IN SILICO POTENTIAL VACCINE AGAINST OPRF IN MULTIDRUG-RESISTANT PSEUDOMONAS AERUGINOSA

Asefeh Kofeiti¹, Mehdi kalantar², Fateme Sefid³*
¹Departemen 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

Abstract
Pseudomonas aeruginosa is an important cause of nosocomial infection and may lead to septicemia and death. P. aeruginosasepticaemia is associated with the highest mortality rate of all Gram-negative infections. Because of the general resistance of the organism to antibiotics, research has been focused on immunotherapy. There are several bacterial cell components incorporated into subunit vaccines. Vaccine studies have often focused on lipopolysaccharide (LPS) and the outer membrane proteins (OPRs) due to its potent stimulation of the immune response. The major OPRs, OprF interested in the potential of OPRs as vaccines. Determination of OprF tertiary structure and theoretical methods for epitope prediction has been led to synthesis of such peptides that are important for immunodiagnostic tests and vaccines. Bioinformatic tools to better understanding and characterizing the oprF structure of P. aeruginosa was used. For homology modeling, BLAST was run on the sequence in order to find the best template. The template was then served to model the 3D structure. Also, Secondary structure of the protein was predicted. Moreover, topology, signal peptide and B cell epitopes of oprF were predicted. In conclusion, protein epitopes were selected as vaccine candidates. These regions contain functional exposed amino acids with higher properties score of B cell epitopes. In these regions, the majority of amino acids are hydrophile, flexible, accessible, and favorable for B cells with a view to point of secondary structure.

Keywords: Pseudomonas aeruginosa, Vaccine candidate, OprF, Bioinformatic

1. Introduction
Pseudomonas aeruginosa continues to be a cause of life-threatening infections in patients treated in intensive care units(Wessel, Liew, Kwon et al., 2013; Glaser, Pupko, Paz et al., 2003). Whereas healthy individuals are not in
general susceptible to infection by this organism, the risk to granulocytopenic patients, such as cancer patients undergoing cytostatic therapy or transplant patients receiving immunosuppressive (with extensive burns, for instance, or recovering from major surgery) is dramatically increased (Worgall, Krause, Rivara et al., 2005; Lim, De Vos, Brauns et al., 1997; Glaser et al., 2003). The origin of the invading microorganisms is most frequently the patient’s own microflora. Damage to the first line of defense, such as the skin or the mucous membranes, enables the colonizing bacteria to enter the bloodstream and cause septicemia (Ochs, McCusker, Bains et al., 1999; Sugawara, Steiert, Rouhani et al., 1996; Glaser et al., 2003).

P. aeruginosasepticaemia is associated with the highest mortality rate of all Gram-negative infections. Because of the general resistance of the organism to antibiotics, research has been focused on immunotherapy (D-ring & Pier, 2008; Glaser et al., 2003).

The two major antigenic surface-associated components of P. aeruginosa are the lipopolysaccharides (LPS) and the outer membrane proteins (OPRs). LPS-based vaccines have been successfully tested in animal models as well as in clinical trials. However, the severe side effects observed in vaccinated individuals have made it necessary to develop subunit vaccines (Qian, Wu, Muratova et al., 2007; Glaser et al., 2003).

The major OPRs, OprF interested in the potential of OPRs as vaccines, because OPRs have been shown to be highly conserved and antigenically related in all 17 serogroups of the International Antigenic Typing Scheme(Knapp, Hungerer, Broker et al., 1999; Glaser et al., 2003).

Due to the accessibility on the bacterial surface, LPS and OM proteins of P. aeruginosa are particularly important targets for vaccine studies (Tamber & Hancock, 2004; Priebe & Goldberg, 2014; Sherman, Stefansson, Fox et al., 2001; Glaser et al., 2003).

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 (Rawling, Martin, & Hancock, 1995; Rodriguez-Herva & Ramos, 1996; Ponomarenko & Van Regenmortel, 2009; Glaser et al., 2003).

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(Blythe & Flower, 2005; Greenbaum,
Andersen, Blythe et al., 2007; Yasser & Honavar, 2013; Larsen, Lund, & Nielsen, 2006; Chen, Liu, Yang et al., 2007; Haste Andersen, Nielsen, & Lund, 2006; Zhang, Wang, Kim et al., 2008a; Glaser et al., 2003). 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 oprF structure of P. aeruginosa and select appropriate regions as effective B cell epitops.

2. Methods

2.1. Sequence availability and homology search

The oprf protein sequence with accession No. WP_004885687.1 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(Glaser et al., 2003; 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(Glaser et al., 2003; 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 modelling. 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 modelling 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 qualititatively estimated by GMQE and QMEAN4 (Benkert, Tosatto, & Schomburg, 2008) scores.

2.9. Identification of functionally and structurally important residues

Consurf(Glaser 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 et al., 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., 2008b) 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 344 residues 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 oprF subunit sequence. All hits were of Pseudomonas. Putative conserved domains were detected within this sequence. Most of the sequences belong to ompA_C-like (Peptidoglycan binding domains similar to the C-terminal domain of outer-membrane protein OmpA). OmpA-like domains (named after the C-terminal domain of OmpA protein) have been shown to non-covalently associate with peptidoglycan, a network of glycan chains composed of disaccharides, which are crosslinked via short peptide bridges.

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: 4RLC-A, Max score: 148, Query coverage: 44%, Max ident: 50%) 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 summarized in table1.

Table1. OprF various physical and chemical parameters.

<table>
<thead>
<tr>
<th>Residuenumber</th>
<th>Molecular weight</th>
<th>Theoretical pI</th>
<th>Instability index</th>
<th>Aliphatic index</th>
<th>GRAVY</th>
</tr>
</thead>
<tbody>
<tr>
<td>oprF</td>
<td>344</td>
<td>36549.1</td>
<td>4.69</td>
<td>30.77 (stable)</td>
<td>69.74</td>
</tr>
</tbody>
</table>
### 3.4. Subcellular localization

OprF subcellular localization predicted by CELLO was outer membrane with the highest reliability index (4.848).

### 3.5. Topology and signal peptide prediction

OprF signal peptide cleavage site was predicted between positions 24 and 25 of protein sequence. Topology and signal peptide prediction of SPOCTOPUS server is shown in Figure 1.

![Topology prediction of SPOCTOPUS server](image)

- **Figure 1.** Topology prediction of SPOCTOPUS server.

  **Topology output:** This is a graphic representation of the most likely topology as predicted by SPOCTOPUS.

  **Network output:** The two diagrams show the estimated preference for each residue to be located in different structural regions. The top diagram shows the preference of being either in:

  - the hydrophobic part of the membrane, 0-13Å from the membrane center (M)
  - the membrane water-interface, 11-18Å from the membrane center (I)
  - a close loop region, 13-23Å from the membrane center (L)
  - a globular region, further than 23Å from the membrane (G)

  The bottom diagram shows the estimated preference of a particular residue to be located either on the inside (i) or on the outside (o) of the membrane.

  The raw data underlying these two plots can be found in the SPOCTOPUS network file.

### 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.
Attribution of secondary structure components in the protein is alpha helix (19%), beta strand (35%) and random coil (46%).

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 models. All the models were selected for further analyses.

3.8. Models evaluations

The 3D models estimated qualititatively by two servers revealed that there was a consensus on a single model. Results are shown in Table-2. 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-residue 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.
Table-2: OprF predicted 3D models estimated qualititatively.

<table>
<thead>
<tr>
<th>model</th>
<th>figure</th>
<th>Template</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>4rlc.1.A</td>
<td>53.33%</td>
<td>0.46</td>
<td>0.44</td>
<td>0.29</td>
<td>-3.60</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="image" /></td>
<td>4rlc.1.A</td>
<td>50.99%</td>
<td>0.45</td>
<td>0.44</td>
<td>0.26</td>
<td>-4.23</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="image" /></td>
<td>2k0l.1.A</td>
<td>18.71%</td>
<td>0.28</td>
<td>0.45</td>
<td>0.25</td>
<td>-9.27</td>
</tr>
</tbody>
</table>

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 2.

3.10. ligand binding site predictions

Ligand binding sites determined using COFACTOR software, indicate involvement of conserved residues include 31, 32, 70, 72, 87 and 112 in binding site with the highest Cscore\textsuperscript{LB} (the confidence score of predicted binding site) (Figure 3).

![image](image4.png)

The conservation scale:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Average</td>
<td>Conserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure-2. ConSurf, identification of functionally and structurally important residues Results.**
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 4.

3.12. Prediction of B cell epitopes by integrated strategy

Four linear along with 11 discontinuous B cell epitopes were predicted by ElliPro software (Table 3,4). Two discontinuous and 2 linear epitopes with the highest PI (protrusion index) are shown in Figure 5.

Table-3: 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>209</td>
<td>344</td>
<td>CPDTPANVTVDADGCPAVAEVVRVELDVKFDFDKSVVKPSYSYGDIKNLADFMQQYPQTSTTVEGHTDSVGPDAYNQKLSERRANAVKQVLNQYGVGASRVNSVGYGESRPVADNATESGRAVNRRVVEAEVEAQAK</td>
<td>136</td>
<td>0.792</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>7</td>
<td>MKLKNTL</td>
<td>7</td>
<td>0.592</td>
</tr>
<tr>
<td>3</td>
<td>156</td>
<td>165</td>
<td>QYNIDQGNTE</td>
<td>10</td>
<td>0.583</td>
</tr>
<tr>
<td>4</td>
<td>117</td>
<td>132</td>
<td>QSIGQDARGGRDGSTF</td>
<td>16</td>
<td>0.555</td>
</tr>
</tbody>
</table>
Table 4: 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>_:R334, _:E335, _:A337, _:E338, V339, E340, A341, Q342, A343, K344</td>
<td>11</td>
<td>0.983</td>
</tr>
<tr>
<td>2</td>
<td>R318, P319, V320, A321, D322, N323, A324, T325, E326, S327, G328, R329, A330, V331, N332</td>
<td>15</td>
<td>0.942</td>
</tr>
<tr>
<td>3</td>
<td>Y302, G303, V304, G305, A306, S307, R308, V309, N310, S311, V312, G313, Y314, G315, E316, S317</td>
<td>16</td>
<td>0.862</td>
</tr>
<tr>
<td>4</td>
<td>M260, Q261, V262, P263, D264, Q265, T266, S267, T268, T269, V270, E271, G272, H273, T274, D275, S276, V277, G278, P279, D280, A281, Y282, N283, Q284</td>
<td>25</td>
<td>0.846</td>
</tr>
<tr>
<td>5</td>
<td>Q296, V297, L298, V299, N300, Q301</td>
<td>6</td>
<td>0.772</td>
</tr>
<tr>
<td>6</td>
<td>L286, S287, E288, R290, A291, N292, A293, V294, K295</td>
<td>9</td>
<td>0.759</td>
</tr>
<tr>
<td>7</td>
<td>Y250, G251, D252, I253, K254, N255, L256, A257, D258, F259</td>
<td>10</td>
<td>0.75</td>
</tr>
<tr>
<td>8</td>
<td>V232, E233, L234, D235, V236, K237</td>
<td>6</td>
<td>0.684</td>
</tr>
<tr>
<td>9</td>
<td>F238, D239, F240, D241, K242, S243, V244, V245, K246, P247, S248, S249</td>
<td>12</td>
<td>0.641</td>
</tr>
<tr>
<td>11</td>
<td>K4, N5, T6, L7, G8, Y40, D42, S43, D44, Q117, S118, I119, G120, Q121, D122, A123, R124, G125, G126, R127, D128, G129, S130, T131, F132, Q156, Y157, N158, I159, D160, Q161, G162, N163, T164, E165</td>
<td>35</td>
<td>0.548</td>
</tr>
</tbody>
</table>

[Graphs showing hydrophilicity, accessibility, antigenicity, and flexibility]
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*3
Email: sefid@shahed.ac.ir