ANALYSIS AND IDENTIFICATION OF ENSA AS A TARGET MOLECULE FOR DIABETES MELLITUS USING BIOINFORMATICS TOOLS

Sudha D.*1, Shoba K.2 and Asha S.3

1M. Sc Biochemistry, Dept. of Biochemistry, DKM College for Women, Vellore.
2Assistant Professor, Dept. of Biochemistry, DKM College for Women, Vellore.
3Assistant Professor, Dept. of Biotechnology, DKM College for Women, Vellore.

ABSTRACT
Diabetes mellitus, which is simply referred to as diabetes these days, is a metabolic disorder which in particular affects the metabolism of carbohydrates. The condition requires medical treatment and, more often than not, a number of lifestyle changes. The target gene for Diabetes mellitus is identified through NCBI data base. Using BLASTP the similarities of the target gene was identified. The primary analysis were done through coils, protparam and secondary analysis were done through YASPIN, SMART. The target gene had been modeled in Mdeller9v6 server and visualized in Rasmol. The results obtained from the investigation would have a scope for drug designing in the future studies.

The current work focus on functional characterization of target gene which is responsible for Diabetes mellitus.

KEY WORDS: Diabetes mellitus, Ncbi, BLASTP, YASPIN,SMART, Mdeller9v6, ENSA.

INTRODUCTION
Diabetes has been known for centuries, although it has not been fully understood, and the disease takes its name from the Greek for "passing through" because of one of its main symptoms - excessive urine production. During the fifteenth century the word Mellitus was from the Latin for "honey" when it was noted that many patients with diabetes had high levels of sugar in their blood and urine.
Diabetes mellitus, which is simply referred to as diabetes these days, is a metabolic disorder which in particular affects the metabolism of carbohydrates. The condition requires medical treatment and, more often than not, a number of lifestyle changes.

To function properly the human body requires a source of energy and derives this from the food that we eat. A normal diet comprises of a mixture of carbohydrates, proteins and fats with carbohydrates accounting for up to three-quarters of this mix. There are a wide variety of high carbohydrate (sometimes referred to as high starch) foods and these include bread, bran, cereal, beans, rice and pasta.

Food is broken down by the digestive process into a variety of organic compounds and one of these, which forms the body's prime source of energy, is glucose. Glucose is then carried to various parts of the body by the blood and is transferred to the cells of the body to fuel both cell growth and cell repair.

An essential element in the transfer process is the presence of insulin in the bloodstream. Insulin is produced by specialized cells (known as beta-cells) which are located in an area of the pancreas called the Islets of Langerhans.

Blem with the insulin producing beta-cells of the pancreas and are unable to produce sufficient insulin to transfer glucose from the bloodstream to the cells of the body. This means that it is necessary to closely monitor levels in the blood and to administer insulin so that glucose can be transferred and the glucose levels in the blood returned to normal.

**METHODOLOGY**

The gene coding protein sequence of the ENSA was retrieved in fasta format from NCBI. Sequence comparison studies was done using BLASTP program to find out the similarities between the target and template. Structural analysis of ENSA protein was done using primary, secondary analysis tools. The three dimensional structure prediction of ENSA protein was performed using an automated fold recognition modeling server called MODELLER. The evaluation process of modeled structure was done using Flipper server. The modeled protein structure was viewed in Molecular visualization tools, rasmol.

**RESULTS AND DISCUSSION**

NCBI

>gi|49168640|emb|CAG38815.1| ENSA [Homo sapiens]
MSQKREEENPAEETGEEKQDTQEKEGILPERAEAEAKLAKYPQLQKPGGSDFLMKR
LQKGQKYFDSDGY
NMAKAKMKKNQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGGQVE

>gi|49168639|emb|CR536578.1| Homo sapiens full open reading frame cDNA clone
RZPD0834E0422D for gene ENSA, endosulfine alpha; complete cds, incl. stop codon
ATGTCCAGAAACGAGAAGAGAAGAACCCTGCGAGGAGAGACCAGAAGACCGGCGAGGAGAA
GCAGGACACGCAGGAGA
AAGAAGGTTATTCTGCTGAGAGCTGAAGGCAAAGCTAAAGGCCAAATACC
CAAGCCTAGGACAAAA
GCCTGGAGGCTCCGACCTTCCTCATGAAGAGACTCCAGAAGGCAAAAGTGACTTT
TGACTCAGGAGACTAC
AACATGGCCAAAGGCAAGATGAAGAATAAGCAGCTGCCAAGTGCAGGACCAGA
CAAGAACCTGTTGACCTG
GTGATCACATCCCCACCCACAGATCTGCCCAGAGAAAGTCTCGGTCGTCGAC
CAGCAAGCTTTGCGGG
TGGCCAAAGTTGAATGA
The above results show the fasta format of ENSA sequence.

**Primary analysis**

**ProtParam**

**User-provided sequence**

```
MSQKREEENPAEETGEEKQDTQEKEGILPERAEAEAKLAKYPQLQKPGGSDFLMKR
LQKGQKYFDSDGY
10 20 30 40 50
NMAKAKMKKNQLPSAGPDKNLVTGDHIPTPQDLPQRKSSLVTSKLAGGGQVE

60
SDFLMKRLQK

70 80 90 100 110
GQKYFDSDGY

120
VTSKLAGGGQVE
```

**Number of amino acids:** 121

**Molecular weight:** 13417.0
Theoretical pI: 7.86

Aminoacidcomposition

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala (A)</td>
<td>8</td>
<td>6.6%</td>
</tr>
<tr>
<td>Arg I</td>
<td>4</td>
<td>3.3%</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>4</td>
<td>3.3%</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>7</td>
<td>5.8%</td>
</tr>
<tr>
<td>Cys I</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Gln (Q)</td>
<td>10</td>
<td>8.3%</td>
</tr>
<tr>
<td>Glu (E)</td>
<td>13</td>
<td>10.7%</td>
</tr>
<tr>
<td>Gly (G)</td>
<td>11</td>
<td>9.1%</td>
</tr>
<tr>
<td>His (H)</td>
<td>1</td>
<td>0.8%</td>
</tr>
<tr>
<td>Ile (I)</td>
<td>2</td>
<td>1.7%</td>
</tr>
<tr>
<td>Leu (L)</td>
<td>10</td>
<td>8.3%</td>
</tr>
<tr>
<td>Lys (K)</td>
<td>17</td>
<td>14.0%</td>
</tr>
<tr>
<td>Met (M)</td>
<td>4</td>
<td>3.3%</td>
</tr>
<tr>
<td>Phe (F)</td>
<td>2</td>
<td>1.7%</td>
</tr>
<tr>
<td>Pro (P)</td>
<td>9</td>
<td>7.4%</td>
</tr>
<tr>
<td>Ser (S)</td>
<td>8</td>
<td>6.6%</td>
</tr>
<tr>
<td>Thr (T)</td>
<td>5</td>
<td>4.1%</td>
</tr>
<tr>
<td>Trp (W)</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Tyr (Y)</td>
<td>3</td>
<td>2.5%</td>
</tr>
<tr>
<td>Val (V)</td>
<td>3</td>
<td>2.5%</td>
</tr>
<tr>
<td>Pyl (O)</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Sec (U)</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>(B)</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>(Z)</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>(X)</td>
<td>0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Total number of negatively charged residues (Asp + Glu): 20
Total number of positively charged residues (Arg + Lys): 21

Atomic composition

<table>
<thead>
<tr>
<th>Component</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon</td>
<td>578</td>
</tr>
</tbody>
</table>
Hydrogen  H  942  
Nitrogen  N  166  
Oxygen  O  192  
Sulfur  S  4  

**Formula:** $C_{578}H_{942}N_{166}O_{192}S_{4}$  
**Total number of atoms:** 1882  

**Extinction coefficients**  
This protein does not contain any Trp residues. Experience shows that this could result in more than 10% error in the computed extinction coefficient.  

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

- Ext. coefficient  4470  
- Abs 0.1% (=1 g/l)  0.333  

**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 65.24  
This classifies the protein as unstable.  

**Aliphatic index:** 52.48  

**Grand average of hydropathicity (GRAVY):** -1.255  
Protparam results shows the molecular weight, amino acid composition and biophysical properties of target protein.
Coils output for ENSA

SMART

Confidently predicted domains, repeats, motifs and features

Name Begin End E-value

low complexity 6 17 -
The region of low compositional complexity chosen starts at position 6 of the query sequence ends at position 17.

The sequence is: SDEXPAELTEGEE

From the results of SMART it shows the domain region in the ENSA protein, indicated by pink color.

HOMOLOGY MODELLING

Modeller9v6

Template – 2J13

Target – ENSA Protein

The above results shows the homology modeled structure of ENSA protein.

CONCLUSION

Molecular modeling and cheminformatics are becoming an essential component of drug discovery in every pharmaceutical industry. Rational drug design methods accelerate the
process by speeding up the discovery of new chemical entities that may become new drug. Determining protein function from genomic sequence is a central goal of bioinformatics. The fully sequenced genome of numerous organisms offers large amounts of information about cellular biology. It is a challenge of bioinformatics to use this information in discovering the functions of protein. According to recent trends homology modeling has become very valuable recently.

The main objective of the present study is to model receptors for the diabetes mellitus. This disease affected by various factors in humans. The functional Characterization of ENSA Protein which is responsible for diabetes mellitus were done through use of sophisticated bioinformatics software’s and tools. The future studies on this protein may lead to find out more insights.

REFERENCES


8. Shoba K., Manjuladevi M, Dr. Mazher sultana, Biochemical analysis and gene expression profiling on collagenase protein in fiddler crab, World journal of pharmacy and pharmaceutical sciences, issn 2278 – 4357, volume 6, issue 3, 747-756


13. Shoba.k and Dr. Mazher sultana, Three - dimensional structure and motif prediction studies on collagenase protein in fiddler crab, International journal


