Homology modeling – a step towards vaccine development by analyzing structure of haemophilus influenza protein, transcriptional regulator H10994

Anila Farid
Lecturer, Department of Biochemistry, Abbottabad International Medical College, Abbottabad

Dr Madeeha Jadoon
Assistant Professor, Department of Biochemistry, Women Medical and Dental College, Abbottabad
Corresponding author email: Dr.jadoon73@gmail.com

Dr Sofia Shoukat
Assistant Professor, Department of Biochemistry, Ayub Medical College, Abbottabad

Prof Dr Uzma Faryal
Professor and HOD, Department of Biochemistry, Women Medical and Dental College, Abbottabad

Dr Bilal Karim
Assistant Professor Biochemistry, Islamabad Medical and Dental College, Islamabad

Dr Anwar Shahzad
Associate Professor, Community Medicine and Public Health AIMIT, Abbottabad

Abstract---Introduction: Haemophilus influenza is a type of bacterium that is non motile, gram negative and causes poisoning and infection including pneumonia,bronchitis etc. In order to study the resistivity of H.influenza protein, transcription regulator: H10433, homology modeling is an important step to predict structure. Material and methods: Bioinformatics such as CMR, BLAST, modeller Prcheck and Prosa was carried out to find 3D structure of protein. Results and Discussion: H.influenza has 1792 proteins. Out of these, 456 hypothetical proteins were found. Homology modeling of transcriptional regulator H10994 was done it consists of 8 helices and
7 beta sheets. Ramachandran plot has shown that it consists of 95.2% particles in maximum allowed regions, 2.9% particles in fewer allowed region, 1.4% particles in inadequate allowed region, 5% particles in disallowed region. Conclusion: By homology modeling of *H. influenza*, transcriptional regulator protein (HI0433), structure was designed which has provided enough information for vaccine development to control its transcription for causing disease.

**Keywords**---Haemophilus influenza, Homology modeling, vaccine development.

**Introduction**

Haemophilus influenza is a type of bacterium that is non motile, gram negative and causes poisoning and infection including pneumonia, bronchitis etc. It is transmitted through respiratory droplets or from close contact. Naturewise, It is generally aerobic. Different types of H.influenza are known. Mostly six or seven types are talked about. Numerous antibiotics like ceftriaxone and ceftotoxamine are introduced into the blood when infection is too serious. It exhibits much resistance to Macrolide. H.influenzae consists of many strains which are highly resistant to antibiotics. Thus making treatment highly challenging. Bioinformatics aids in analyzing large scale genomic data and understanding different purposes of proteins.

In order to study the resistivity of H.influenza protein, transcription regulator: HI0433, homology modeling is an important step to predict structure. It is also known as comparative modeling which is preferred to predict 3D structure of protein. It is based on its amino acid sequence and known structure of protein (template). It gives important evidence that how resistivity can be reduced and vaccine is developed to cure infection effectively. Homology modeling is very fast technique to make or construct protein models. Template is used to build models of target proteins. Protein Data Bank (PDB) helps in searching template.

The process involves aligning target with template sequence and then using that alignment for unknown protein model building. The level of resemblance between template and target sequence determines the accuracy of homology modeling. The predicted structure is more reliable if similarity is high. A protein transcriptional regulator is responsible for infection. So it is selected for homology modeling in order to develop vaccine against it. Homology modeling helped a lot in discovery of ligands to modulate their activity. Recognition of template from PDB data bank of known protein, template and query sequence alignment, building of model on alignment basis and last refinement of built model are four steps for homology modeling.

**Material and Methods**

Complete protein sequence of Haemophilus influenzae was downloaded from database of CMR (comprehensive microbial resources (http://cmr.jcvi.org). It is free website which put on view details entire prokaryotic genome. H.I contains
many proteins which do not show clear function (hypothetical protein). We distinct hypothetical protein from total genome and put aside in word document format.

Interproscan is tool helps in visualizing function of hypothetical protein sequence. Different structure of proteins were modeled after function annotation of protein. In model building, template plays essential part. Blast against PDB (Protein data bank) helps in finding template. The template which has maximum identity, Low E-value, high query coverage is preferred to be selected with equivalent target proteins.

Target template alignment file helps in building target protein model. Blast helps in target and template alignment. Modeller 9.10 program is used to construct 3D homology model of protein. DS viewer helps in visualizing newly build model. 10 models were created. One best model was selected. Procheck helped in checking reliability of model and the energy is checked through Prosa A.

**Results and Discussion**

BLAST worked for alignment *E.Coli* Yebc protein (PDB i.d. 1KON), crystal structure *E.Coli* Yebc protein was selected as template for homology modeling. E value, query coverage and sequence identity were the criteria for the selection of template. Similarity was 41% between target and template. 75% sequence identity and 100% query coverage was shown by both sequences. Modeller 9.10 was run to obtain 10 models. Out of 10 models, the best model was selected. The model was visualized by DS viewer. It consists of 8 alpha helices (red) and 7 beta sheets (blue).

The reliability of models was examined through Procheck. 10 files were set up by Procheck. A valuable accent among them was Ramachandran Plot. It has 4 regions which are distributed into maximum, fewer, inadequately and disallowed regions. 95.2% debris present in maximum allowed region, 2.9% in fewer allowed region, 1.4% in inadequate allowed region and in disallowed region, 5% particle were found. The energy was accessed through Prosa which was highly negative (3e-140). It showed the stability of model.
Blast result of *H. influenzae* strain 86 028NP hypothetical protein HI0433

Three Dimensional structure of *H. influenza* hypothetical protein HI0433. (8 helices 7 beta sheets)
**Plot statistics**

<table>
<thead>
<tr>
<th>Region</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum allowed regions</td>
<td>95.2% particles</td>
</tr>
<tr>
<td>Fewer allowed region</td>
<td>2.9% particles</td>
</tr>
<tr>
<td>Inadequate allowed region</td>
<td>1.4% particles</td>
</tr>
<tr>
<td>Disallowed region</td>
<td>5% particles</td>
</tr>
</tbody>
</table>

**Conclusion**

By homology modeling of *H. influenza*, transcriptional regulator protein (HI0433), structure was designed which has provided enough information for vaccine development to control its transcription for causing disease.

**References**


