In silico Prediction and Docking of Tertiary Structure of LuxI, an Inducer Synthase of Vibrio fischeri

Mohammed Zaghlool Saeed Al-Khayyat*,1, Ammar Ghanem Ameen Al-Dabbagh2

Abstract

Background: LuxI is a component of the quorum sensing signaling pathway in Vibrio fischeri responsible for the inducer synthesis that is essential for bioluminescence.

Methods: Homology modeling of LuxI was carried out using Phyre2 and refined with the GalaxyWEB server. Five models were generated and evaluated by ERRAT, ANOLEA, QMEAN6, and Procheck.

Results: Five refined models were generated by the GalaxyWEB server, with Model 4 having the greatest quality based on the QMEAN6 score of 0.732. ERRAT analysis revealed an overall quality of 98.9%, while the overall quality of the initial model was 54%. The mean force potential energy, as analyzed by ANOLEA, were better compared to the initial model. Stereochemical quality estimation by Procheck showed that the refined Model 4 had a reliable structure, and was therefore submitted to the protein model database. Drug Discovery Workbench V.2 was used to screen 2700 experimental compounds from the DrugBank database to identify inhibitors that can bind to the active site between amino acids 24 and 110. Ten compounds with high negative scores were selected as the best in binding.

Conclusion: The model produced, and the predicted acetyltransferase binding site, could be useful in modeling homologous sequences from other microorganisms and the design of new antimicrobials.

Keywords: Docking, Homology modeling, LuxI, Quorum sensing

Introduction

Quorum sensing was first discovered in Vibrio fischeri (V. fischeri), a bacterium that symbiotically produces light with certain marine animals (1-2). It is a signal transduction pathway that provides cell-cell communication and acts as a mechanism for coordinating gene expression in response to cell density (3). This system exists in Gram-positive and Gram-negative bacteria and controls virulence, biofilm formation, and antibiotic production (4, 5, 6).

In V. fischeri, quorum sensing is regulated by two gene products: LuxI and LuxR (7). LuxI is an acylated-homoserine lactone (acyl-HSL) (8). The acyl portion of acyl-HSL is derived from fatty acid precursors conjugated to acyl carrier protein (acyl-ACP), and the HSL moiety is derived from S-adenosylmethionine (SAM). LuxI promotes the formation of an amide bond joining the acyl side chain from acyl-ACP to SAM. Lactonization of the ligated intermediate, with the subsequent release of methylthioadenosine (MTA), results in the formation of acyl-HSL (9).

Homology or comparative modeling of a protein is a method of structure prediction based on amino acid sequence similarity to closely-related known structures (10). Genome-large scale sequencing projects revealed millions of sequences that could not be analyzed by X-ray crystallography and NMR spectroscopy techniques due to time restraints and other technical difficulties (11, 12). Because of these difficulties, researchers utilize bioinformatics to model unknown protein structures (13). These approaches help to identify active sites, design ligands and mutants, predict antigenic epitopes, and determine...
protein functions (14-17). This study focuses on homology modeling to predict the three-dimensional (3D) structure of LuxI in V. fischeri (strain ATCC 7006 01 / ES114) (18). This protein may be a potential target for antimicrobial design.

Materials and Methods

Sequence retrieval, physiochemical properties, and secondary structure

The amino acid sequence of LuxI was obtained from the UniProt database available at http://www.uniprot.org/, (Accession number: P35328) and used in FASTA file format in the analysis. The physiochemical properties of LuxI were characterized using the ProtParam tool of the Expasy server (Biozentrum, University of Basel, Switzerland) at http://web.expasy.org/protparam/ (19). These parameters include molecular weight, amino acid composition, theoretical isoelectric point (pI), extinction coefficient, and instability index (20-21). The secondary structure was predicted by SSpro8 of SCRATCH, a program specialized to predict secondary and disordered regions (Donald Bren School of Informatics and Computer Sciences, California, USA) at: http://scratch.proteomics.ics.uci.edu/ (22). Structures determined using the method of Kabsch and Sander (23) were alpha-helix, 3-10-helix, extended strand, turn, bend, bridges, and the rest.

Functional domain prediction

MOTIF is a program to identify motifs from GenomeNet, Japan using the Pfam and Prosite data bases, at: http://www-genome.cs.titech.ac.jp/tools/motif/ which uses Pfam. Pfam is a data base of protein domain alignment derived from the protein sequence secondary database of the Swiss Institute of Bioinformatics (SWISS-Prot) and translated to nucleic acid. The secondary database is stored in the European Molecular Biology Laboratory database (TrEMBL) (24).

Homology modeling, refinement, and evaluation of the 3D structure

The protein tertiary structure was built by PHYRE2 (Protein Holomogy/analogy Recognition Engine version 2) from Imperial College London available at (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (25). The structures generated were refined by the GalaxyWEB server at the Computational Biology Lab in the Department of Biochemistry, Seoul National University (http://galaxy.seoklab.org/) (26). The refined models were evaluated by several validation tools to select the best model and assess the quality of that model. ERRAT is a protein structure verification algorithm for evaluating the progress of crystallographic model building and refining maintained by the National Health Institute, University of California, USA (http://services.mbi.ucla.edu/ERRAT/) (27). The Z-score was determined by the PROSA web tool from the Center of Applied Molecular Engineering, Division of Bioinformatics, University of Salzburg, Salzburg, Austria. It measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations found in native proteins and available at (28). The SWISS-MODEL workspace server was also used from Biozentrum, University of Basel, Switzerland (29). This workspace contains several evaluation tools integrated within ANOLEA of Pontifical Catholic University, Chile, the QMEAN6 server (Qualitative Model Energy ANalysis), which estimates the global and local quality of the models from Biozentrum, University of Basel, Switzerland, and Procheck from the European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, UK (30-32).

Submission of the model

The best refined 3D model was submitted into the protein model database (PMDB) (http://bioinformatics.cineca.it/PMDB) (33).

Molecular Docking

The compounds used to screen Acyl-HSL synthases inhibitors were obtained from the DrugBank database (http://www.drugbank.ca). This database is supported by the Canadian Institutes of Health Research, Alberta Innovates - Health Solutions, and by The Metabolomics Innovation Centre (TMIC), Canada (34). Docking of the compounds was performed in CLC Drug Discovery workbench 2.0 (CLC Bio, QIAGEN Company, Denmark). Each compound was subjected to 100 iterations.

Results

To predict functional motifs and domains in LuxI, MOTIF software was used. Results show that LuxI has
an autoinducer synthetase family signature domain and contains an acetyltransferase domain between residues 24 and 110 (Fig. 1A).

SSpro8 adopts full DSSP-8 classification of the secondary structure (23). The secondary structure (Fig. 1B) predicted that 30.05% of the protein is comprised of \( \alpha \)-helices, 28.5% of extended \( \beta \)-strands, 15\% of \( \beta \)-turns, 2.1\% of 3-10 helices, 2.1\% of bends, and the remaining 22.28\% as random coils.

**Fig. 1.** (A) The amino acid sequence of LuxI in *V. fischeri* ES114 showing the acetyltransferase domain in red (B) The secondary structure predicted by SSpro8 where H: alpha-helix, G: 3-10-helix, E: extended strand, T: turn, S: bend, C: the rest.

The 3D structure of the protein was built by Phyre2 and refined by GalaxyWEB server. This server can detect unreliable regions and perform *ab initio* modeling to improve models (26). Five refined models were generated by GalaxyWEB server. Model 4 (Fig. 2) had the best quality according to the QMEAN6 scores (Table 1) and was submitted into the PMDB with the ID: PM0079876.

**Fig. 2.** The three-dimensional structure of LuxI, Model 4, produced by Phyre2 and refined by GalaxyWEB servers.
Table 1. The refined models produced by GalaxyWEB with their scores

<table>
<thead>
<tr>
<th>Model</th>
<th>ProSA Z-scores</th>
<th>ERRAT Quality (%)</th>
<th>QMEAN6 score</th>
<th>Procheck Ramachandran plot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Co1 (%)</td>
</tr>
<tr>
<td>Initial</td>
<td>-7.12</td>
<td>54.6</td>
<td>0.717</td>
<td>87.4</td>
</tr>
<tr>
<td>Model 1</td>
<td>-7.45</td>
<td>95.7</td>
<td>0.710</td>
<td>89.7</td>
</tr>
<tr>
<td>Model 2</td>
<td>-7.41</td>
<td>94.0</td>
<td>0.706</td>
<td>90.3</td>
</tr>
<tr>
<td>Model 3</td>
<td>-7.35</td>
<td>91.9</td>
<td>0.701</td>
<td>89.1</td>
</tr>
<tr>
<td>Model 4</td>
<td>-7.25</td>
<td>98.9</td>
<td>0.732</td>
<td>89.1</td>
</tr>
<tr>
<td>Model 5</td>
<td>-7.23</td>
<td>93.5</td>
<td>0.699</td>
<td>89.1</td>
</tr>
</tbody>
</table>

1Residues in the most-favored regions, 2residues in the additionally-allowed regions, 3residues in the generously-allowed regions, 4residues in the disallowed regions.

ERRAT is a novel method that can detect incorrect regions of protein structures according to errors leading to random distributions of atoms, which can be distinguished from correct distributions (27). Fig. 3 shows the refined Model 4 with quality of 98.913%, which is greater than the initial model containing many erroneous regions and a quality of 54.595%.

In ANOLEA profile (Fig. 4), the initial model had many areas of high energy, which were greatly improved in the refined model, suggesting greater reliability. The Z-scores of all the models are similar to the normal values commonly found in native structures determined by NMR spectroscopy and X-ray crystallography (Fig. 5).

Table 3 compares stereochemical parameters of the initial model and Model 4 where all Ramachandrans and Chi1-chi2 plots were better in Model 4 and no bad contacts. Ramachandran plot of the Model 4 is shown in Fig. 6.

Fig. 3. ERRAT plot shows error values for residues. The Y-axis represents the error value and the X-axis represents the amino acid residues of the protein model. An error value exceeding 99% confidence level indicates a poorly-modeled region (A) The initial model with quality of 54.595% (B) the refined Model 4 with quality of 98.913%.
Fig. 4. ANOLEA plots showing high energy zones as red (A) The initial model (B) The refined Model 4 with the high energy zones greatly minimized and improved.

Fig. 5. Z-score plot of LuxI (dot) determined by ProSA. The Z-score is -7.25, within the range of experimental native structures of similar sizes.
Fig. 6. Ramachandran plot of the fourth model is determined by Procheck. The most favoured regions are marked as A, B, and L. The additional allowed regions are marked as a, b, l, and p. All non-glycine and proline residues are shown as filled black squares, whereas glycines (non-end) are shown as filled black triangles. Disallowed residues are coloured red.

Table 2. The QMEAN6 and component scores and their Z-scores of the initial and fourth models with respect to experimental structures of similar sizes

<table>
<thead>
<tr>
<th>Scoring function term</th>
<th>Initial Model</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw score</td>
<td>Z-score</td>
</tr>
<tr>
<td>Cβ interaction energy</td>
<td>-69.61</td>
<td>-1.38</td>
</tr>
<tr>
<td>All-atom pairwise energy</td>
<td>-4084.92</td>
<td>-1.55</td>
</tr>
<tr>
<td>Solvation energy</td>
<td>-21.78</td>
<td>0.28</td>
</tr>
<tr>
<td>Torsion angle energy</td>
<td>-36.60</td>
<td>-1.22</td>
</tr>
<tr>
<td>Secondary structure agreement</td>
<td>83.9%</td>
<td>0.13</td>
</tr>
<tr>
<td>Solvent accessibility agreement</td>
<td>80.3%</td>
<td>0.19</td>
</tr>
<tr>
<td>QMEAN6 score</td>
<td>0.717</td>
<td>-0.53</td>
</tr>
</tbody>
</table>

Table 3. Stereochemical qualities of the initial and fourth models by Procheck

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial model</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Ramachandrans, out of 191</td>
<td>5 labeled residues</td>
<td>3 labeled</td>
</tr>
<tr>
<td>Chi1-chi2 plots, out of 121</td>
<td>3 labeled residues</td>
<td>0 labeled</td>
</tr>
<tr>
<td>Main-chain parameters</td>
<td>6 better</td>
<td>6 better</td>
</tr>
<tr>
<td>Side-chain parameters</td>
<td>5 better</td>
<td>5 better</td>
</tr>
<tr>
<td>Residue properties, bad contacts</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Overall G-factor</td>
<td>-0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>Main chain bond lengths</td>
<td>99.1% within limits</td>
<td>97.5% within limits</td>
</tr>
<tr>
<td>Main chain bond angles</td>
<td>89.9% within limits</td>
<td>88.6% within limits</td>
</tr>
<tr>
<td>Planar groups</td>
<td>100.0% within limits</td>
<td>97.3% within limits</td>
</tr>
</tbody>
</table>
During the last decade, quorum sensing system has been proposed as a target for antimicrobial agents and controlling the expression of virulence factors in bacteria (35). About 2700 compounds from the experimental library of DrugBank database were screened for ligands inhibitory to LuxI. The compounds with the ten highest scores are shown in Table 3.

**Discussion**

Many different LuxI-type proteins have been identified in Proteobacteria; these proteins are 190-230 amino acids in length and share 30-35% similarity. Ten residues are conserved in the terminal 110 amino acids with a conserved threonine at position 10 of the 110c-terminal amino acids. This threonine might be involved in stabilizing interactions with the fatty acyl biosynthetic precursors (36, 37).

The physiochemical properties of LuxI were computed using ProtParam. LuxI is comprised of 193 amino acids with a molecular weight of 22.014 kDa and a calculated pl of 5.7, allowing purification of the protein by isoelectric focusing (38). The extinction coefficient of LuxI at 280 nm is 23045 M$^{-1}$ cm$^{-1}$, assuming all cysteine pairs come from cystines, and 22920 M$^{-1}$ cm$^{-1}$ when all cysteines are reduced. The extinction coefficient indicates how much light a protein can absorb at a given wavelength. Estimation of this parameter is useful in spectrophotometric analysis of the protein (20). The instability index provides an estimate of the stability of the protein in a test tube. The instability index of this protein is 44.95, suggesting that it is slightly unstable (21).

In homology modeling, sequence identities greater than 40% can produce good overall quality models; however, if the target-template sequence identity is less than 40%, the predicted models will deviate significantly (39-40). Low sequence similarity and high structural divergence indicate the models may contain errors (41-42). Typical error sources are misplaced side chains, backbone distortions, alignment errors, or selecting a template of incorrect fold (43-45). Melo et al. (30) suggested that errors in the model are either in or close to regions that connect secondary structure core components and are of high energy.

In the process of protein structure prediction alternative models are generated, from which the most accurate model is selected using a scoring function (46). The scoring function relies on the principle that the native state of a protein has a minimum free energy (47). Various assessment tools have been developed (48), which aid in the assessment of problems that may arise in evaluation, such as whether the model has the correct fold, the overall geometric accuracy of the protein, and the geometric accuracy of individual protein regions (40).

The QMEAN6 scoring function consists of a linear combination of six descriptors. Two distance-dependent interaction potentials of mean force based on C$\beta$ atoms and all atom types are used to assess long range interactions; a torsion angle potential over three consecutive amino acids is applied to analyze the local backbone geometry of the model, a solvation potential to estimate the burial status of the residues, the agreement of the predicted and the calculated secondary structures, and solvent accessibility prediction. The raw score is between 0-1 (31). The QMEAN6 score of Model 4 was the greatest of the models tested. Because

### Table 4. Docking results of the two inducers against the experimental compounds in DrugBank

<table>
<thead>
<tr>
<th>Drug</th>
<th>DrugBank ID</th>
<th>Weight</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>K2C</td>
<td>DB08036</td>
<td>311.34</td>
<td>-98.74</td>
</tr>
<tr>
<td>Approved 873</td>
<td>DB03031</td>
<td>539.78</td>
<td>-95.90</td>
</tr>
<tr>
<td>740</td>
<td>DB07220</td>
<td>453.56</td>
<td>-88.63</td>
</tr>
<tr>
<td>Approved 1890</td>
<td>DB04502</td>
<td>600.73</td>
<td>-86.88</td>
</tr>
<tr>
<td>Approved 1155</td>
<td>DB03471</td>
<td>395.54</td>
<td>-86.74</td>
</tr>
<tr>
<td>Approved 292</td>
<td>DB02089</td>
<td>503.38</td>
<td>-83.62</td>
</tr>
<tr>
<td>4CP</td>
<td>DB07105</td>
<td>405.90</td>
<td>-80.82</td>
</tr>
<tr>
<td>4HD</td>
<td>DB07111</td>
<td>344.49</td>
<td>-79.95</td>
</tr>
<tr>
<td>GSK</td>
<td>DB07847</td>
<td>465.95</td>
<td>-75.82</td>
</tr>
</tbody>
</table>
QMEAN6 can provide both global and local (per residue) estimates of model quality (46), this model can be regarded as the best model obtained. Table 2 shows the contribution of each descriptor to the overall score. The pseudo-energies of the contributing terms are given together with their Z-scores with respect to scores obtained for high-resolution experimental structures of similar sizes solved by X-ray crystallography.

In the ProSA web tool the Z-score of a protein is defined as the energy separation between the native fold and the average of an ensemble of the misfolds in the units of standard deviation of the ensemble (49). A Z-score outside a range characteristic for native proteins of similar sizes indicates erroneous structure.

This score is displayed in a plot that contains the Z-scores of all experimentally-determined proteins (28).

A Ramachandran plot is an x-y plot of phi/psi dihedral angles between N-Cα and Cα-C planar peptide bonds in a protein’s backbone. According to Ramachandran plots analyzed for over 118 structures at 2.0 Å resolution, a good quality model can obtained when greater than 90% of residues fall into the most favored region (32).

The 3D structure of LuxI could be used to model inducers with homologous sequences in other microorganisms. This study also shows that the acetyltransferase site of action may be exploited in ligand design to inhibit quorum sensing in pathogenic organisms with homologous systems.

Acknowledgment

The author wishes to thank Dr. Raywa Zaghlool said at Technical institute of Mosul for advice and notes.

References

In silico Study of LuxI of Vibrio Fischeri


