RESEARCH ARTICLE

Interaction of N-3-oxododecanoyl homoserine lactone with transcriptional regulator LasR of Pseudomonas aeruginosa: Insights from molecular docking and dynamics simulations [version 1; peer review: awaiting peer review]

Hovakim Grabski¹, Lernik Hunanyan¹, Susanna Tiratsuyan¹,², Hrachik Vardapetyan¹

¹Department of Medical Biochemistry and Biotechnology, Russian-Armenian University, Yerevan, 0051, Armenia
²Faculty of Biology, Yerevan State University, Yerevan, 0025, Armenia

Abstract

Background: In 2017, the World Health Organization announced a list of the most dangerous superbugs. Among them is Pseudomonas aeruginosa, an opportunistic human pathogen with high levels of resistance to antibiotics that is listed as one of the ‘ESKAPE’ pathogens, which are the leading cause of nosocomial infections. A major issue is that it mostly affects vulnerable patients such as those suffering from AIDS, cystic fibrosis, cancer and severe burns. P. aeruginosa creates and inhabits surface-associated biofilms which increase resistance to antibiotics and host immune responses and contribute to the ineffectiveness of current antibacterial treatments. It is therefore imperative to find new antibacterial treatment strategies against P. aeruginosa. The LasR protein is a major transcriptional activator of P. aeruginosa and plays a pivotal role in biofilm formation and the activation of many virulence genes, although detailed characteristics of the LasR protein are not currently known. In the present study, we aimed to analyse the molecular properties of the LasR protein as well as its interactions with the signalling molecule N-3-oxododecanoyl homoserine lactone (3OC12-HSL).

Methods: We used a combination of molecular docking, molecular dynamics (MD) simulations and machine learning techniques to study the interaction of the LasR protein with the 3OC12-HSL ligand. We assessed conformational changes occurring upon their interaction and analysed the molecular details of their binding.

Results: A new possible interaction site for 3OC12-HSL and LasR was found, involving conserved residues from the ligand binding domain (LBD), beta turns in the short linker region (SLR) and the DNA-binding domain (DBD). This interaction is referred to as the LBD-SLR-DBD bridge or ‘the bridge’ interaction.

Conclusions: This study may enable future experimental studies to detect the interaction of signalling molecules with “the bridge” of the LasR protein and suggests a potential new interaction site to assist antibacterial drug design.
Keywords
Pseudomonas aeruginosa, transcriptional regulator, LasR, antibiotic resistant, ESKAPE, 3OC12-HSL, homology modelling, molecular docking, molecular dynamics, machine learning techniques

Corresponding author: Hovakim Grabski (hovakim.grabski@rau.am)

Author roles: Grabski H: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; Hunanyan L: Supervision; Tiratsuyan S: Investigation, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; Vardapetyan H: Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The authors gratefully acknowledge financial support by the Russian-Armenian University [NIR 25/15]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Grabski H et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Grabski H, Hunanyan L, Tiratsuyan S and Vardapetyan H. Interaction of N-3-oxododecanoyl homoserine lactone with transcriptional regulator LasR of Pseudomonas aeruginosa: Insights from molecular docking and dynamics simulations [version 1; peer review: awaiting peer review] F1000Research 2019, 8:324 (https://doi.org/10.12688/f1000research.18435.1)

### Abbreviations

PDB, Protein Data Bank; MD, molecular dynamics; PCA, principal component analysis; 3O12-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; IQS, integrated quorum sensing; QS, quorum sensing; AHL, acyl homoserine lactone; HCN, hydrogen cyanide; HTH, helix-turn-helix; MSA, multiple sequence alignment; RMSD, root-mean-square-deviation; LBD, ligand binding domain; DBD, DNA binding domain.

### Introduction

*Pseudomonas aeruginosa* (*P. aeruginosa*) is a Gram-negative, monoflagellated, obligate aerobe. This species of bacteria is one of the ‘ESKAPE’ pathogens, which includes six bacterial pathogens which are commonly associated with antimicrobial resistance and are the leading cause of nosocomial infections throughout the world. It can be found in many diverse environments such as in soil, plants and hospitals. *P. aeruginosa* is an opportunistic human pathogen because it rarely infects healthy people, mostly affecting patients suffering from AIDS, cystic fibrosis, cancer and burns. Most of the deaths caused by cystic fibrosis are due to this pathogen. The pathogenicity of *P. aeruginosa* is due to virulence factors such as: the synthesis of proteases, hemolysins, exotoxin A, pyocyanin, hydrogen cyanide (HCN), and rhamnolipids; the possession secretion systems such as in soil, plants and hospitals. *P. aeruginosa* is a common characteristic of many bacteria as the bacterial population density increases in the body during pathology. These bacteria possess a system of regulating gene expression in response to population density, called quorum sensing (QS), that uses hormone-like molecules called autoinducers (AIs) which accumulate in the extracellular matrix. When a threshold is reached, the AIs bind to their cognate receptors and then a response regulator modulates the expression of QS virulence genes which regulate adaptation, colonization, antibiotic resistance, plasmid conjugation and other bacterial processes.

*Pseudomonas aeruginosa* has four QS systems. The type one system is regulated by LuxI/LuxR-type proteins. AIs diffuse freely across the bacterial membrane and bind to the transcriptional activator LuxR. Type two and three systems are also LuxI/LuxR-type systems, the first LasI that produces 3O12-HSL and the second, RhlII that synthesizes C4-HSL; both acyl homoserine lactones (AHL) which regulate virulence and biofilm formation. A fourth integrated QS (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.

QS regulates the expression of virulence factors such as HCN. Exposure to HCN can lead to neuronal necrosis through the inhibition of cytochrome c oxidase, the terminal component of the aerobic respiratory chain. In *P. aeruginosa*, the hcnABC operon is responsible for HCN biosynthesis by the enzyme HCN synthase. Three transcriptional regulators (LasR, ANR and RhI1R) control the transcription of the hcnABC gene cluster although it has been proposed that LasR is the crucial activator of hcnABC genes following mutagenesis experiments.

The transcriptional activator protein LasR regulates target gene expression by recognizing a conserved DNA sequence termed a lux box. LasR has two domains: 1) a ligand binding domain at the N-terminus (LBD); 2) a DNA-binding domain at the C-terminus (DBD). LasR has a DNA-binding helix-turn-helix (HTH) motif. Binding of the AI 3-O-C12HSL stabilizes LasR and promotes its dimerization. After that, the resulting LasR-AI homodimer complex binds to the promoter of the target DNA and induces gene transcription.

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

There remains a need for a further understanding of the LasR monomer because to date there is no detailed information about LasR monomer interactions. Hence, we analysed the molecular details of the interactions of 3O12-HSL with the LasR protein. So far, this is the first report that shows that 3O12-HSL can interact with conserved amino acid residues from the ligand binding domain (LBD), beta turns in the short linker region (SLR) and the DNA-binding domain of LasR. For simplification, we refer to this region as the LBD-SLR-DBD bridge or “the bridge”. This could mean that modulation of the transcriptional regulator LasR is more nuanced than previously thought. This study may be utilized for the development of new therapeutic agents against *P. aeruginosa* targeting both the LBD and the LBD-SLR-DBD bridge of LasR in order to inhibit the expression of virulence genes.

### Methods

Figure 1. Flow chart providing an overview of the methodology used in this study.

### Analysis of the LasR protein sequence

This part of the methodology was based on the work by Chowdhury et al. to reproduce the full structure of LasR protein. The amino acid sequence of the LasR protein of *P. aeruginosa* was obtained from UniprotKB (Uniprot ID: P25084). The crystal structure of the LasR protein LBD (amino acid residues 7 to 169) from *P. aeruginosa* was acquired from the Protein Data Bank (PDB ID: 3IX3). However, the entire structure of the LasR protein was required in order to have a better understanding of its molecular properties.

For this reason, the amino acid sequence (Uniprot ID: P25084) was used to search the PDB databank using BLASTp in order to identify suitable templates for homology modelling.
crystal structure of QscR bound to 3OC12-HSL was found to be the best fit based on sequence identity, query coverage, and E-value from the BLAST results (PDB ID: 3SZT). The full list of similar structures obtained from the BLAST search are shown in Table 1.

Reconstruction of the LasR monomer
As the current available crystal structures of LasR only contain the ligand binding domain, we needed to reconstruct the full structure of LasR with the missing DNA-binding domain (DBD). The HHpred web server was used for homology modelling of the missing LasR DBD as was performed by Chowdhury et al. The four templates used for homology modelling of the LasR DBD have the PDB identifiers 3SZT, 3IX3, 1FSE and 3ULQ. The experimental structure of LasR and these templates were used to reconstruct the final model of LasR that contains both the LBD and the DBD. The final reconstructed model of the LasR protein was verified using PROCHECK, Verify3D 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. This score indicates a suitable computational model. The Ramachandran plot revealed that no amino acid residues were located in the disallowed regions. The ProSA value of this model was -6.69 which suggests that the quality of the homology model is high. Although the obtained model is slightly different from that obtained by Chowdhury et al., it is still viable.

Analysis of docking conformations and trajectories
The analysis of docking conformations and trajectories was performed using custom Python scripts (see Extended data). These scripts use pandas, scikit-learn and MDTraj libraries. pandas facilitates data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. scikit-learn is a machine-learning library, which provides a set of common algorithms to Python users through a consistent interface. The MDTraj Python library facilitates the combination of machine learning libraries such as scikit-learn for trajectory analysis. MDTraj serves as a bridge, connecting molecular dynamics (MD) data with statistics and graphics libraries developed for general data science users.

Diagrams of the interaction of 3OC12-HSL with LasR were obtained using LigPlot+ v.1.4 using default parameters, which automatically generates 2D diagrams of ligand–protein interactions from 3D coordinates by loading the coordinate files in pdb format. Plot visualization was carried out using matplotlib v2.0.1, seaborn v0.8.1. The seaborn library aims to make visualization a central part of exploring and understanding data. It also provides a concise, high-level interface for drawing statistical graphics. Figures and videos were prepared with PyMOL v1.9.0.0, VMD 1.9.3 and UCSF Chimera v1.11.2. VMD is often used for this purpose as it has the ability to read trajectory files created during simulations in formats produced by many different software packages. PyMOL and UCSF Chimera are also commonly used by researchers and produce high quality images but are typically less straightforward to use for viewing and making videos of the trajectories.

The analysis protocol was similar to the work of Wolf et al. and Zamora et al., using PCA and cluster analysis modules within scikit-learn v0.18.2. 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. The process of applying PCA to a protein trajectory is called essential dynamics (ED). PCA
analysis, performed using Cartesian coordinates, has become an important tool for the examination of conformational changes. Default parameters within the scikit-learn library were used. Cluster analysis is another unsupervised technique that tries to identify structures within 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. kmeans algorithm was used for molecular docking analysis. Agglomerative clustering using Ward’s linkage was used for molecular dynamics trajectory analysis with the connectivity matrix built on the basis of the k-Neighbors graph. Default parameters within the scikit-learn library were used.

Molecular dynamics simulations of LasR protein systems

Molecular dynamics (MD) simulations of all systems were conducted with the GROMACS suite, version 5.1.2, utilizing the Amber ff99SB-ILDN force field, which exhibits considerably better agreement with NMR data than other forcefields. In all cases, short-range non-bonded interactions were cut off at 1.4 nm. The Particle mesh Ewald (PME) method was used for the calculation of long-range electrostatics.

LasR monomer simulation. In order to generate the starting structure of the LasR monomer before docking, the LasR molecule was placed in a dodecahedron box of TIP3P water. After that, 100 mM NaCl was added, including neutralizing counterions. Following two minimizations by the method of steepest descent algorithm, the LasR monomer was equilibrated in two stages. The first stage involved simulating for 200 picoseconds under a constant volume (NVT) ensemble. The second stage involved simulating for 200 picoseconds under a constant pressure (NpT), maintaining pressure isotropically at 1.0 bar. Production simulation was conducted for 100 nanoseconds in the absence of any restraints. The temperature was sustained at 300 K using the v-rescale algorithm. For isotropic regulation of the pressure, the Parrinello-Rahman barostat was used. This combination of thermostat and barostat use ensures that an adequate NpT ensemble is sampled. Finally, the trajectory of the LasR monomer simulation was used for the calculation of chemical shifts. SPARTA+ v2.90 and SHIFTX2 v1.10 were used to predict the chemical shifts of the protein backbone atoms with the help of the MDTraj wrapper functions. Both programs have been proven to be remarkably accurate, especially for predicting 13C shifts. This was used to validate whether the force field is appropriate for further investigation. The final extracted structure that was used for docking and MD simulations was also verified with Gaia from the Chiron webserver, which compares the intrinsic structural properties of in silico protein models to high resolution crystal structures.

Molecular dynamics simulations using docking poses. The 3D model of 3OC12-HSL was obtained from PubChem (CID: 127864) (Figure 2). It has been shown that docking has its limitations. For this reason, after finishing the molecular docking simulations of 3OC12-HSL with LasR, the docking poses were extracted to perform molecular dynamics (MD) simulations. The ligand parameters for the molecular dynamics simulations were calculated for the General Amber Force Field (GAFF) using the ACPYPE tool with AM1-BCC partial charges and generation of a GAFF topology. A time step of 2 femtoseconds was used during heating and the production run coordinates were recorded every 1 picosecond. Three simulations of 300 nanoseconds were performed. Further details about the simulation protocol can be found in ‘LasR monomer simulation’ section.

LasR–3OC12-HSL ligand blind docking experiments

To build the LasR–ligand complex, the ligand 3OC12-HSL was docked with the LasR monomer using AutoDock Vina v1.1.2. However, AutoDock Vina currently uses several simplifications that affect the obtained results. The most notable simplification, as the creators note, is the use of a rigid receptor. The amount of computation used during a docking experiment is determined by a parameter called ‘exhaustiveness’ in AutoDock Vina. However, the default exhaustiveness value is 8 and the creators claim that increasing this value leads to an exponential increase in the probability of finding the energy minima. The conformational space of the whole protein was searched using grid box dimensions 60x62x48Å. The following exhaustiveness values were tested in this study: 8, 16, 32, 64, 128, 256, 512, 1024, 2048 and 4096. The ligand and target structures containing hydrogen atoms were prepared using the AutoDockTools from MGLTools v1.5.6.

Principal component analysis (PCA) and cluster analysis using k-means clustering was performed (Figures S4 and S5, Extended data). The results demonstrated that the number of
interaction sites does not change for exhaustiveness values from 1024 to 4096. An exhaustiveness value of 1024 was chosen as it provides good results, good speed and thorough sampling of the docked configurations. The exhaustiveness value was set to 1024 and the maximum number of binding modes to generate was set to 20. After that, 100 independent docking calculations were carried out with random initial seeds. The target was taken from the molecular dynamics simulation of the LasR monomer. Water molecules were removed from the structure.

We also performed another set of docking simulations using AutoDock Vina v1.1.2\textsuperscript{10}, rDock v2013.1\textsuperscript{16} and FlexAID v2.48\textsuperscript{49}. We performed 10 docking simulations using each program. Each individual run was set to generate 20 docked poses. For AutoDock Vina, we performed one docking run for each exhaustiveness value: 8, 16, 32, 64, 128, 256, 512, 1024, 2048 and 4096.

Docking was also performed using the rDock program\textsuperscript{49} which was previously found to be successful in predicting the binding mode for the CCDC-Astex set of 85 diverse protein–ligand complexes in approximately 80\% of cases\textsuperscript{66}. A cavity with a radius of 36 Å was set and centred on the LasR monomer to cover the whole protein surface and define the docking volume. AutoDock Vina and rDock were chosen based on the performance of these academic programs\textsuperscript{49}.

FlexAID (parameters are available in Extended data\textsuperscript{66}) was also used as, unlike other programs, it is robust against increasing structural variability\textsuperscript{67}. FlexAID performs better than AutoDock Vina in all scenarios against non-native complex structures\textsuperscript{51}. FlexAID uses a soft and smooth scoring function based on contact surfaces which is different from Autodock Vina and rDock. FlexAID was run using default parameters.

**Binding energy calculation**

g\textsubscript{mmbpsa} v1.6\textsuperscript{69} 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\textsubscript{mmbpsa} calculates the relative binding energy rather than the absolute binding energy, so the entropy contribution (TΔS) is not evaluated\textsuperscript{69}. From the two simulations with length of 300 nanoseconds, we cut out 10 nanoseconds with an interval of every 120 picoseconds for the calculation of binding energies. g\textsubscript{mmbpsa} uses the MM-PBSA approach, which has grown to be one of the most widely used methods to compute interaction energies and is often employed to study biomolecular complexes\textsuperscript{65}.

**Sequence conservation analysis**

To find out whether 3OC12-HSls interact with conserved amino acid residues, we performed multiple sequence alignment (MSA) using the R msa package v1.4.5\textsuperscript{80}. ClustalW\textsuperscript{81}, Clustal Omega\textsuperscript{82} and MUSCLE\textsuperscript{83} within the msa package were used to carry out multiple sequence alignments. ClustalW and MUSCLE are commonly and widely used, while Clustal Omega is a more recent and powerful method that is used when aligning a large number of sequences\textsuperscript{82}. We performed multiple sequence alignment of the LasR protein and closest homologue sequences, based on the work by Chowdhury \textit{et al.}\textsuperscript{1}. The consensus threshold parameter in msa was set to 75, the rest was set to default. Table 2 shows the sequences used for sequence alignment.

**Protein-protein docking**

When docking homology models, it is preferable to have experimental evidence suggesting the general interaction site (within ~10 Å). Representative structures from molecular dynamics simulations were used for protein-protein docking using ClusPro v2.0\textsuperscript{14}. ClusPro was chosen because it gave equivalent results to the best human predictor group according to the latest CAPRI experiments carried out in 2013\textsuperscript{10}. From the experimental X-ray data, it was found that ‘Trp152’, ‘Lys155’ and ‘Asp156’ play an important role during dimerization. The distance between ‘Trp152’ from chain A and ‘Asp156’ from chain B of the crystallographic structure was 0.276 nm and the distance between ‘Asp156’ from chain A and ‘Lys155’ from chain B was 0.279 nm. These residues were used as attraction constraints.

**Results**

**Conformational changes of LasR without 3OC12-HSL**

We performed a simulation run of 100 nanoseconds using a standard MD protocol for the assessment of conformational changes of the LasR monomer. 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 the MD run (Figure S6, Extended data\textsuperscript{86}). Another suitable measurement for the stability of the LasR monomer structure is the radius of gyration, which also shows the structure is stable (Figure S7, Extended data\textsuperscript{86}).
During the examination of MD trajectories, principal component analysis (PCA)\(^6\) is usually used for the visualization of the motion of the system (Figure 3a). Generally, in order to capture more than 70% of the variance in the trajectory data, the first handful of components are sufficient\(^8\). PCA can uncover the fundamental movements contained in an MD trajectory, however, it does not group the snapshots into different clusters\(^4\). This can be accomplished by clusterization of the PCA data.

Cluster analysis was performed on the LasR monomer simulation for the selection of an initial starting structure to study docking. Identifying a distinct, representative structure of the cluster allows blind docking to be performed on the whole structure. It should be also noted that cluster analysis allows the evaluation of the frequent conformations of LasR. For the clustering analysis, we chose agglomerative clustering using Ward’s linkage\(^6\). This algorithm appears to be the most appropriate as it is deterministic, allowing for reproducibility of the resulting clusters, thus minimizing the amount of bias.

We performed several rounds of agglomerative clustering using Ward’s linkage. The accuracy of the clustering was assessed with the help of the Davies–Bouldin index (DBI)\(^5\), Dunn index\(^9\), silhouette score\(^7\) and Calinski-Harabasz pseudo-F statistic (pSF)\(^8\) metrics. An optimal number of clusters were chosen, simultaneously accounting for low DBI, high silhouette, high Dunn index and high pSF values.

The distribution of clusters over the simulation is visualized in Figure 3b and the four clusters 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 the end. The clusterization defined by the first two principal components (Figure 3b) provides a coherent picture and it is also supported by good DBI, Dunn index, silhouette score and pSF values (Figure S8, Extended data\(^2\)). Figure 3c shows clearly that the simulation has converged.

For validation of the molecular dynamics simulation quality, theoretical NMR shifts were calculated using Sparta\(^+\)\(^1\) and ShiftX\(^2\). For the NMR shifts calculation, snapshots from cluster 3 (Figure 3b) were used because of their stability. We compared NMR values of C-alpha atoms from the experimental NMR of the LBD with theoretical NMR values of the LBD from our molecular dynamics simulations (BMRB: 6271)\(^6\). Unfortunately, the current available NMR information only includes the 173 amino acid residues of the LBD. In Figure 4, one can see that simulated NMR shift values from the two programs are close to the experimental data, thus demonstrating the quality of the MD simulation of the LasR monomer.

It should be noted this protein becomes partially soluble when the ligand interacts with it\(^6\). The differences between the theoretical chemical shifts and experimental values could be attributed to: 1) the theoretical model does not involve the interaction with the ligand 2) the experimental NMR only includes information about the LBD and not the DBD, which can affect the LBD.

After that, the representative structure (Figure 5c) was extracted from cluster 3 (Figure 3c) to study docking. The representative structure of the LasR system and its differences from the homology modelled structure (Figure 5b) are visible in Figure 5c.

**Table 2. LasR and LasR homologue sequences from closely related species used for multiple sequence alignment.**

<table>
<thead>
<tr>
<th>Reference Sequence</th>
<th>Accession</th>
<th>Name</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. WP_054058449.1</td>
<td>WP_054058449.1</td>
<td>LuxR family transcriptional regulator</td>
<td>Pseudomonas fuscovaginae</td>
</tr>
<tr>
<td>2. NP_250121.1</td>
<td>NP_250121.1</td>
<td>Transcriptional regulator LasR</td>
<td>Pseudomonas aeruginosa PAO1</td>
</tr>
<tr>
<td>3. KFC75736.1</td>
<td>KFC75736.1</td>
<td>LuxR-type transcriptional regulator</td>
<td>Massilia sp. LC238</td>
</tr>
<tr>
<td>4. WP_018433960.1</td>
<td>WP_018433960.1</td>
<td>LuxR family transcriptional regulator</td>
<td>Burkholderia sp. JPY251</td>
</tr>
<tr>
<td>5. WP_042326260.1</td>
<td>WP_042326260.1</td>
<td>LuxR family transcriptional regulator</td>
<td>Paraburkholderia ginsengisoli</td>
</tr>
<tr>
<td>6. WP_012426170.1</td>
<td>WP_012426170.1</td>
<td>LuxR family transcriptional regulator</td>
<td>Paraburkholderia phytofirmans</td>
</tr>
<tr>
<td>7. WP_027776298.1</td>
<td>WP_027776298.1</td>
<td>LuxR family transcriptional regulator</td>
<td>Paraburkholderia caledonica</td>
</tr>
<tr>
<td>8. WP_027214716.1</td>
<td>WP_027214716.1</td>
<td>LuxR family transcriptional regulator</td>
<td>Burkholderia sp. WSM2232</td>
</tr>
<tr>
<td>9. WP_041729325.1</td>
<td>WP_041729325.1</td>
<td>LuxR family transcriptional regulator</td>
<td>Burkholderia sp. CGE1003</td>
</tr>
<tr>
<td>10. CBI71275.1</td>
<td>CBI71275.1</td>
<td>UnaR protein</td>
<td>Paraburkholderia unamae</td>
</tr>
<tr>
<td>11. CAP91064.1</td>
<td>CAP91064.1</td>
<td>BraR protein</td>
<td>Paraburkholderia kururiensis</td>
</tr>
<tr>
<td>12. WP_003082999.1</td>
<td>WP_003082999.1</td>
<td>Transcriptional activator protein LasR</td>
<td>Pseudomonas</td>
</tr>
<tr>
<td>13. WP_012076422.1</td>
<td>WP_012076422.1</td>
<td>LuxR family transcriptional regulator</td>
<td>MULTISPECIES: Pseudomonas</td>
</tr>
<tr>
<td>14. WP_050376898.1</td>
<td>WP_050376898.1</td>
<td>LuxR family transcriptional regulator</td>
<td>MULTISPECIES: Pseudomonas</td>
</tr>
<tr>
<td>15. WP_050395760.1</td>
<td>WP_050395760.1</td>
<td>LuxR family transcriptional regulator</td>
<td>Pseudomonas aeruginosa</td>
</tr>
</tbody>
</table>
Figure 3. LasR monomer simulation analysis a) Simulation trajectory projected onto its first two principal components. The black scale indicates the time progression from 0 ns (white) to 100 ns (black) b) Clustering results of ward-linkage algorithm formed by first two PCs c) Colour-coded RMSD of the simulation obtained from Ward-linkage cluster analysis.

Figure 4. Plot of chemical shifts of Cα-atom vs. residue number.
was performed using PDBSum\textsuperscript{79} (Figure S9, Extended data\textsuperscript{80}). Finally, the extracted centroid structure was validated using Gaia\textsuperscript{81} (Figure S10, Extended data\textsuperscript{80}), which also shows the model is of a high quality. Therefore, in this study, a model has been developed to include the dynamics of the full-length LasR molecule (residues 1 to 239) and it has been shown that the dynamics of the complete C-terminal region of LasR modulate the N-terminal region, as will be discussed later.

**Docking analysis of 3OC12-HSL with LasR monomer**

In this study, a molecular docking approach was used to inspect the possible binding modes of 3OC12-HSL with the LasR monomer. PCA and cluster analysis were performed on docking data (Figure 6a, b) and the results show three binding sites. Cluster 2 corresponds to experimental data, while cluster 1 and cluster 3 do not (Figure 6c). These results clearly support the findings of Bottomley et al.\textsuperscript{14}, who also demonstrated that 3OC12-HSL binds to the LBD.

We performed several rounds of k-means clustering. The accuracy of the cluster analysis was evaluated using the DBI\textsuperscript{59}, Dunn index\textsuperscript{59}, silhouette score\textsuperscript{72} and pSF\textsuperscript{73} metrics (Figure S11, see Extended data\textsuperscript{80}). An optimal number of clusters were chosen for docking analysis, simultaneously accounting for 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 3OC12-HSL binding sites. The representative structures were produced by identifying the centroid conformations of the clusters using the following algorithm:

1. Compute all pairwise root mean square deviations (RMSDs) between the conformations.

2. Transform the distances into similarity scores and they are calculated as

$$S_{ij} = e^{-\beta d_{ij}^2 \text{scale}}$$  \hspace{1cm} (eq. 1)

Where $s_{ij}$ is the pairwise similarity, $d_{ij}$ is the pairwise distance, and $d_{\text{scale}}$ is the standard deviation of the values of $d$ to make the computation scale invariant.

3. Then the centroid is defined with $\beta=1$ as

$$\text{argmax}_{i} \sum S_{ij}$$  \hspace{1cm} (eq. 2)

Representative structures from each cluster were extracted for further use in MD simulations. The binding energy of the representative structure of cluster 1 is -5.4 kcal/mol and the mean binding affinity for the whole cluster is -5.257 ±0.233 kcal/mol. Cluster 1 contains 839 docked poses from a total 2000 (41.95%). For cluster 2, the binding affinity of the representative structure is -5.1 kcal/mol and the mean for the whole cluster is -5.593 ±0.386 kcal/mol. These results correspond to the experimental binding site data\textsuperscript{14}. Cluster 2 contains 864 docked poses from a total of 2000 (43.2%). For cluster 3, the representative structure features the highest binding affinity of -5.7 kcal/mol and the mean binding affinity for the whole cluster is -5.264 ±0.27 kcal/mol. Cluster 3 contains 297 docked poses from 2000 (14.85%) which does not correspond to the experimental binding site.

It has been suggested that molecular docking is not appropriate for the prediction of binding affinity or binding poses of protein-ligand complexes. However, docking analysis can still provide important information regarding potential binding sites and the resulting poses can be used for MD simulations\textsuperscript{96,80}.

We also ran another set of simulations to corroborate the blind docking with other molecular docking software, including AutoDock Vina\textsuperscript{82}, rDock\textsuperscript{66} and FlexAID\textsuperscript{82} (Figure 7). All three programs, which each use different algorithms, suggest that there are two potential binding sites. One is within the LBD, which corresponds to experimental data\textsuperscript{84}, and the other is within the ‘bridge’. PCA of the results from these docking programs was performed for easier comprehension (Figure S12, Extended data\textsuperscript{84}).

**Binding modes of 3OC12-HSL by MD simulations**

In total, 900 nanoseconds of MD simulations have been performed and used for the analysis of 3OC12-HSL interaction with the LasR monomer. The representative structures were taken from docking results (Figure 6c) and used as starting points for MD simulations with LasR. Simulations were conducted for a sufficient time to allow the positions of 3OC12-HSL to reach equilibrium in the LasR molecule.

The stability of the molecule was evaluated by calculating the RMSD of the backbone atoms. RMSD was calculated with reference to the initial snapshot. Figure 8a shows that Simulation 2 and 3 experience a substantial RMSD deviation from the initial starting point. Simulation 2 corresponds to cluster 2 in the docking simulations, while Simulation 3 corresponds to cluster 3 (Figure 8a). For Simulation 1, which corresponds to cluster 1, the 3OC12-HSL molecule did not fixate and reach equilibrium and therefore no further analysis was performed (Figure 8a) for this simulation. Simulation 2 shows that after 230 nanoseconds, the structure becomes stable. For Simulation 3, the structure changes its conformation at 100 nanoseconds and becomes stable between 260 nanoseconds and 300 nanoseconds.
PCA and cluster analysis were performed on simulation 2. Hydrogen bond analysis was also performed, based on cut-offs for distance and angles according to the Wernet-Nilson criteria\(^{81}\) using MDTraj\(^{31}\). During the simulation, 3OC12-HSL forms hydrogen bonds with Tyr56, Ser129 and Trp60, which is in agreement with experimental data\(^{18}\) (Figure S14, Extended data\(^{26}\)). Over the course of simulation 2, 3OC12-HSL also establishes a large number of hydrophobic contacts with the amino acid side chains of the LasR LBD (Figure 9b). This phenomenon is not unexpected, given the large hydrophobic surface area of the LasR LBD and the low solubility of 3OC12-HSL in water. 3OC12-HSL has hydrophobic interactions mainly with amino acids from helix 3, helix 5 and sheet 1 side chains. RMSD analysis of the conformations of LasR with and without 3OC12-HSL bound to the LBD gives a value of 7.027 Å between the conformations (Figure S16, Extended data\(^{26}\)). RMSD of the modelled full structure of LasR with ligand superimposed onto the experimentally obtained tertiary structure of LBD of LasR with ligand gives a value of 1.692 Å, where RMSD <2 Å in relation to the experimentally studied structure (Figure 9c). Residues that participate in hydrophobic interactions and hydrogen bonding are shown in Figure 9e and Figure S18 and 19 (Extended data\(^{26}\)).

The second binding mode involves the interaction of 3OC12-HSL with the LBD-SLR-DBD bridge. PCA, cluster and hydrogen bond analysis were also performed on simulation 3. Over the course of simulation 3, 3OC12-HSL establishes hydrogen bonds and a large number of hydrophobic contacts with amino acid side chains in the beta turns of the LasR SLR. In simulation
Figure 8. Analysis for 3 independent runs of 300 ns using representative structures from 3OC12-HSL docking poses as starting points. 

a) RMSD evolution

b) RMSF of Cα-atoms of LasR for each individual run.
Figure 9. Visualizations of the interactions of 3OC12-HSL with LasR. Conformations taken from centroids from cluster analysis a) 3OC12-HSL with LasR LBD b) Insertion of 3OC12-HSL in LBD of LasR c) Superimposition of the modelled interaction (red) of 3OC12-HSL with LasR LBD and crystallographic structure (green) d) 3OC12-HSL with the “bridge” of LasR e) Putative hydrogen bonds of 3OC12-HSL with Lys182 and Leu177.

Energy calculations to assess residue interactions and sequence alignment show that 3OC12-HSL interacts with highly conserved amino acid Trp60, which agrees with experimental data. The interaction also involves nine other conserved amino acids: Tyr64, Asp73, Pro74, Val76, Phe102, Phe103, Ala105 and Gly113 (Figure 10b). In total, 16 amino acids that contribute to the interaction have over 75% conservation.

Protein docking of MD models
It is known that LasR binds to the corresponding promoter as a dimer after interacting with the autoinducer. Thus, monomeric LasR–3OC12-HSL complexes were subjected to dimerization as an additional experiment. The protein docking experiments using structures from MD runs were performed using ClusPro v2.0 because of its success in the CAPRI (Critical Assessment of Predicted Interactions) scoring experiment. We used the centroid conformations of the LasR bridge-3OC12-HSL and LasR LBD-3OC12-HSL complexes from MD simulations.
Table 3. Relative binding energies calculated from simulation data using g_mmpbsa.

<table>
<thead>
<tr>
<th>Sim №</th>
<th>Binding sites</th>
<th>Van der waals (kJ/mol)</th>
<th>Electrostatic interactions (kJ/mol)</th>
<th>Polar salvation (kJ/mol)</th>
<th>Non-polar salvation (kJ/mol)</th>
<th>Binding energy (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>LBD-SLR-DBD bridge</td>
<td>-205.141 ± 11.803</td>
<td>-142.974 ± 29.032</td>
<td>201.889 ± 20.117</td>
<td>-20.230 ± 0.842</td>
<td>-166.456 ± 20.492</td>
</tr>
</tbody>
</table>

Figure 10. Energy contributions of residues and multiple sequence alignments demonstrating the conserved amino acid residues of LasR that interact with 3OC12-HSL. a) All amino acids that interact with LBD b) Conserved amino acids that interact with LBD. Blue boxes and arrows point the amino acid residues that interact with 3OC12-HSL. c) All amino acids that interact with the “bridge”. Black boxes and arrows point the residues that interact with 3OC12-HSL. d) Conserved amino acids that interact with “bridge”. Black boxes and arrows point the residues that interact with 3OC12-HSL.

Simulations. Then we performed protein docking using protein conformations without the ligands and used constraints based on the residue distances between monomers from the experimental data. For the selection of the model, we used an approach recommended by the authors of ClusPro, which suggests finding the most populated clusters. For LasR bridge-3OC12-HSL complex docking, the top model contained 80 members and the scores by ClusPro for the docking model were -951.4 for the centre and -1332.0 and for the lowest energy region, suggesting a favourable binding mode (Figure 11a). For the LBD LasR-3OC12-HSL complex, the top model contained 122 members and the scores for the docking model were -1440.7 for the centre and -1517.9 for the lowest energy region (Figure 11b). Both docking structures feature negative energy and this could
suggest the possibility of multiple docking conformations. It is possible that the multiple binding modes of 3OC12-HSL imply that LasR has multiple dimerization interfaces like QscR (Figure 11c), which is a post-translational repressor of LasR. In the absence of 3OC12-HSL, it forms a heterodimer with LasR. It appears this protein may have a much more delicate regulation than was initially thought.

Discussion

Pseudomonas aeruginosa is one of the ESXAVE pathogens, which are the leading cause of nosocomial infections throughout the world. However, the molecular mechanisms of the LasR protein in regulating quorum sensing machinery have not been fully elucidated. LasR has been problematic to structurally analyse for the last decade because of the insolubility of the full protein. All of the available crystallographic structures contain only the LBD and lack the DBD. The lack of a full crystal structure has prevented us from furthering our understanding of LasR as knowledge of its full 3D structure is vital. It is therefore very important to study the interaction of the autoinducer with the full structure of the LasR protein. Adequate knowledge of this interaction would provide the opportunity to conduct improved structure-based drug discovery studies.

Results from our analysis show that the interaction of 3OC12-HSL with LasR has two binding modes. The results were cross-checked using multiple docking programs, molecular dynamics simulations and machine learning techniques. The 3OC12-HSL has a long hydrophobic tail and is therefore capable of binding to both the LBD and the “bridge” of LasR. These results suggest that both the C-terminal and the N-terminal regions of LasR interact with 3OC12-HSL. The binding of 3OC12-HSL provokes conformational transitions in the structure of LasR. Binding energy analysis showed that the “bridge” does not compete with LBD. An MM-PBSA approach was used for the calculation of relative binding energy, although this finding needs confirmation through further research. The calculation of absolute binding energy could give more reliable results, though this type of calculation is more time consuming. The results of our in silico experiments for the interaction of 3OC12-HSL with the LasR LBD is comparable with the results of other authors, although currently there are no studies investigating the DBD.

This work is connected to recent experimental studies concerning LasR regulation. It should be noted that the LasR quorum-sensing system of P. aeruginosa is regulated by pre-quorum and post-quorum regulating systems. The first system is controlled by post-translational repressors QscR, QslA and Qse. In the absence of 3OC12-HSL, they can each form a heterodimer with LasR. Structural and functional studies of bacterial QS anti-activators revealed three different modes of action for LuxR-type transcriptional regulation: the destruction of the LBD dimerization interface in P. aeruginosa, the occupation of the AHL binding pocket and binding to the DBD. All these interactions indirectly change the conformation of the DBD, thus allosterically preventing DNA binding. It was previously shown that QslA can bind to LasR and thereby inhibit its DNA-binding capability. Moreover, it has been shown that QslA suppresses formation of the LasR homodimer complex by occupying the same surface area that is used for dimerization.

It has also been shown that QscR interacts with 3OC12-HSL. QscR has multiple dimerization interfaces and possesses both an LBD and DBD. The heterodimer QslA-LasR affects the dimerization interface and prevents receptor activation and DNA promoter binding. According to one of the mechanisms, the anti-activator QslA blocks the regions that are necessary for transcription factors multimerization. Moreover, 3OC12-HSL does not overcome the repression of LasR exerted by QslA. This indicates that 3OC12-HSL interacts with helixes that form part of these dimerization interfaces. It may be that the multiple binding modes of 3OC12-HSL imply that LasR has multiple dimerization interfaces like QslA and QscR.

There is also another aspect to be discussed. One major strategy for LasR inhibition is the development of small-molecule antagonists that mimic the native autoinducer and inactivate LasR. However, the mechanism by which they inactivate LasR is unknown. It has been suggested that LasR interactions with these antagonists promote a robust, unnatural fold of LasR that...
does not permit DNA binding\textsuperscript{30}. Further research studying flavonoids as potential inhibitors shows that they do not bind in the ligand-binding pocket of LasR\textsuperscript{30}. It has also been shown that they prevent LasR from binding to promoter DNA, but do not stop dimerization\textsuperscript{30}. Unfortunately, in this study it was not possible to assay the LasR DBD due to insufficient protein yields\textsuperscript{31}. The author suggest that the AI and flavonoid inhibitors can simultaneously bind to LasR and therefore postulate that flavonoids do not use the canonical AI binding site. We suggest that flavonoids and other inhibitors may interact with the bridge, thereby preventing LasR from binding to promoter DNA.

Our model might explain the molecular mechanisms proposed by recent experimental studies involving the study of flavonoids and other inhibitors which suggest the possibility of another binding site\textsuperscript{30,31}. It may also explain why computational drug design has not been successful in targeting this protein. However, the lack of a full, experimentally-derived structure for LasR is problematic and requires further experimental and computational research. We expect that our molecular insights on LasR can shed more light on this protein and assist in the development of new treatments against \textit{P. aeruginosa}.

**Conclusion**

From the simulations, it appears that the AI ligand 3OC12-HSL can bind to both the LBD and the “bridge” of the transcriptional regulator LasR. The interaction with the “bridge” is a novel site. Binding energy analysis shows that the interaction of 3OC12-HSL with the “bridge” and its interaction with the LBD is not competitive. Conserved amino acids such as Leu236, Phe219, Leu177, Lys182 and Trp19 contribute the most during the interaction with “bridge”. This could suggest that the interaction of 3OC12-HSL with the “bridge” is necessary for the DNA binding capability of LasR. One possible explanation for these multiple binding sites is that LasR may have multiple dimerization interfaces. This study may reveal new insights into the interactions of the native 3OC12-HSL ligand with transcriptional regulator LasR in \textit{P. aeruginosa}. Results from this study may aid future drug development endeavours.

**Data availability**

**Underlying data**

All data underlying the results are available as part of the article and no additional source data are required.

**Extended data**

Zenodo: Interaction of N-3-oxododecanoyl homoserine lactone with transcriptional regulator LasR of \textit{Pseudomonas aeruginosa}.

Insights from molecular docking and dynamics simulations. [https://doi.org/10.5281/zenodo.25865526](https://doi.org/10.5281/zenodo.25865526)

This project contains the following extended data:

- AdditionalFile1_SupportingInformation.pdf (Figures S1-20)
- AdditionalFile2_ClustalW.pdf (Full sequence alignment of LasR using ClustalW algorithm)
- AdditionalFile3_ClustalOmega.pdf (Full sequence alignment of LasR using Clustal Omega algorithm)
- AdditionalFile4_Muscle.pdf (Full sequence alignment of LasR using Muscle algorithm)
- HSL_dock.zip (Molecular docking simulations input and output for AutoDock Vina)
- HSL_exhaustiveness.zip (Molecular docking simulations input and output for the determination of exhaustiveness (AutoDock Vina))
- LasR_HSL_multidock.zip (Molecular docking simulations input and output for AutoDock Vina, rDock and FlexAID)
- MD100ns.tar.gz (Molecular dynamics simulation data of LasR protein for 100 nanoseconds)
- HSL_LasR_simulation_1.zip (1st MD simulations data for 3OC12-HSL with LasR) HSL_LasR_simulation_2.zip (2nd MD simulation data for 3OC12-HSL with LasR)
- HSL_LasR_simulation_3.zip (3rd MD simulation data for 3OC12-HSL with LasR)
- scripts.zip (Scripts that were used for the analysis of the data based on the molmolpy set of scripts\textsuperscript{28})

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

**Grant information**

The authors gratefully acknowledge financial support by the Russian-Armenian University [NIR 25/15].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Acknowledgments**

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

**References**


22. UniProt Consortium: \textit{UniProt: a worldwide hub of protein knowledge}. \textbf{Publisher Full Text}


17. Lee J, Zhang L, \textit{et al.}: \textit{A cell-cell communication signal integrates quorum sensing and stress response}. \textbf{Publisher Full Text}


2. Waskom M, Botvinnik O, O’Kane D, \textit{et al.}: \textit{Mwaskom/Seaborn: V0.8.1 (September 2017)}. \textbf{Published Full Text}

1. Waskom M, Botvinnik O, O’Kane D, \textit{et al.}: \textit{Mwaskom/Seaborn: V0.8.1 (September 2017)}. \textbf{Published Full Text}


The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com