IN SILICO DETERMINATION AND VALIDATION OF OPRL STRUCTURE AND B CELL EPITOPES AS VACCINE CANDIDATE IN PSEUDOMONAS AERUGINOSA

Asefeh Kofeiti¹, Mehdi kalantar², Fateme Sefid*³

¹Departament of biology, Science and Art University, Yazd, Iran.
²Shahid Sadoughi University of Medical Sciences and Health, Yazd, Iran.
³Department of Biology, Shahed University, Tehran-Qom Express Way, Iran.
Email: sefid@shahed.ac.ir

Received on 15-05-2016 Accepted on 18-06-2016

Abstract

Pseudomonas aeruginosa is an opportunistic, non-fermentative, gram-negative rod which is an important causes of nosocomial infection leading to septicemia and death. The mortality rate is higher than bacteremia’s caused by other gram-negative opportunistic pathogens. One of the most important features of the bacterium is its resistance to various antibacterial agents, and even newly developed antibiotics have failed to reduce the mortality rate associated with this organism. There is increasing interest in bacterial virulence factors as a basis for effective vaccines and immune therapies. Extracellular products and cell structures of P. aeruginosa contain numerous antigens with different specificity. These antigens have different roles for the formation of host protective immunity against P. aeruginosa. Due to the accessibility on the bacterial surface, LPS and OM proteins of P. aeruginosa are particularly important targets for vaccine studies. Thus identification of epitopes and functional sites on the surface of this protein is important for vaccine design against Pseudomonas aeruginosa. We predict OprL 3D structure with homology modeling approach. In this regard ElliPro predicts linear and discontinuous antibody epitopes based on a protein antigen's 3D structure. ElliPro associates each predicted epitope with a score, defined as a PI (Protrusion Index) value averaged over epitope residues. For each residue, a PI value is defined as percentage of the protein atoms enclosed in the ellipsoid, which approximates the protein surface, at which the residue first becomes lying outside the ellipsoid.

Keywords: Pseudomonas aeruginosa, Vaccine candidate, OprL, Bioinformatic

1. Introduction

Pseudomonas aeruginosa is a typical species of the genus Pseudomonas which is a member of the family Pseudomonadaceae(Sherman, Stefansson, Fox et al., 2001).
P. aeruginosais an opportunistic, non-fermentative, gram-negative rod which is an important cause of nosocomial infection leading to septicemia and death. The mortality rate is higher than bacteremias caused by other gram-negative opportunistic pathogens (Tamber & Hancock, 2004; Priebe & Goldberg, 2014).

One of the most important features of the bacterium is its resistance to various antibacterial agents, and even newly developed antibiotics have failed to reduce the mortality rate associated with this organism (Rawling, Martin, & Hancock, 1995; Rodriguez-Herva & Ramos, 1996). Results of bacteriological investigations conducted in different hospitals have shown Pseudomonas spp., mainly P. aeruginosa, to be among the most frequently isolated organisms from pus, sputum, blood and other clinical material of patients with Gram-negative hospital infections (Knapp, Hungerer, Broker et al., 1999; Qian, Wu, Muratova et al., 2007). P. aeruginosa is also associated with complications in patients with cystic fibrosis (CF), or following surgery, trauma, and thermal burn (Dilling & Pier, 2008).

There is increasing interest in bacterial virulence factors as a basis for effective vaccines and immunotherapies (Ochs, McCusker, Bains et al., 1999). Several extracellular products from P. aeruginosa such as exotoxin A, exoenzyme S, phospholipase and hemolysins have been studies as potential virulence factors (Worgall, Krause, Rivara et al., 2005). Extracellular products and cell structures of P. aeruginosa contain numerous antigens with different specificity. These antigens have different roles for the formation of host protective immunity against P. aeruginosa. A number of proteins with different MWs and functions have been located in the OM of P. aeruginosa (Sugawara, Steiert, Rouhani et al., 1996; Lim, De Vos, Brauns et al., 1997). Due to the accessibility on the bacterial surface, LPS and OM proteins of P. aeruginosa are particularly important targets for vaccine studies (Wessel, Liew, Kwon et al., 2013). The property of an antigen to bind specifically complementary antibodies is known as the antigen’s antigenicity; likewise, the ability of an antigen to induce an immune response is called its immunogenicity (Ponomarenko & Van Regenmortel, 2009; Blythe & Flower, 2005). Attempts should be made to discover peptides that could mimic protein epitopes and possess the same immunogenicity as the whole protein. Subsequently, theoretical methods for epitope prediction have been developed leading to synthesis of such peptides that are important for development of immunodiagnostic tests and vaccines (Greenbaum, Andersen, Blythe et al., 2007; Yasser & Honavar, 2013; Larsen, Lund, & Nielsen, 2006; Chen, Liu, Yang et al., 2007; Haste Andersen, Nielsen, & Lund, 2006). The present study was designed to in silico resolving the major obstacles in the control or in prevention of the diseases caused by P. aeruginosa. We exploited bioinformatic tools to better understanding and characterizing the oprL structure of P. aeruginosa and select appropriate regions as effective B cell epitops.
2. Methods

2.1. Sequence availability and homology search

The oprlreference protein sequence with accession No. NP_249664.1 acquired from NCBI at http://www.ncbi.nlm.nih.gov/protein was saved in FASTA format for further analyses. The sequences served as a query for protein BLAST at http://blast.ncbi.nlm.nih.gov/Blast.cgi against non-redundant protein database. Probable putative conserved domains of the query protein were also searched for, at the above address.

2.2. Template search

The query protein sequences were used as an input data for the PSI-BLAST against protein data bank (PDB) at http://blast.ncbi.nlm.nih.gov/Blast.cgi to identify its homologous structures.

2.3. Primary sequence analysis

protparam(Gasteiger, Hoogland, Gattiker et al., 2005) online software at http://expasy.org/tools/protparam.html was employed for estimation and determination of properties such as molecular weight, theoretical pI, amino acid composition, total number of negatively and positively charged residues, instability index and aliphatic index.

2.4. Subcellular localization

Subcellular localization of the protein was predicted by CELLO (Gardy & Brinkman, 2006) at http://cello.life.nctu.edu.tw/.

2.5. Topology and signal peptide prediction

SignalP 4.1 server (Dyrlov Bendtsen, Nielsen, von Heijne et al., 2004) at http://www.cbs.dtu.dk/services/SignalP/ was invoked to predict the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms. The method incorporates a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks. SPOCTOPUS(Viklund & Elofsson, 2008) at http://spoctopus.cbr.su.se/ was also employed to determine membrane protein topology and signal peptides.

2.6. Secondary structure prediction

Secondary structure of the protein was predicted by PSIPRED(McGuffin, Bryson, & Jones, 2000) at http://bioinf.cs.ucl.ac.uk/psipred/. The PSIPRED Protein Sequence Analysis Workbench aggregates several UCL structure prediction methods into one location.

Phyre2(Kelley, Mezulis, Yates et al., 2015) server at http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index employed to validate PSIPRED predictions.
2.7. 3D structure prediction

The SWISS-MODEL (Guex & Peitsch, 1997) Workspace at http://swissmodel.expasy.org/ is a web-based integrated service dedicated to protein structure homology modeling. It assists and guides the user in building protein homology models at different levels of complexity.

Building a homology model comprises four main steps: identification of structural template(s), alignment of target sequence and template structure(s), model building, and model quality evaluation. These steps can be repeated until a satisfying modeling result is achieved.

Each of the four steps requires specialized software and access to up-to-date protein sequence and structure databases.

2.8. Models evaluations

All 3D models of the proteins built were qualitatively estimated by GMQE and QMEAN4 (Benkert, Tosatto, & Schomburg, 2008) scores.

2.9. Identification of functionally and structurally important residues

Consurf (Glaser, Pupko, Paz et al., 2003) program at http://consurf.tau.ac.il/ was used for annotating functional residues of protein structure in twilight zone.

2.10. Ligand binding site predictions

Cofactor (Roy, Yang, & Zhang, 2012) at http://zhanglab.ccmb.med.umich.edu/COFACTOR/ is a structure-based method for biological function annotation of protein molecules. Important amino acid involved in ligand binding site is predicted by this server.

2.11. Single-scale amino acid properties assay

IEDB (Zhang, Wang, Kim et al., 2008) at http://tools.immuneepitope.org/tools/bcell/iedb_input parameters such as hydrophilicity, flexibility, accessibility, turns and antigenic propensity of polypeptide have been correlated with the location of B cell epitopes.

This has led to a search for empirical rules that would allow the position of B cell epitopes to be predicted from certain features of the protein sequence.

2.12. B cell epitope prediction

3. Result and Discussion

3.1. Sequence availability and homology search

The protein sequence with 168 residue obtained from NCBI and saved in FASTA format. Protein sequence serving as query for BLAST produced a set of sequences as the highest similar sequence. BLAST search revealed numerous hits to the oprL subunit sequence.

All hits were of Pseudomonas. Putative conserved domains were detected within this sequence and are shown in Figure 1.

![Figure 1: Putative conserved domains have been detected. Most of the sequences belong to ompA_C-like superfamily.](image)

3.2. Template search

PSI-BLAST against protein data bank (PDB) results displayed several hits as homologous structures. The first hit possessing the highest score was selected as a template for homology modelling. The first hit (Accession: 4G4V-A, Max score: 134, Query coverage: 59%, Max ident: 63%) possessing the highest score was selected as a template.

3.3. Primary sequence analysis

The protein sequence served as input for the computation of various physical and chemical parameters. The computed parameters included the molecular weight, theoretical pI, instability index, aliphatic index and grand average of hydropathicity (indicates the solubility of the proteins: positive GRAVY (hydrophobic), negative GRAVY (hydrophilic)) are listed below.

- Number of amino acids: 168
- Molecular weight: 17925.0
- Theoretical pI: 5.95
- Total number of negatively charged residues (Asp + Glu): 23
- Total number of positively charged residues (Arg + Lys): 21

Atomic composition:
<table>
<thead>
<tr>
<th>Element</th>
<th>Formula</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td>C</td>
<td>774</td>
</tr>
<tr>
<td>Hydrogen</td>
<td>H</td>
<td>1235</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>N</td>
<td>229</td>
</tr>
<tr>
<td>Oxygen</td>
<td>O</td>
<td>249</td>
</tr>
<tr>
<td>Sulfur</td>
<td>S</td>
<td>6</td>
</tr>
</tbody>
</table>

**Formula:** $C_{774}H_{1235}N_{229}O_{249}S_6$

**Total number of atoms:** 2493

**Extinction coefficients:**

Extinction coefficients are in units of $M^{-1} \text{ cm}^{-1}$, at 280 nm measured in water.

**Ext. coefficient** 14440

Abs 0.1% (=1 g/l) 0.806, assuming all pairs of Cys residues form cystines

**Ext. coefficient** 14440

Abs 0.1% (=1 g/l) 0.806, assuming all Cys residues are reduced

**Estimated half-life:**

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).

> 20 hours (yeast, in vivo).

> 10 hours (Escherichia coli, in vivo).

**Instability index:** The instability index (II) is computed to be 25.79

This classifies the protein as stable.

**Aliphatic index:** 73.27

**Grand average of hydropathicity (GRAVY):** -0.432

### 3.4. Subcellular localization

**oprF** Subcellular localization predicted by CELLO was outer membrane with the highest reliability index (4.016).

### 3.5. Topology and signal peptide prediction

OprL signal peptide cleavage site was predicted between positions 19 and 20 of protein sequence. Topology and signal peptide prediction of SPOCTOPUS server is shown in Figure 2.
3.6. Secondary structure prediction

Secondary structure of the proteins was predicted by PSIPRED. Coil, helix and strands are components constituting secondary structure of the proteins. The secondary structure could be used to validate the tertiary structures. Phyre2 secondary prediction result is shown in Figure 3.

Figure 3. Phyre 2 secondary prediction.
3.7. 3D structure prediction

Building a homology model comprises four main steps: identification of structural template(s), alignment of target sequence and template structure(s), model building, and model quality evaluation. These steps can be repeated until a satisfying modelling result is achieved.

Each of the four steps requires specialized software and access to up-to-date protein sequence and structure databases. Swiss model software recruited for homology modeling introduced 3 model. All the models were selected for further analyses.

3.8. Models evaluations

The 3D models estimated qualititatively by tow servers revealed that there was a consensus on a single model. Results are shown in Table 1.

QMEAN is a composite scoring function for the estimation of the global and local model quality. QMEAN consisting of four structural descriptors: The local geometry is analyzed by a torsion angle potential over three consecutive amino acids. Two pairwise distance-dependent potentials are used to assess all-atom and C-beta interactions. A solvation potential describes the burial status of the residues. The pseudo energies returned from the four structural descriptors and the final QMEAN4 score get directly related to what we would expect from high resolution X-ray structures of similar size using a Z-score scheme.

The score of a model in also shown in relation to a set of high-resolution PDB structures (Z-score). The plot relates the obtained global QMEAN4 value to scores calculated from a set of high-resolution X-ray structures.

Local estimates of the model quality based on the QMEAN scoring function are shown as per-reside plot. Each residue is assigned a reliability score between 0 and 1, describing the expected similarity to the native structure. Higher numbers indicate higher reliability of the residues. GMQE (Global Model Quality Estimation) is a quality estimation which combines properties from the target-template alignment. The resulting GMQE score is expressed as a number between zero and one, reflecting the expected accuracy of a model built with that alignment and template. Higher numbers indicate higher reliability. Once a model is built, the GMQE gets updated for this specific case by also taking into account the QMEAN4 score of the obtained model in order to increase reliability of the quality estimation.

3.9. Identification of functionally and structurally important residues

Consurf annotated functional residues on the 3D structure of oprF in twilight zone. Results are shown in figure 4.
3.10. ligand binding site predictions

Ligand binding sites determined using COFACTOR software, indicate involvement of conserved residues include 31, 32, 65, 67, 72 and 118 in binding site with the highest Cscore\textsuperscript{LB} (the confidence score of predicted binding site). The calculated BS-score for this predicted binding site was 1.62.

BS-score is a measure of local similarity (sequence & structure) between template binding site and predicted binding site in the query structure. Based on large scale benchmarking analysis, observed that a BS-score $>1$ reflects a significant local match between the predicted and template binding site. (Figure 5).

**Table 1. OprL predicted 3D models estimated qualititatively.**

<table>
<thead>
<tr>
<th>model</th>
<th>figure</th>
<th>Seq Identity</th>
<th>Seq Similarity</th>
<th>Coverage</th>
<th>GMQE</th>
<th>QMEAN4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td>42.86%</td>
<td>0.41</td>
<td>0.75</td>
<td>0.58</td>
<td>-3.39</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Image" /></td>
<td>50.40%</td>
<td>0.43</td>
<td>0.74</td>
<td>0.57</td>
<td>-1.97</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Image" /></td>
<td>50.40%</td>
<td>0.43</td>
<td>0.74</td>
<td>0.56</td>
<td>-2.82</td>
</tr>
</tbody>
</table>

The conservation scale:

```
0 1 2 3 4 5 6 7 8 9
```

Variable    Average    Conserved

**Figure-4: ConSurf, identification of functionally and structurally important residues Results.**
Figure-5: Cofactor ligand binding site prediction. involvement of conserved residues in binding site.

3.11. single-scale amino acid properties assay

IEDB server predict several properties such as hydrophilicity, accessibility, antigenicity, flexibility and beta turn secondary structure in the protein sequence.

Propensity scale methods assign a propensity value to each amino acid which measures the tendency of an amino acid to be part of a B-cell epitope (as compared to the background).

To reduce fluctuations, the score for each target amino acid residue in a query sequence is computed as the average of the propensity values of the amino acids in a sliding window centered at the target residue. hydrophilicity, accessibility, antigenicity, flexibility and secondary structure properties have fundamental role in B cell epitope prediction. Relying on just one of these properties, reliable results could not be achieved. Results are shown in figure 6.

3.12. Prediction of B cell epitopes by integrated strategy

When a living organism encounters a pathogenic virus or microbe, the B cells of the immune system recognize the pathogen’s antigens by their membrane-bound immunoglobulin receptors and, in response, produce antibodies specific to these antigens.

The term antigen refers to any entitya cell, a macromolecular assembly, or a moleculethat may be bound by either a B-cell receptor or an antibody molecule. The binding portion of an antigen is called a B-cell epitope or an antigenic determinant.

If an antigen is a protein, an epitope may be either a short peptide from the protein sequence or a patch of atoms on the protein surface in the three-dimensional space.

Five linear along with 8 discontinuous B cell epitopes were predicted by ElliPro software (Table 2,3). Two discontinuous and 2 linear epitopes with the highest PI (protrusion index) are shown in Figure-7.
Table 2. Linear Epitopes Predicted by Ellipro.

<table>
<thead>
<tr>
<th>No.</th>
<th>Start</th>
<th>End</th>
<th>Peptide</th>
<th>Number of residues</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>138</td>
<td>168</td>
<td>LELVSYGKERPVATGHDEQSWAQNRRVELKK</td>
<td>31</td>
<td>0.837</td>
</tr>
<tr>
<td>2</td>
<td>93</td>
<td>111</td>
<td>GSGQRVVLEGHTDERGTRE</td>
<td>19</td>
<td>0.682</td>
</tr>
<tr>
<td>3</td>
<td>75</td>
<td>85</td>
<td>DLKPEAMRALD</td>
<td>11</td>
<td>0.665</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>64</td>
<td>GSLSDEAALRAI</td>
<td>12</td>
<td>0.625</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>44</td>
<td>PNAGY</td>
<td>5</td>
<td>0.558</td>
</tr>
</tbody>
</table>

Table 3: Discontinuous Epitopes Predicted by Ellipro.

<table>
<thead>
<tr>
<th>No.</th>
<th>Residues</th>
<th>Number of residues</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>_:R163, _:V164, _:E165, _:L166</td>
<td>4</td>
<td>0.97</td>
</tr>
<tr>
<td>2</td>
<td>_:W158, _:A159, _:Q160, _:N161</td>
<td>4</td>
<td>0.92</td>
</tr>
<tr>
<td>3</td>
<td>_:E155, _:Q156, _:S157</td>
<td>3</td>
<td>0.893</td>
</tr>
<tr>
<td>4</td>
<td>_:Q137, _:E139, _:L140, _:V141, _:S142, _:Y143, _:G144, _:K145, _:E146, _:R147, _:P148, _:V149, _:A150, _:T151, _:G152, _:H153, _:D154</td>
<td>17</td>
<td>0.756</td>
</tr>
<tr>
<td>5</td>
<td>_:A83, _:L84, _:D85</td>
<td>3</td>
<td>0.73</td>
</tr>
<tr>
<td>6</td>
<td>_:G53, _:S54, _:L55, _:S56, _:D57, _:E58, _:A59, _:A60, _:L61, _:A63, _:I64, _:G93, _:S94, _:G95, _:Q96, _:R97, _:V98, _:V99, _:L100, _:E101, _:G102, _:H103, _:T104, _:D105, _:E106, _:R107, _:T109, _:R110, _:E111</td>
<td>29</td>
<td>0.675</td>
</tr>
<tr>
<td>7</td>
<td>_:G33, _:A34, _:G36, _:G37, _:P40, _:N41, _:A42, _:G43, _:Y44, _:G45, _:D75, _:L76, _:P78, _:E79, _:A80, _:M81, _:R82</td>
<td>17</td>
<td>0.599</td>
</tr>
<tr>
<td>8</td>
<td>_:K25, _:G26, _:G27, _:A29, _:S30</td>
<td>5</td>
<td>0.599</td>
</tr>
</tbody>
</table>

[Graphs showing hydrophilicity and accessibility]
**Fig 6.** IEDB linear B cell epitope prediction results for oprF protein.

**Fig 7.** From left to right, 2 linear and 2 discontinuous epitopes with the highest PI score predicted by Ellipro server are shown. Epitopes mapped on 3D models using Discovery Studio Visualizer 2.5.5 software.

**References**


14. Google Patents. Ref Type: Generic


**Corresponding Author:**

Fateme Sefid*

**Email:** sefid@shahed.ac.ir