FSE and 3ULQ. The aforementioned templates were used to reconstruct the final model of LasR. The final reconstructed model of LasR protein was verified using PROCHECK [25] (Additional file 1: Figures S1, S2), Verify3D [26] and Ramachandran Plots. More than 96.65% of the amino acid residues in the final reconstructed model had a 3D–2D score > 0.2 as indicated by Verify3D as

shown in Additional file 1: Figure S3 for a good computational model [26]. Ramachandran Plot (Additional file 1: Figure S1) revealed that no amino acid residues were found in the disallowed regions. The PROSA value of this model was -6.69 which suggests that the quality of the homology model is good [27], although the obtained model is slightly different [1], it is still viable.

Constructing 3-O-C12-HSL model

Figure 1 show the 3D model of 3-O-C12-HSL, which was acquired from PubChem [28] to study the interactions of the ligand with LasR.

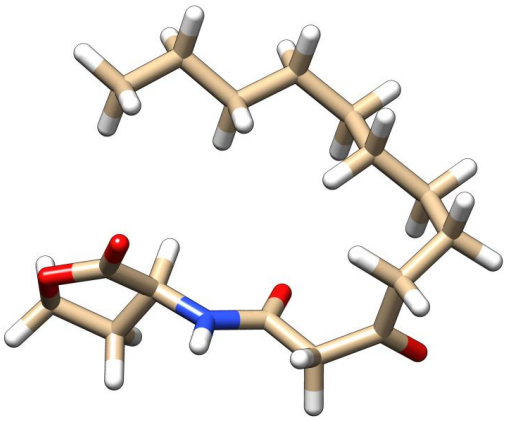

Fig. 1 The stick representation of the native ligand of LasR (3O-C12-HSL).

The ligand parameters for molecular dynamics simulations were calculated for the General Amber Force Field [29] using the acpype tool [30] with AM1-BCC partial charges [31].

Molecular Dynamics Simulations of LasR protein systems

MD simulations of all systems were conducted with the GROMACS suite, version 5.1.2 [32], utilizing the Amber ff99SB-ILDN force field [33]. In all cases, Short-range non-bonded interactions were cut off at 1.4 nm. Particle Mesh Ewald (PME) [34, 35] was used for the calculation of long-range electrostatics.

Set 1. LasR monomer simulation. In order to generate the starting structure of LasR monomer before docking, it was placed in a dodecahedron box of TIP3P water [36], to which 100 mM NaCl was added, including neutralizing counter-ions. Following two steepest descents minimization, LasR monomer was equilibrated in two stages. The first stage involved simulating

for 200 ps under a constant volume (NVT) ensemble. The second stage involved simulating for 200 ps under a constant-pressure (NPT) for maintaining pressure isotropically at 1.0 bar. Production MD simulation was conducted for 100 ns in the absence of any restraints. The temperature was sustained at 300 K using V-rescale [37] algorithm. For isotropic regulation of the pressure, the Parrinello-Rahman barostat [38] was used.

Set 2. Molecular dynamics simulations using docked structures. It has been shown that docking has its limitations [39]. For this reason, after finishing molecular docking simulations we chose the structures of LasR bound to 3-O-C12-HSL to perform molecular dynamics (MD) simulations. A time step of 2 fs was used during heating and the production run and coordinates were recorded every 1 ps. Three simulations of 300 ns were performed. Further details about the simulation protocol can be found above.

LasR–3-O-C12-HSL ligand blind docking experiments

To build the LasR–ligand complex, the ligand 3-O-C12-HSL was docked with LasR monomer using Autodock Vina [40]. However, AutoDock Vina currently uses several simplifications that affect the obtained results. The most notable simplification as the creators' note [40] is the use of a rigid receptor. Vina provides a parameter called 'Exhaustiveness' to change the amount of computational effort used during a docking experiment [40, 41]. But the default exhaustiveness value is 8 and the creators claim that it should increase the time linearly and decrease the probability of not finding the minimum exponentially [40, 41], hence increasing this value leads to an exponential increase in the probability of finding the energy minima.

The whole protein conformational space was searched, using grid box dimensions 60×62× 48 A˚. Following exhaustiveness values were tested in this study: 8, 16, 32, 64, 128, 256, 512, 1024, 2048 and 4096.

Principal component (PC) [42] and cluster analysis using K-means algorithm [43] was performed (Additional file 1: Figures S4, S5). The results demonstrated that the number of interaction sites doesn't change in the interval using exhaustiveness from 1024 to 4096. Exhaustiveness value of 1024 was chosen as it provides good results, good speed and thorough sampling of the docked configurations. Exhaustiveness value was increased to a value of 1024 and a maximum number of binding modes to generate set to 20. After that 100 independent docking calculations were carried out with random initial seeds.

We also used other docking programs for comparison using rDock [44] and FlexAid [45]. We performed 10 docking simulation for each program and generated 20 docked poses.

Analysis of docking conformations and trajectories

We performed the analysis of docking conformations and trajectories by a combination of Autodock Vina [40], Gromacs [32] in addition to custom python scripts, which uses pandas [46], scikit-learn [47] and MDTraj [48].

Here is the list of program packages we used for analysis:

- The MDTraj python library [48] was used for the trajectory analysis and LigPlot+ for 2D diagrams of 3-O-C12-HSL–LasR interactions [49].
- Plot visualization was done using matplotlib [50] and Seaborn package [51].
- Figures and videos were prepared with PyMOL [52], VMD [53] and UCSF Chimera [54].

The analysis protocol approach was similar to the works of Wolf et al. [55] and Zamora et al. [56] and the following techniques were used for the analysis:

- Principal component analysis (PCA) is an unsupervised statistical technique that is often used to make data easy to explore and visualize. PCA tries to explain the maximum amount of variance with the fewest number of principal components [42]. The process of applying PCA to a protein trajectory is called Essential Dynamics (ED) [56-58]. PC analysis, performed on Cartesian coordinates has become an important tool for the examination of conformational changes.
- Cluster analysis is another unsupervised technique that tries to identify structures within the data. It is a data exploration tool for dividing a multivariate dataset into groups. Clustering algorithms can be grouped into partitional and hierarchical clustering approaches [59].

NMR calculations

Finally, the trajectory of LasR monomer simulation was used for the calculation of chemical shifts. Sparta+ [60] and ShiftX2 [61] were used to predict the chemical shifts of backbone atoms of the protein with the help of wrapper functions of MDTraj [48].

Binding energy calculation.

g_mmpbsa [62] was used for the evaluation of binding energies. It was also used for the estimation of the energy contribution of each residue to the binding energy. g_mmpbsa calculates the relative binding energy rather than the absolute binding energy, so the entropy contribution (TΔS) is not evaluated [62]. We cut out from the 300ns simulations production trajectories of 10ns with the interval of every 120 ps for the calculation of all energy components of binding energies.

Sequence Conservation

To find out whether 3-O-C12-HSLs interact with conservative amino acid residues we performed multiple sequence alignment (MSA) using msa package [63]. ClustalW [64], Clustal Omega [65] and Muscle [66] within msa package were used for multiple sequence alignments. We performed multiple sequence alignment of the LasR protein between the closely related species of *P.Aeruginosa*.

The following sequences were used for sequence alignment:

- WP_054058449.1 LuxR family transcriptional regulator [*Pseudomonas fuscovaginae*]
- NP_250121 transcriptional regulator LasR [*Pseudomonas aeruginosa* PAO1]
- KFC75736.1 LuxR-type transcriptional regulator [Massilia sp. LC238]
- WP_018433960 LuxR family transcriptional regulator [Burkholderia sp. JPY251]
- WP_042326260.1 LuxR family transcriptional regulator [Paraburkholderia ginsengisoli]
- WP_012426170.1 LuxR family transcriptional regulator [Paraburkholderia phytofirmans]
- WP_027776298.1 LuxR family transcriptional regulator [Paraburkholderia caledonica]
- WP_027214716.1 LuxR family transcriptional regulator [Burkholderia sp. WSM2232]
- WP_041729325.1 LuxR family transcriptional regulator [Burkholderia sp. CCGE1003]
- CBI71275.1 UnaR protein [Paraburkholderia unamae]
- CAP91064.1 BraR protein [Paraburkholderia kururiensis]
- WP_003082999.1 transcriptional activator protein LasR [*Pseudomonas*]
- WP_012076422.1 MULTISPECIES: LuxR family transcriptional regulator [*Pseudomonas*]

- WP_050376898.1 MULTISPECIES: LuxR family transcriptional regulator [*Pseudomonas*]
- WP_050395760.1 LuxR family transcriptional regulator [Pseudomonas aeruginosa]

Protein-Protein Docking

When docking homology models, it is best if there is an experimental evidence to suggest the general interaction site (within ~10 Å). Representative structures from molecular dynamics simulations were used for protein-protein docking using ClusPro [67]. From the experimental X-ray data, it was found that 'Trp152', 'Lys155' and 'Asp156' play an important role during dimerization.

The distances between 'Trp152' from chain A and 'Asp156' from chain B was 0.276 nm, 'Asp156' from chain A and 'Lys155' from chain B of the crystallographic structure was 0.279nm. These residues were used as attraction constraints.

Entire flowchart

The whole methodology is presented as a flowchart for a better comprehension:

- LasR protein sequence taken from UniprotKB (sequence ID P25084) and performed blastp of LasR to identify suitable templates for reconstruction.
- HH-pred web server was used for the reconstruction of LasR monomer structure and validated using Verify 3D, Procheck, Prosa.
- The 3D model of 3-O-C12-HSL acquired from PubChem web server.
- Docking of 3-O-C12-HSL ligand with the LasR monomer performed using Autodock Vina, later we corroborated with rDock, FlexAid.
- PCA and cluster analysis performed on docking conformations.
- Extraction of centroid conformations from cluster analysis.
- Ligand parameters generated using Acpype interface in the framework of the AMBER force field.
- Using centroid conformations as starting points for molecular dynamics simulations using Gromacs.
- Analysis of molecular dynamics trajectory files using MDTraj.

- Relative binding energy calculation and contribution of residues to binding using g_mmpbsa.
- Sequence alignment using msa package with ClustalW, Clustal Omega and Muscle algorithms.
- Performed protein-protein docking using ClusPro.

Results and discussion

Conformational changes of LasR without 3-O-C12-HSL

We performed a simulation run of 100 ns using a standard MD protocol for the assessment of conformational changes of LasR monomer (Additional file 2). The overall stability of the molecule was assessed using the root-mean-square-deviation (RMSD) of the protein atoms. RMSD was calculated with reference to the initial frame through the time of MD run (Additional file 1: Figure S6). Another suitable measurement for the stability of the LasR monomer structure is the radius of gyration (Additional file 1: Figure S7).

During the examination of MD trajectories, Principal Component (PC) analysis [42] is usually used for the visualization of the motions of the system (Figure 2). Generally to capture more than 70% of the variance in the trajectory data the first handful of components are sufficient [56]. PC analysis can uncover the fundamental movements contained in an MD trajectory, however, it does not group the snapshots into different clusters [59]. This can be accomplished by the clusterization of the PC data.

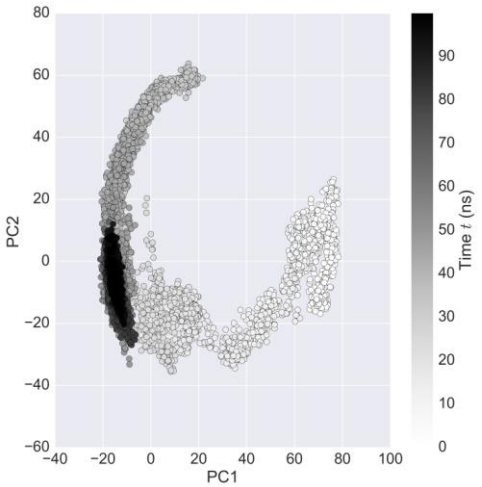

Fig. 2 LasR monomer simulation trajectory projected onto its first two principal components. The black scale indicates the time progression from 0 ns (white) to 100 ns (black).

For the selection of initial starting structure of the LasR for docking study, we performed cluster analysis on the LasR monomer run for the selection of a starting point. By identifying a distinct representative structure of the most populated cluster, this will allow performing blind docking on the whole structure. It should be also noted that cluster analysis allows evaluating the frequent conformations of LasR. For the clustering analysis, we chose the Agglomerative Clustering algorithm with Ward linkage [68] as the most appropriate as they are deterministic, allowing for reproducibility of resulting clusters, thus minimizing the amount of bias.

We performed several rounds of Agglomerative clustering with Ward linkage (details can be found in the methods section). The accuracy of the clustering was assessed by the help of the Davies–Bouldin Index (DBI) [69], Dunn Index [70], Silhouette Score [71], and the pseudo-F statistic (pSF or Calinski Harabasz) [72] metrics (Additional file 1: Figure S8). An optimal number of clusters were chosen, simultaneously accounting for a low DBI, high Silhouette, high Dunn Index and high pSF values.

The distribution of clusters over simulation is visualized in Figure 3 and the four cases are cluster 4 (black) at the beginning of the simulation (after equilibration), cluster 2 (green) in the middle, cluster 1 (light green) and cluster 3 (dark blue) at its end.

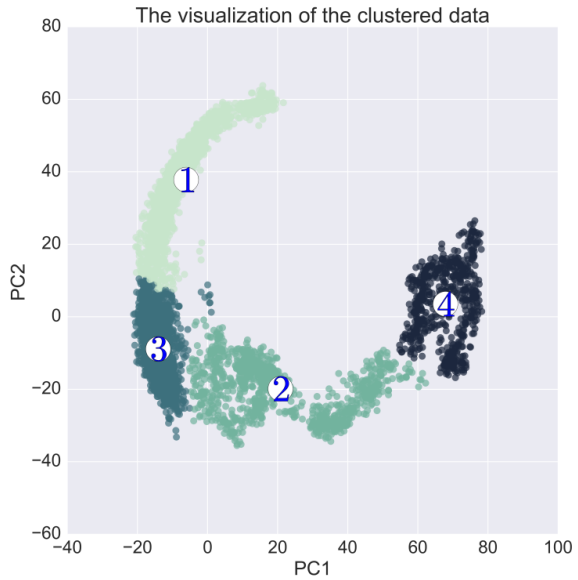

Fig. 3 Clustering results of ward-linkage algorithm formed by first two PCs. The entire MD trajectory data was used for the clusterization.

The clusterization defined by the first two PCs (Figure 4) provides a coherent picture and it also supported by a good DBI, Dunn, Silhouette Score and pSF value (Additional file 1: Figure S8). Figure 4 shows clearly that the simulation has converged.

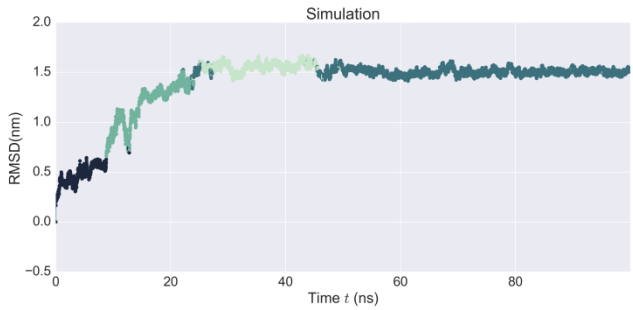

Fig. 4 Colour-coded RMSD of the simulation obtained from Ward-linkage cluster analysis. The colours represent the clusters that are demonstrated in Figure 4.

For the validation of the quality of molecular dynamic simulation, theoretical NMR shifts were calculated using Sparta+ [60] and ShiftX2 [61]. For the NMR shifts calculation, snapshots from cluster 3 (Figure 3 and Figure 4) were used.

In Figure 5 you can see that there is a strong linear relationship between experimental and simulated ShiftX2 NMR shift values (Additional file 1: Figure S9), thus this demonstrates the quality of MD simulation.

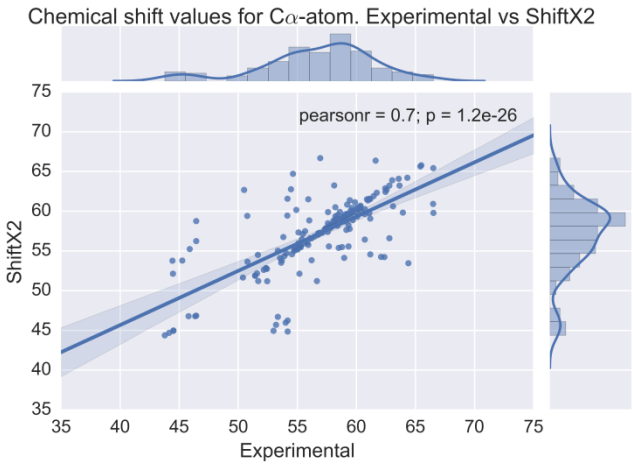

Fig. 5 Chemical shifts of Cα- atoms. Experimental vs ShiftX2 predictions.

Figure 6 shows that there is a strong linear relationship as well between experimental and simulated Sparta+ NMR shift values (Additional file 1: Figure S9).

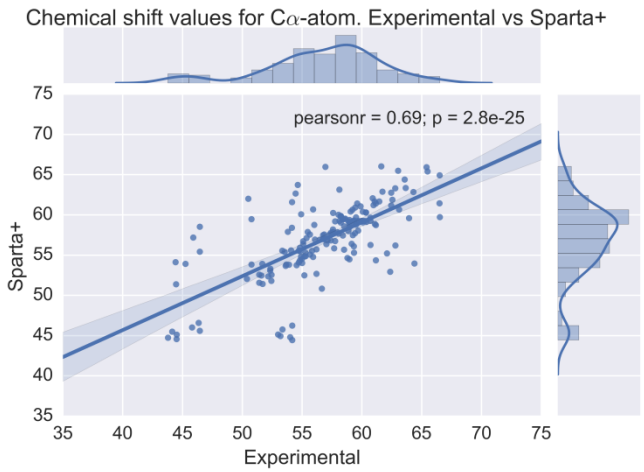

Fig. 6 Chemical shifts of Cα- atoms. Experimental vs Sparta+ predictions.

After that, for docking study, the representative structure (Figure 7c) was extracted from cluster 3 (Figure 3 and Figure 4).

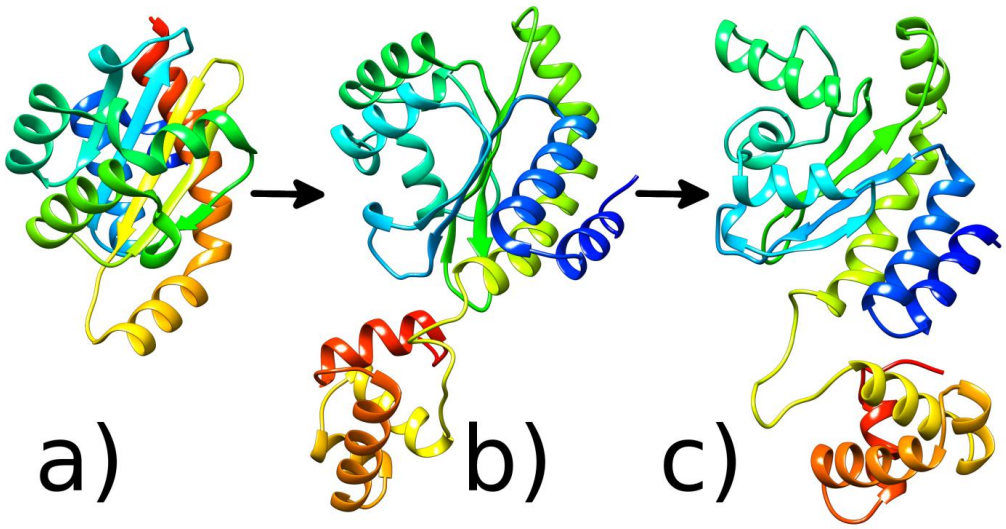

Fig. 7 3D LasR protein structures: a) Crystal structure of the N-terminal LBD of LasR protein b) Homology Modelling c) Representative Structure after 100ns MD run. Images generated with UCSF Chimera [49].

The representative structure of the LasR system and its differences from the homology modelled structure are highlighted in Figure 8c. Secondary structure analysis of the representative structure was performed as well using PDBSum [73] (Additional file 1: Figure S10). To our opinion, this is a complete study to include the dynamics of the full-length LasR

molecule (residues 1 to 239). It has also been shown that the dynamics of the complete C-terminal region of LasR modulate N-terminal region, as will be discussed later.

Docking analysis of 3-O-C12-HSL with LasR monomer

In this study, molecular docking approach was used to inspect the possible binding modes of 3-O-C12-HSL in LasR monomer. PCA and cluster analysis were performed on docking data (Figure 8). The results show three binding sites, cluster 2 corresponds to experimental data, while cluster 1 and cluster 3 do not (Figure 8). These results clearly support the findings of Bottomley et al. [18], where it was also demonstrated that 3-O-C12-HSL binds to LBD.

We performed several rounds of K-Means clustering (details are available in the section of methods). The accuracy of the cluster analysis was evaluated using the DBI [69], Dunn Index [70], Silhouette score [71] and the pSF [72] metrics (Additional file 1: Figure S11). An optimal number of clusters were chosen for docking results, simultaneously accounting for a low DBI, high Dunn, high Silhouette and high pSF values.

We generated 2000 docked poses and performed representative structure extraction for use in MD simulations of the LasR 3-O-C12-HSL binding sites. The resulting representative structures from each cluster are shown in Figure 9. These cluster representative structures were produced by finding the centroid conformations.

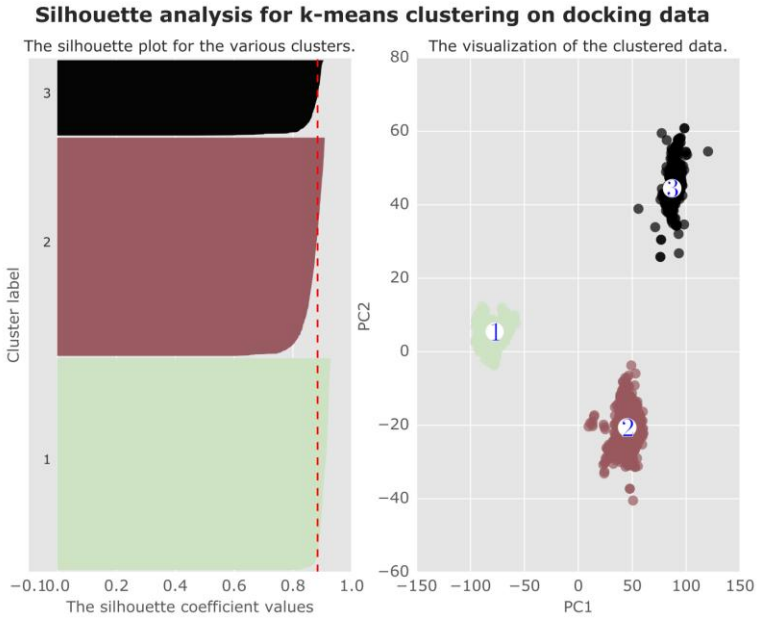

Fig. 8 Silhouette plot (left) and clustering results (right) on the docking data using k-means algorithm formed by first two PCs.

Representative structures from each cluster were extracted. The binding energy for the representative structure of cluster 1 is -5.4 kcal/mol, as for the mean binding affinity for the whole cluster is -5.257 ±0.233 kcal/mol (Additional file 1: Figure S12). Cluster 1 contains 839 docked poses from 2000, about 41.95%. For cluster 2 the binding affinity for the representative structure is -5.1 kcal/mol and for the whole cluster -5.593 ±0.386 kcal/mol (Additional file 1: Figure S13) and this one corresponds to the experimental binding site data [18]. Cluster 2 contains 864 docked poses from 2000, about 43.2%. For cluster 3, the representative structure features the highest binding affinity -5.7 kcal/mol and for the whole cluster -5.264 ±0.27 kcal/mol (Additional file 1: Figure S14). Cluster 3 contains 297 docked poses from 2000, about 14.85%, which is a rather unstudied area.

There is a huge amount of literature that suggests that molecular docking is not appropriate for the prediction of binding affinity or binding poses of protein-ligand complexes, however, they can still provide important information [39, 74].

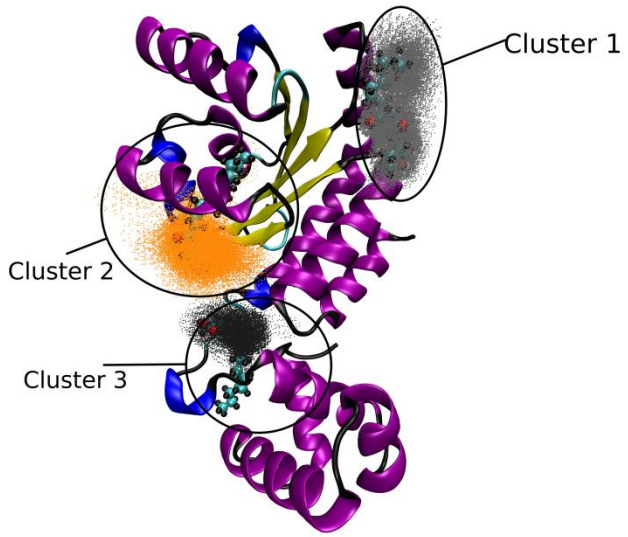

Fig. 9 3D visualization of the analysed docking data with their representative structures and clusters.

Later we corroborated the blind docking with other molecular docking software as well, which includes Autodock Vina [40], rDock [44] and FlexAid [45] (Figure 10). PC analysis of the

various docking programs was performed for easier comprehension (Additional file 1: Figure S15).

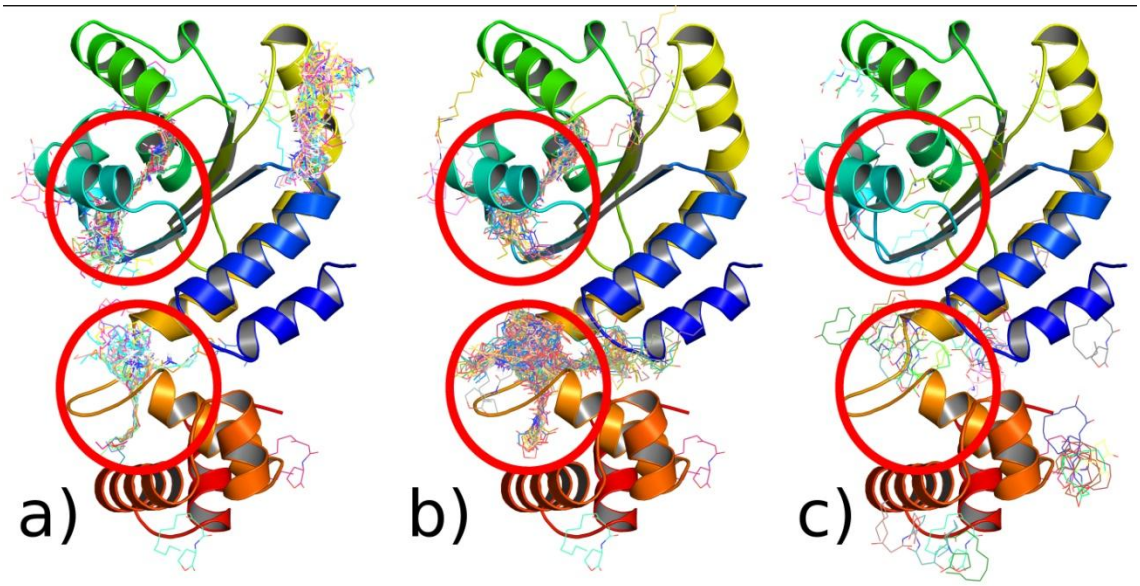

Fig. 10 Blind docking with various molecular docking programs. Red circles - binding interactions that are characteristic to all programs a) Autodock Vina b) rDock c) FlexAid.

Binding Modes of 3-O-C12-HSL

We performed three 300 ns simulations using standard MD protocol. Overall, 900ns of simulation data was used for the analysis of 3-O-C12-HSL interaction with LasR monomer. The representative structures were taken from docking results (Figure 9) and used as starting points for MD simulation with LasR. Simulations were conducted for sufficient time to allow the positions of 3-O-C12-HSL to reach equilibrium in LasR molecule.

The overall stability of the molecule was assessed using the mass-weighted RMSD of the backbone atoms. RMSD was calculated with reference to the initial snapshot for the different independent MD runs. Figure 11 shows that Simulation 2 and 3 experience a substantial RMSD deviation from the initial starting point. Simulation 2 corresponds to cluster 2 in docking simulations, while Simulation 3 corresponds to cluster 3 (Figure 11). Simulation 1 which corresponds to cluster 1, the molecule of 3-O-C12-HSL did not fixate and reach equilibrium, so further research was not performed (Figure 11) (Additional file 3). Simulation 2 (Additional file 4) shows that after 230ns, the structure becomes stable. While Simulation 3 (Additional file 5) changes its conformation in 100ns and becomes stable between 260ns and 300ns.

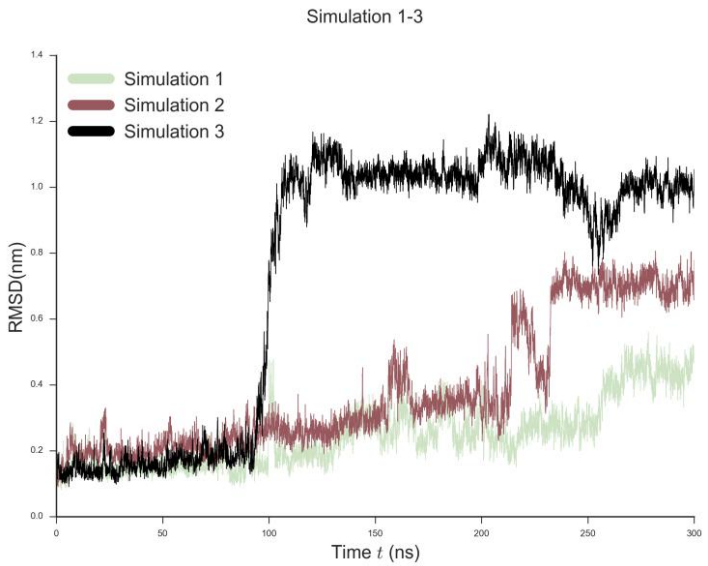

Fig. 11 RMSD evolution during 300 ns of simulation for 3 independent runs using representative structures of 3-O-C12-HSL docking poses as starting points.

The root-mean-square fluctuation (RMSF) was used for the assessment of the flexibility of the LasR monomer. In Figure 12 the average per-residue RMSF for each cluster simulation runs is plotted.

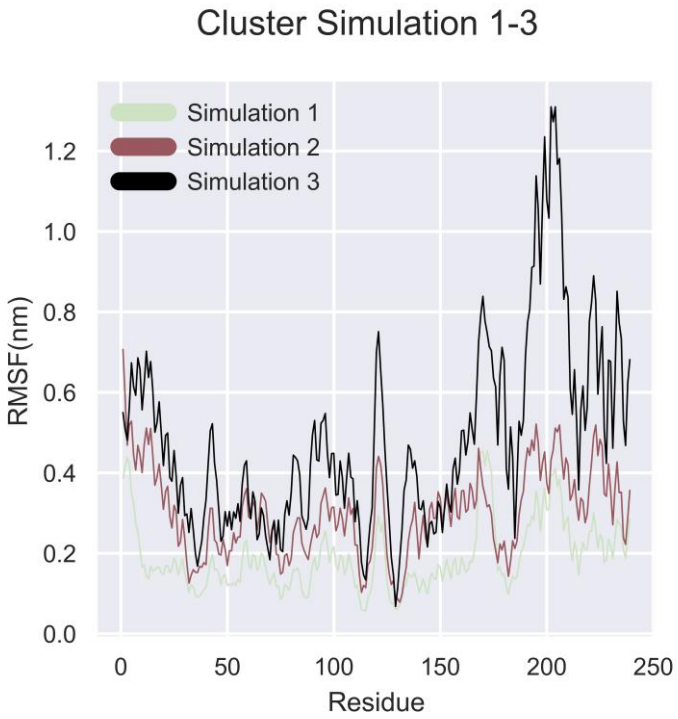

Fig. 12 RMSF of Cα- atoms of LasR for each individual run.

From Figure 12 it visible that the residues from 165 to 176, which correspond to beta turns in the SLR of LasR, are of high mobility (Additional file 1: Figure S10). Simulations 2 and 3 show that 3-O-C12-HSL has two binding modes, one with LBD, which corresponds to experimental data and simulation 3 with the LBD-SLR-DBD bridge of LasR.

PCA and Cluster analysis were performed on simulation 2 (Additional file 1: Figure S16) as well as hydrogen bond analysis based on cutoffs distance and angle according to the criterion of Wernet-Nilson [75] using MDTraj [48]. Over the course of cluster 1, 3-O-C12-HSL with LasR established an average of 0.655±0.651 of hydrogen bonds, while in cluster 2 the average is 0.042±0.202 (Additional file 1: Figure S16). Over the course of simulation 2, 3-O-C12-HSL establishes a large number of hydrophobic contacts with amino acid side chains of the LBD of LasR protein (Additional file 1: Figure S18), a phenomenon that is not unexpected, given the large hydrophobic surface area of the LasR LBD and the low solubility of 3-O-C12-HSL in water. In simulation 2, 3-O-C12-HSL has hydrophobic interactions mainly with amino acids from H3, H5 and S1 (Additional file 6, 7, 8). RMSD analysis of the conformations between LasR with/without 3-O-C12-HSL in LBD is equal to 7.027 Å (Figure 13).

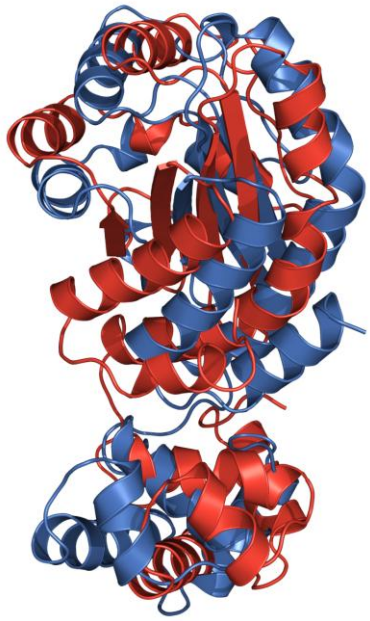

Fig. 13 LasR monomer structures: without 3-O-C12-HSL (blue colour), with 3-O-C12-HSL bound to LBD (red colour).

Residues that participate in hydrophobic interactions are shown in Figure 14 and Additional file 1: Figure S18.

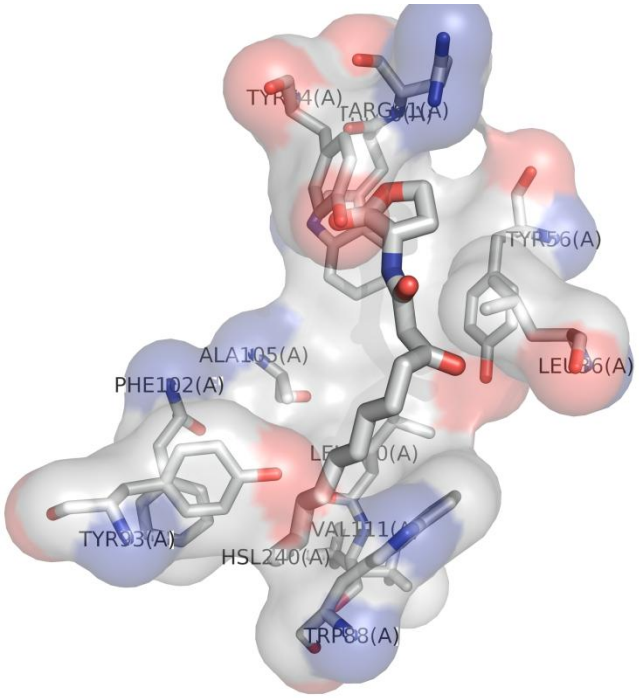

Fig. 14 Insertion of 3-O-C12-HSL in LBD of LasR, from centroid snapshot of cluster 2 in simulation 2 (Additional file 1: Figure S16).

The second binding mode involves the interaction of 3-O-C12-HSL with the LBD-SLR-DBD bridge. PCA, cluster, and hydrogen bond analysis were also performed on simulation 3 (Additional file 1: Figure S19). Over the course of cluster 2, 3-O-C12-HSL with LasR established an average of 1.506±0.742 hydrogen bonds, for cluster 3 the average 0.228±0.492 (Additional file 1: Figure S19), while in cluster 1 the average is 0.652±0.654. Over the course of simulation 3, 3-O-C12-HSL establishes hydrogen bonds and a large number of hydrophobic contacts with amino acid side chains in the beta turns in the SLR of LasR protein (Additional file 1: Figure S21). In simulation 3, 3-O-C12-HSL forms hydrogen bonds mainly with 'Lys182' and 'Leu177' of the beta turns in the SLR of LasR (Additional file 1: Figure S21). RMSD analysis of the conformations between LasR monomer and LasR bound to 3-O-C12-HSL in DBD is equal to 1.677 Å (Figure 15).

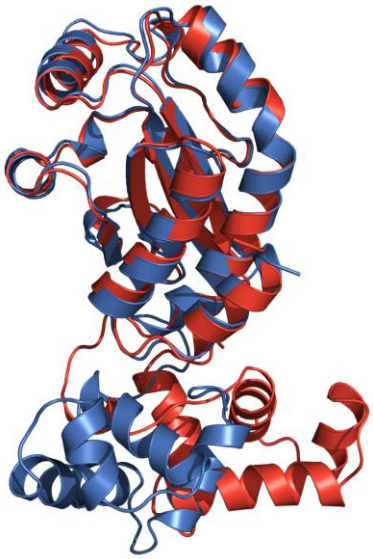

Fig. 15 LasR monomer structures: without 3-O-C12-HSL (blue colour), with 3-O-C12-HSL bound to LBD-SLR-DBD bridge (red colour).

Residues that participate in hydrogen and hydrophobic interactions are shown in Figure 16.

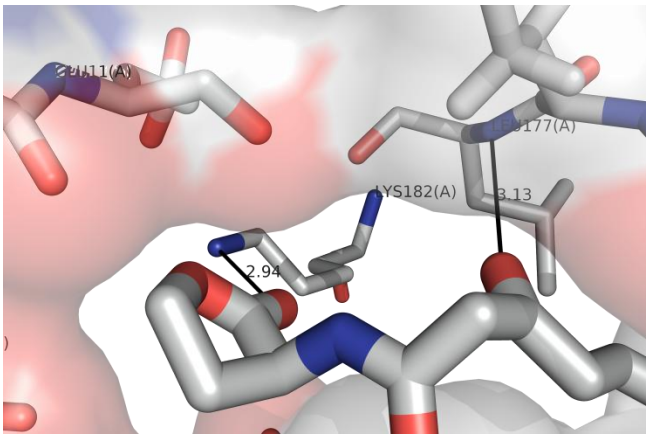

Fig. 16 Putative hydrogen bonds of 3-O-C12-HSL with Lys182 and Leu177 from centroid snapshot of cluster 1 of simulation 3 (Additional file 1: Figure S19).

Complete data sets for MD simulations are available as Supporting Information (Additional file 1: Figures S16-S21).

Binding Energy of 3-O-C12 to LasR and Sequence Conservation

In order to analyse the binding sites in detail, MMPBSA [62] binding energy calculation was performed for each binding site based on the trajectories. The binding energy calculation demonstrated a very interesting results and that is that the interaction with LBD-SLR-DBD bridge is higher than with the LBD (Table 2). More detailed analysis on the energy terms showed that Van der Waals, electrostatic interactions and non-polar solvation energy contribute negatively to the binding energy while polar solvation energy contributes positively. And van der Waals interaction contributes most in the terms of negative contribution for both cases, but for the interaction with the LBD-SLR-DBD bridge or "the bridge" the electrostatic interaction is 3.6 higher than with the LBD interaction.

Table 2 Relative Binding energy using g_mmpbsa on simulation data.

Sim №	**Binding sites**	**Van der waal (kJ/mol)**	**Electrostatic (kJ/mol)**	**Polar salvation (kJ/mol)**	**Non-polar salvation (kJ/mol)**	**Binding energy (kJ/mol)**
2	**LBD**	**-215.673 ±8.007**	**-39.586 ±18.817**	**147.522 ±13.978**	**-20.840 ±0.794**	**-128.578 ±18.757**
3	**LBD-SLR-DBD bridge**	**-205.141 ±11.803**	**-142.974 ±29.032**	**201.889 ±20.117**	**-20.230 ±0.842**	**-166.456 ±20.492**

This suggests that van der Waals, electrostatic interaction and non-polar solvation energy together contribute to the stability of LasR-3-O-C12-HSL binding complex. To asses with which conserved residues in But to find out we needed to find out which amino acid residues were involved with the interaction. We performed analysis of Energy contribution of residues to binding for both simulations obtained from MMPBSA calculation using g_mmpbsa. Sequence alignment was performed using msa package [63]. ClustalW [64], Clustal Omega[65] and Muscle[67] algorithms were used for sequence alignment. Full alignments using the aforementioned algorithms are available as supplementary materials (Additional file 6, 7, 8). The

binding state with LBD with binding energy ~ -128.578 kJ/mol (Table 2), which the residues Tyr64, Tyr56, Trp88, Leu36 and Trp60 contribute most to (Figure 17).

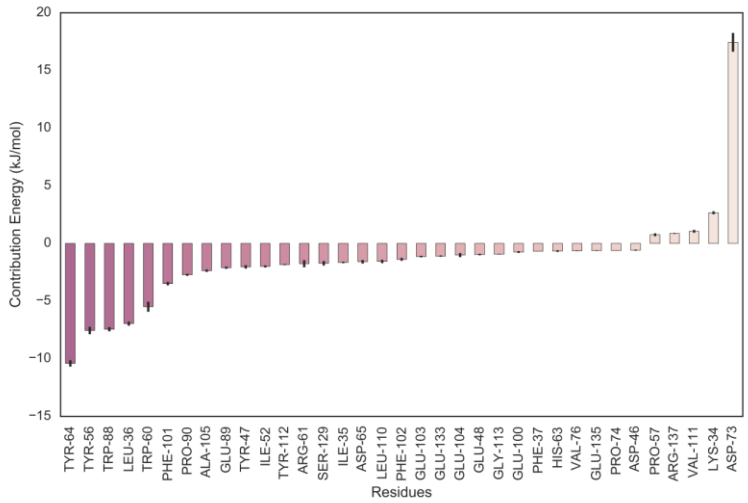

Fig. 17 Energy contribution of LBD amino acid residues from simulation 2.

Residue interaction and sequence alignment show that 3-O-C12-HSL interacts with very conservative amino acid that includes Trp60 which corresponds to experimental data [76]. It also involves other conserved amino acids such as Tyr64, Asp73, Pro74, Val76, Phe102, Phe103, Ala105, and Gly113 (Fig. 18). 3-O-C12-HSL interacts with 10 fully conserved amino acids. In total 16 amino acids participate in the interaction, where amino acid conservation is more than 75%.

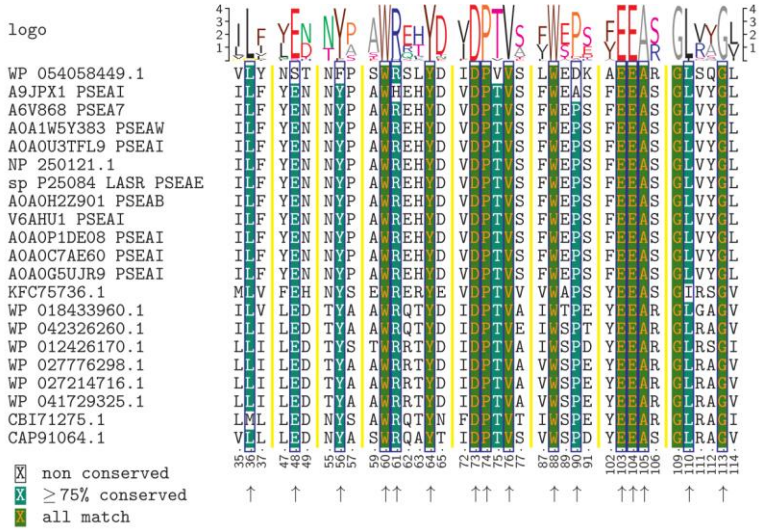

Fig. 18 Interaction of conserved amino acids of LBD of LasR with the autoinducer molecule. Blue boxes and arrows point the amino acid residues that interact with 3-O-C12-HSL.

The new binding state with LBD-SLR-DBD bridge or “the bridge” of LasR with energy ~ -166.456 kJ/mol (Table 2), where residues Leu236, Leu177, Val176, Phe219, Lys182 and Trp19 contribute most (Figure 19).

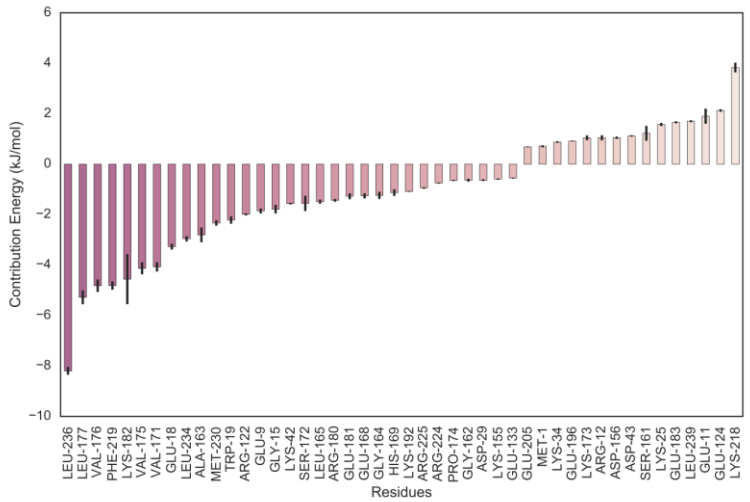

Fig. 19 Energy contribution of LBD-SLR-DBD bridge amino acid residues from simulation 3.

Analysis of residue sequence alignment show that 3-O-C12-HSL interacts with very conservative amino acids such as Leu236, Leu177, Phe219, Trp19 (Fig. 19, 20). 3-O-C12-HSL interacts with 12 fully conserved amino acids. In total 16 amino acids participate in the interaction, where amino acid conservation is more than 75%.

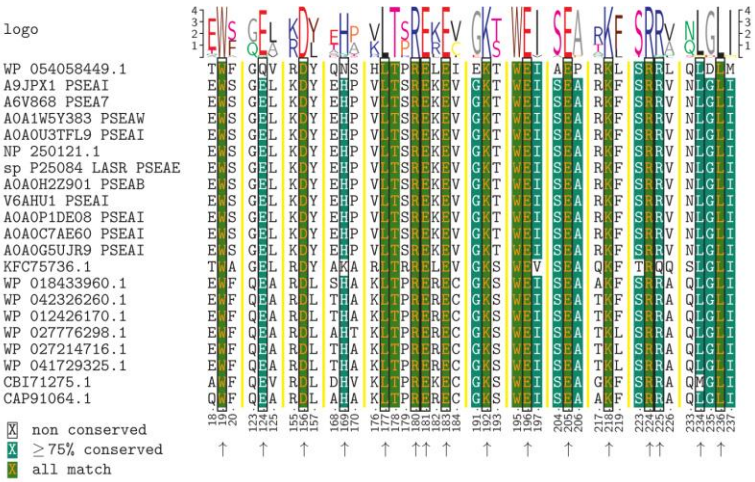

Fig. 20 Interaction of conserved amino acids of LBD-SLR-DBD bridge (“the bridge”) with the autoinducer molecule. Black boxes and arrows point the residues that interact with 3-O-C12-HSL.

In this new binding site Trp19, Asp156 from LBD, Leu177 from SLR and Arg180, Glu181, Glu196, Lys218, Arg224, Lys236 of DBD participate in the interactions with LasR. Lactone head group interacts with conserved Trp19 of LBD. This result clearly suggests that both of the C-terminal and the N- terminal of LasR interact with 3-O-C12-HSL.

Protein Docking of MD models

It is known that LasR binds to the corresponding promoter DNA region as a dimer [1] after interacting with the autoinducer. Thus, the monomeric LasR–3-O-C12-HSL complexes were subjected to dimerization. The protein docking experiments using structures from MD runs were performed with ClusPro 2.0 [67, 77–79] because of its success in the CAPRI (Critical Assessment of Predicted Interaction). Each centroid conformation was extracted from simulations 2 and 3 and was used for docking. For the selection of the model, we used the approach as recommended by the authors of ClusPro [67], suggesting to find the most populated clusters.

From simulation 2 the top model contains 122 members and the scores for the docking model were -1440.7 for the centre and -1517.9 for the lowest energy, suggesting a favourable binding mode (Figure 17).

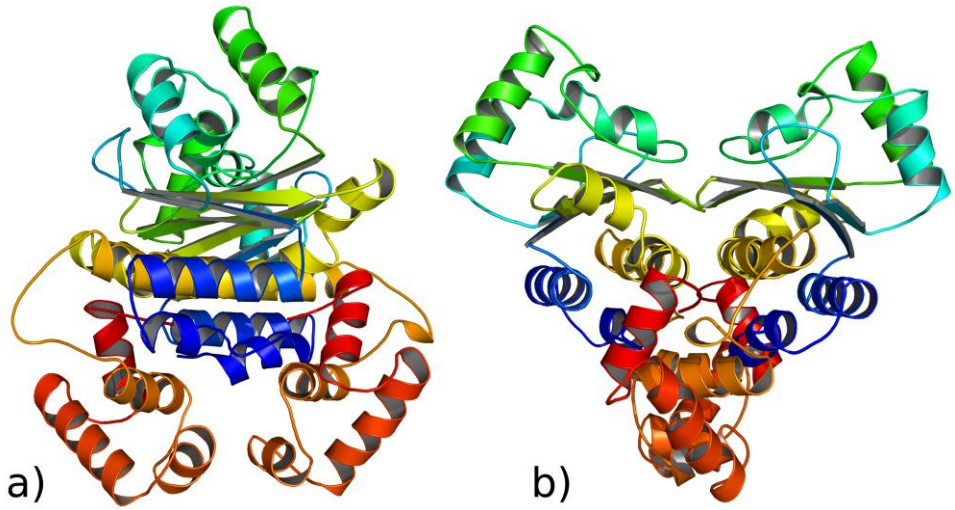

Fig.17 *LasR monomer-LasR monomer docking result based on centroid conformation from cluster 2 of simulation 2 (3-O-C12-HSL with LBD) a) left view b) front view.*

For simulation 3 docking the top model contains 80 Members and the scores for the docking model were -951.4 for the Center and -1332.0 and for the lowest energy, thus suggesting a favourable binding mode (Figure 18).

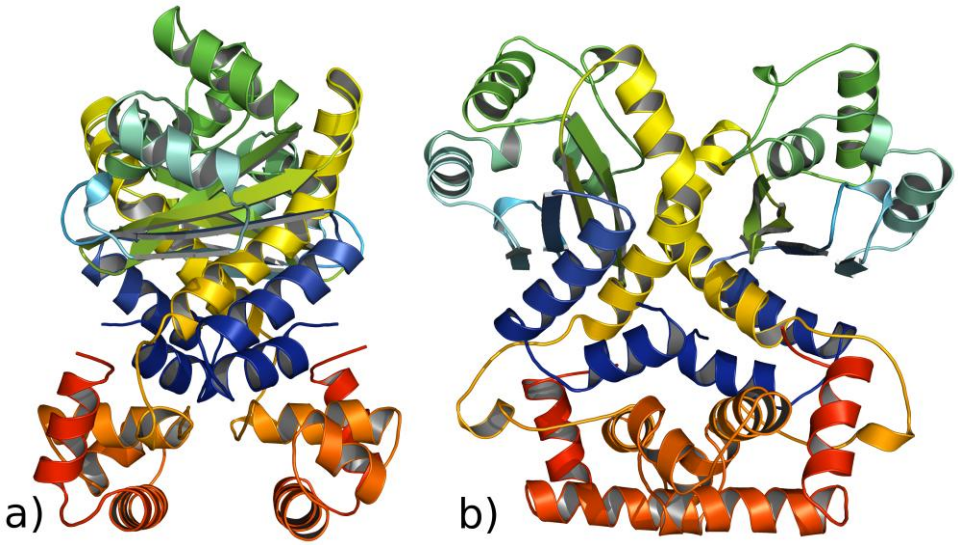

Fig. 18 *LasR monomer-LasR monomer docking result based on centroid conformation from cluster 1 of simulation 3 (3-O-C12-HSL with the "bridge") a) left view b) front view.*

The interaction of 3-O-C12-HSL with the "bridge" creates a possible conformation for dimerization. From our analysis, it is observed that 3-O-C12-HSL has two binding modes to LasR protein. The 3-O-C12-HSL, AI ligand (Figure 1) having a long hydrophobic tail is capable of binding both to LBD and to the "bridge" of LasR . The binding of 3-O-C12-HSL provokes conformational transitions of the transcriptional regulator LasR. The two binding conformations are capable of dimerization.

Conclusion

From the simulations, it can be safely concluded that the AI ligand 3-O-C12-HSL, can bind both to LBD and to "bridge" of transcriptional regulator LasR. This suggests that there are multiple binding modes rather than one. The interaction with the LBD-SLR-DBD bridge is a novel site. The analysis of binding energy shows that the interaction of 3-O-C12-HSL with LDB-SLR-DBD bridge is stronger than with LBD. Conservative amino acids such as Leu236, Phe219, Leu177, Lys182 and Trp19, contribute most during the interaction with LBD-SLR-DBD bridge. This could suggest that for the DNA binding capability, it is necessary the interaction of 3-O-C12-HSL with the "bridge".

Both conformational transitions favour dimerization, so this raises the question what would be the roles of different binding sites. This part needs further research. This study may reveal new insights of the interactions of the native 3-O-C12-HSL ligand with transcriptional

regulator LasR of P. aeruginosa. Results from this study may be used for future drug development endeavours.

Additional files

Additional file 1: Figure S1. PROCHECK plot of LasR model. **Figure S2.** PROCHECK statistics of LasR model. **Figure S3.** Verify3D score plot of LasR model. **Figure S4.** Determination of exhaustiveness after PCA and cluster analysis**. Figure S5.** Exhaustiveness clusterization quality analysis. **Figure S6.** RMSD evolution through time for LasR monomer structure. The RMSD is calculated for the backbone atoms of the residues**. Figure S7.** Rg evolution through time for LasR monomer structure. The Rg is calculated for the backbone atoms of the residues. **Figure S8.** Agglomerative clustering with a different number of clusters for the LasR monomer run. In this case, a cluster count of 4 is the best option. **Figure S9.** The plot of chemical shifts of Cα- atom vs. residue number. **Figure S10.** Secondary structure analysis of LasR protein from simulation representative structure using PDBSum. **Figure S11.** K-means clustering with a different number of clusters for the 3-O-C12-HSL docking data. In this case, a cluster count of 3 is the best option. **Figure S12.** Mean Autodock Vina binding energy score for cluster 1 and percentage from the total of conformations. **Figure S13.** Mean Autodock Vina binding energy score for cluster 2 and percentage from the total of conformations. **Figure S14.** Mean Autodock Vina binding energy score for cluster 3 and percentage from the total of conformations. **Figure S15.** Principal component analysis of ligand (3-O-C12-HSL) conformations obtained from various docking programs. **Figure S16.** PCA and RMSD analysis for simulation 2 of docking cluster 2. Left) PCA analysis Right) Colour-coded RMSD of the simulation obtained from Ward-linkage cluster analysis. **Figure S17.** The frequency of hydrogen bonds of cluster 2 in simulation 2 **Figure S18.** Hydrogen and hydrophobic interactions from centroid snapshot of cluster 2 in simulation 2 using LigPlot+. **Figure S19.** PCA and RMSD analysis for simulation 3 of docking cluster 3. Left) PCA analysis Right) Colour-coded RMSD of the simulation obtained from Ward-linkage cluster analysis. **Figure S20.** The frequency of hydrogen bonds of cluster 1 in simulation 3. **Figure S21.** Hydrogen and hydrophobic interactions from centroid snapshot of cluster 1 in simulation 3 using LigPlot+. (DOCX)
Additional file 2: Movie of LasR monomer MD simulation. (MP4)
Additional file 3: Movie of MD simulation 1. (MP4)
Additional file 4: Movie of MD simulation 2. (MP4)
Additional file 5: Movie of MD simulation 3. (MP4)
Additional file 6: Full sequence alignment of LasR using ClustalW algorithm. (PDF)
Additional file 7: Full sequence alignment of LasR using Clustal Omega algorithm. (PDF)
Additional file 8: Full sequence alignment of LasR using Muscle algorithm. (PDF)

Declarations

Acknowledgments

We thank Yerevan Physics Institute for providing time on the cluster for the molecular dynamics simulations.

Funding

This work was supported by grant NIR 25/15 of the Russian-Armenian University.

Availability of data and materials

The analysed datasets are available as additional files.

Authors' contributions

HG designed the construct and simulations and performed the analyses. HG ST wrote the manuscript. HV ST LH supervised this work. All authors have reviewed and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

References

1. Chowdhury N, Bagchi A. Molecular insight into the activity of LasR protein from Pseudomonas aeruginosa in the regulation of virulence gene expression by this organism. Gene. 2016 Apr 10;580(1):80-7.

2. Botzenhart K, Döring G. Ecology and epidemiology of Pseudomonas aeruginosa. In: Pseudomonas aeruginosa as an Opportunistic Pathogen. Springer US. 1993. p1-18.

3. Pendleton JN, Gorman SP, Gilmore BF. Clinical relevance of the ESKAPE pathogens. Expert Review of Anti-infective Therapy [Internet]. Informa UK Limited; 2013 Mar;11(3):297–308. Available from: http://dx.doi.org/10.1586/eri.13.12

4. Lederberg J. Infectious history. Science. 2000 Apr 14;288(5464):287-93.

5. Irvin RT. Attachment and colonization of Pseudomonas aeruginosa: role of the surface structures. In: Pseudomonas aeruginosa as an Opportunistic Pathogen. Springer US; 1993. p. 19-42.

6. Govan JR, Deretic V. Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. Microbiological reviews. 1996 Sep 1;60(3):539-74.

7. Coggan, K. A., & Wolfgang, M. C. (2012). Global regulatory pathways and cross-talk control Pseudomonas aeruginosa environmental lifestyle and virulence phenotype. Current issues in molecular biology, 14(2), 47.

8. Filloux A. Protein secretion systems in Pseudomonas aeruginosa: an essay on diversity, evolution, and function. Frontiers in microbiology. 2011 Jan 1;2.

9. Van Delden C, Iglewski BH. Cell-to-cell signaling and Pseudomonas aeruginosa infections. Emerging infectious diseases. 1998 Oct;4(4):551.

10. Laville J, Blumer C, Von Schroetter C, Gaia V, Défago G, Keel C, Haas D. Characterization of the hcnABC gene cluster encoding hydrogen cyanide synthase and anaerobic regulation by ANR in the strictly aerobic biocontrol agent Pseudomonas fluorescens CHA0. Journal of Bacteriology. 1998 Jun 15;180(12):3187-96.

11. Solomonson LP. Cyanide as a metabolic inhibitor. Cyanide in biology/edited by B. Vennesland; 1981.

12. Niquet J, Baldwin RA, Allen SG, Fujikawa DG, Wasterlain CG. Hypoxic neuronal necrosis: protein synthesis-independent activation of a cell death program. Proceedings of the National Academy of Sciences. 2003 Mar 4;100(5):2825-30.

13. Lee J, Wu J, Deng Y, Wang J, Wang C, Wang J, et al. A cell-cell communication signal integrates quorum sensing and stress response. Nature Chemical Biology [Internet]. Springer Nature; 2013 Mar 31;9(5):339–43. Available from: http://dx.doi.org/10.1038/nchembio.1225

14. Pessi G, Haas D. Transcriptional Control of the Hydrogen Cyanide Biosynthetic Genes hcnABC by the Anaerobic Regulator ANR and the Quorum-Sensing Regulators LasR and RhlR in Pseudomonas aeruginosa. Journal of Bacteriology. 2000 Dec 15;182(24):6940-9.

15. Lee J, Zhang L. The hierarchy quorum sensing network in Pseudomonas aeruginosa. Protein & Cell [Internet]. Springer Nature; 2014 Sep 25;6(1):26–41. Available from: http://dx.doi.org/10.1007/s13238-014-0100-x

16. Fuqua C, Greenberg EP. Self perception in bacteria: quorum sensing with acylated homoserine lactones. Current opinion in microbiology. 1998 Apr 1;1(2):183-9.
17. Egland KA, Greenberg EP. Quorum sensing in Vibrio fischeri: elements of the luxI promoter. Molecular microbiology. 1999 Feb 1;31(4):1197-204.
18. Bottomley MJ, Muraglia E, Bazzo R, Carfì A. Molecular insights into quorum sensing in the human pathogen Pseudomonas aeruginosa from the structure of the virulence regulator LasR bound to its autoinducer. Journal of Biological Chemistry. 2007 May 4;282(18):13592-600.
19. Fukushima J, Ishiwata T, Kurata M, You Z, Okuda K. Intracellular receptor-type transcription factor, LasR, contains a highly conserved amphipathic region which precedes the putative helix—turn—helix DNA binding motif. Nucleic acids research. 1994 Sep 11;22(18):3706-7.
20. Lintz MJ, Oinuma KI, Wysoczynski CL, Greenberg EP, Churchill ME. Crystal structure of QscR, a Pseudomonas aeruginosa quorum sensing signal receptor. Proceedings of the National Academy of Sciences. 2011 Sep 20;108(38):15763-8.
21. Costas C, López-Puente V, Bodelón G, González-Bello C, Pérez-Juste J, Pastoriza-Santos I, Liz-Marzán LM. Using surface enhanced Raman scattering to analyze the interactions of protein receptors with bacterial quorum sensing modulators. ACS nano. 2015 May 4;9(5):5567-76.
22. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank, 1999–. In: International Tables for Crystallography Volume F: Crystallography of biological macromolecules. Springer Netherlands; 2006. p. 675-684.
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990 Oct 5;215(3):403-10.
24. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic acids research. 2005 Jul 1;33(suppl 2):W244-8.
25. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of applied crystallography. 1993 Apr 1;26(2):283-91.
26. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods in enzymology. 1997 Dec 31;277:396-404.

27. Wiederstein, M., Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic acids research. 2007 Jul 1;35(suppl 2):W407-10.

28. Bolton EE, Wang Y, Thiessen PA, Bryant SH. PubChem: integrated platform of small molecules and biological activities. Annual reports in computational chemistry. 2008 Dec 31;4:217-41.

29. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. Journal of computational chemistry. 2004 Jul 15;25(9):1157-74.

30. da Silva AW, Vranken WF. ACPYPE-Antechamber python parser interface. BMC research notes. 2012 Jul 23;5(1):367.

31. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. Journal of computational chemistry. 2002 Dec 1;23(16):1623-41.

32. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015 Sep 30;1:19-25.

33. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Structure, Function, and Bioinformatics. 2010 Jun 1;78(8):1950-8.

34. Darden T, York D, Pedersen L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. The Journal of chemical physics. 1993 Jun 15;98(12):10089-92.

35. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. The Journal of chemical physics. 1995 Nov 15;103(19):8577-93.

36. 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.

37. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. The Journal of chemical physics. 2007 Jan 7;126(1):014101.

38. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied physics. 1981 Dec;52(12):7182-90.

39. Chen YC. Beware of docking!. Trends in pharmacological sciences. 2015 Feb 28;36(2):78-95.
40. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry. 2010 Jan 30;31(2):455-61.
41. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nature protocols. 2016 May 1;11(5):905-19.
42. Hotelling H. Analysis of a complex of statistical variables into principal components. Journal of educational psychology. 1933 Sep;24(6):417.
43. Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics). 1979 Jan 1;28(1):100-8.
44. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, et al. rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. Prlic A, editor. PLoS Computational Biology [Internet]. Public Library of Science (PLoS); 2014 Apr 10;10(4):e1003571. Available from: http://dx.doi.org/10.1371/journal.pcbi.1003571
45. Gaudreault F, Najmanovich RJ. FlexAID: Revisiting Docking on Non-Native-Complex Structures. Journal of Chemical Information and Modeling [Internet]. American Chemical Society (ACS); 2015 Jul 27;55(7):1323–36. Available from: http://dx.doi.org/10.1021/acs.jcim.5b00078
46. McKinney W. pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing. 2011 Nov:1-9.
47. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011;12(Oct):2825-30.
48. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang LP, Lane TJ, Pande VS. MDTraj: A modern open library for the analysis of molecular dynamics trajectories. Biophysical journal. 2015 Oct 20;109(8):1528-32.
49. Laskowski RA, Swindells MB. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery.

50. Hunter JD. Matplotlib: A 2D graphics environment. Computing In Science & Engineering. 2007 May;9(3):90-5.

51. Waskom, M., Botvinnik, O., Hobson, P., Cole, J. B., Halchenko, Y., Hoyer, S., and Miles, A. (2014) seaborn: v0. 5.0 (November 2014).

52. DeLano WL. The PyMOL molecular graphics system.

53. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. Journal of molecular graphics. 1996 Feb 1;14(1):33-8.

54. 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.

55. Wolf A, Kirschner KN. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11· 23S subdomain. Journal of molecular modeling. 2013 Feb 1;19(2):539-49.

56. Zamora JE, Papadaki M, Messer AE, Marston SB, Gould IR. Troponin structure: its modulation by Ca 2+ and phosphorylation studied by molecular dynamics simulations. Physical Chemistry Chemical Physics. 2016;18(30):20691-707.

57. Hayward S, De Groot BL. Normal modes and essential dynamics. Molecular Modeling of Proteins. 2008:89-106.

58. Amadei A, Linssen A, Berendsen HJ. Essential dynamics of proteins. Proteins: Structure, Function, and Bioinformatics. 1993 Dec 1;17(4):412-25.

59. Tan PN. Introduction to data mining. Pearson Education India; 2006.

60. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. Journal of biomolecular NMR. 2010 Sep 1;48(1):13-22.

61. Han B, Liu Y, Ginzinger SW, Wishart DS. SHIFTX2: significantly improved protein chemical shift prediction. Journal of biomolecular NMR. 2011 May 1;50(1):43.

62. Kumari R, Kumar R, Lynn A. g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. Journal of Chemical Information and Modeling [Internet]. American Chemical Society (ACS); 2014 Jul 28;54(7):1951–62. Available from: http://dx.doi.org/10.1021/ci500020m

63. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. msa: an R package for multiple sequence alignment. Bioinformatics [Internet]. Oxford University Press (OUP); 2015 Aug 26;btv494. Available from: http://dx.doi.org/10.1093/bioinformatics/btv494

64. Thompson JD, Gibson TJ, Higgins DG. Multiple Sequence Alignment Using ClustalW and ClustalX. Current Protocols in Bioinformatics [Internet]. John Wiley & Sons, Inc.; 2002 Aug; Available from: http://dx.doi.org/10.1002/0471250953.bi0203s00

65. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology [Internet]. Wiley-Blackwell; 2014 Apr 16;7(1):539–539. Available from: http://dx.doi.org/10.1038/msb.2011.75

66. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research [Internet]. Oxford University Press (OUP); 2004 Mar 8;32(5):1792–7. Available from: http://dx.doi.org/10.1093/nar/gkh340

67. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S. The ClusPro web server for protein-protein docking. Nature Protocols. 2017 Feb 1;12(2):255-78.

68. Johnson SC. Hierarchical clustering schemes. Psychometrika. 1967 Sep 27;32(3):241-54.

69. Davies DL, Bouldin DW. A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence. 1979 Apr(2):224-7.

70. Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.

71. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics. 1987 Nov 1;20:53-65.

72. Caliński T, Harabasz J. A dendrite method for cluster analysis. Communications in Statistics-theory and Methods. 1974 Jan 1;3(1):1-27.

73. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM. PDBsum: a Web-based database of summaries and analyses of all PDB structures. Trends in biochemical sciences. 1997 Dec 1;22(12):488-90.

74. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G. A critical assessment of docking programs and scoring functions. Journal of medicinal chemistry. 2006 Oct 5;49(20):5912-31.

75. Wernet P, Nordlund D, Bergmann U, Cavalleri M, Odelius M, Ogasawara H, Näslund LÅ, Hirsch TK, Ojamäe L, Glatzel P, Pettersson LG. The structure of the first coordination shell in liquid water. Science. 2004 May 14;304(5673):995-9.

76. Zheng Y, Sintim HO. Molecular Insights into How Ligands Activate or Inactivate LasR. Chemistry & Biology [Internet]. Elsevier BV; 2014 Oct;21(10):1261–3. Available from: http://dx.doi.org/10.1016/j.chembiol.2014.10.001

77. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein–protein docking. Nucleic acids research. 2004 Jul 1;32(suppl 2):W96-9.

78. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics. 2004 Jan 1;20(1):45-50.

79. Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: an FFT-based protein docking program with pairwise potentials. Proteins: Structure, Function, and Bioinformatics. 2006 Nov 1;65(2):392-406.