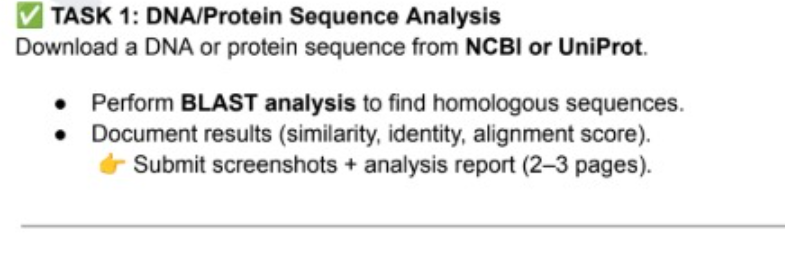
### ****DNA/Protein Sequence Analysis****



## ****Introduction****

## **Insulin is an anabolic protein hormone, composed of amino acids in two chains (A and B), that regulates glucose metabolism and energy storage. It is secreted by pancreatic beta cells and is vital for stimulating glucose uptake by cells and storing it as glycogen. The insulin molecule has a complex 3D structure with α-helices and β-sheets, linked by disulfide bonds, and is derived from a precursor molecule called proinsulin,**,, which undergoes post translational processing to form the active hormone.

## ****Sequence (FASTA Format)****

>NP\_001035835.1 insulin, isoform 2 precursor [Homo sapiens]

MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQASALSLSSSTSTWPEGLDATARAPPALVVTANIGQAGGSSSRQFRQRALGTSDSPVLFIHCPGAAGTAQGLEYRGRRVTTELVWEEVDSSPQPQGSESLPAQPPAQPAPQPEPQQAREPSPEVSCCGLWPRRPQRSQN

****Accession ID:**** NP\_001035835.1  
****Protein Name:**** Insulin, isoform 2 precursor  
****Organism:**** Homo sapiens  
****Protein length:**** 200 amino acids

**Organism:**

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

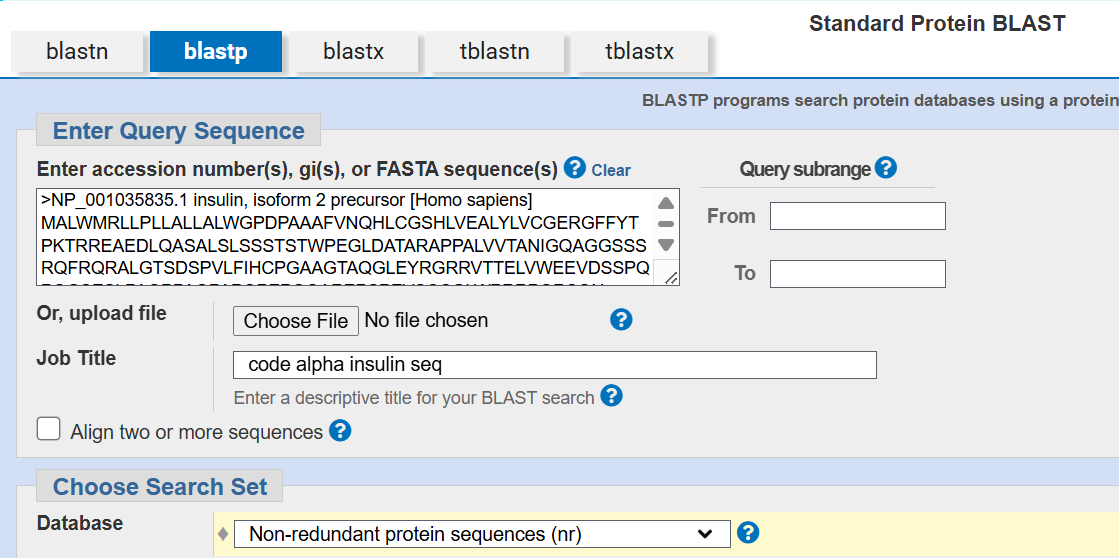
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;

Catarrhini; Hominidae; Homo.

## ****3. Methodology****

* The National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) was used to search and retrieve genome, nucleotide, and protein sequences of various organisms.
* For this study, the FASTA sequence of the insulin protein was obtained from NCBI.
* The retrieved protein sequence was then copied and analyzed using the BLASTp tool available on the NCBI BLAST homepage (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The BLASTp algorithm (protein–protein comparison) was selected to identify homologous protein sequences.
* The query sequence was pasted into the BLASTp input box, and the search was performed using default parameters, including an E-value threshold of 10 and the BLOSUM62 substitution matrix.
* The results were then analyzed to identify homologous sequences and to study their similarity and alignment characteristics

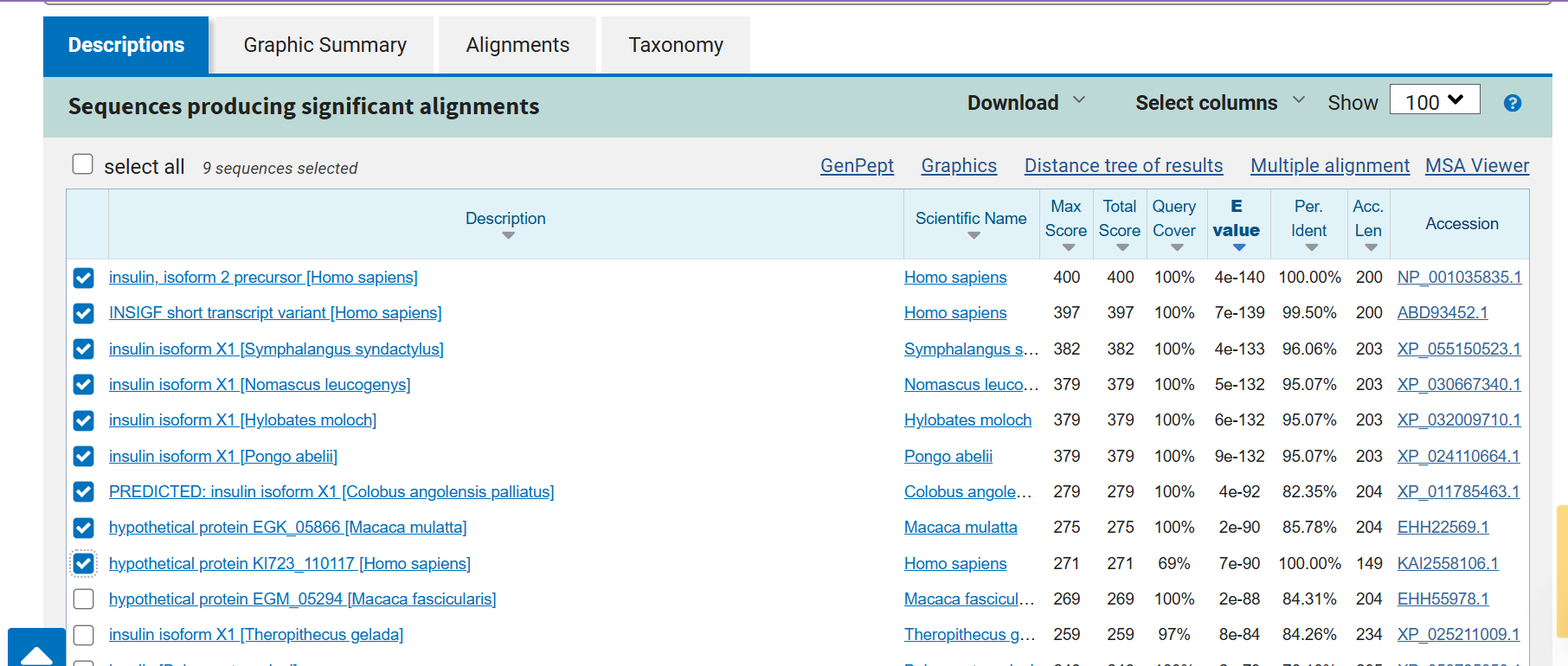
1. % Identity
2. % Similarity
3. E-value
4. Alignment score



### Figure 1: Query Sequence Pasted in BLASTp Interface

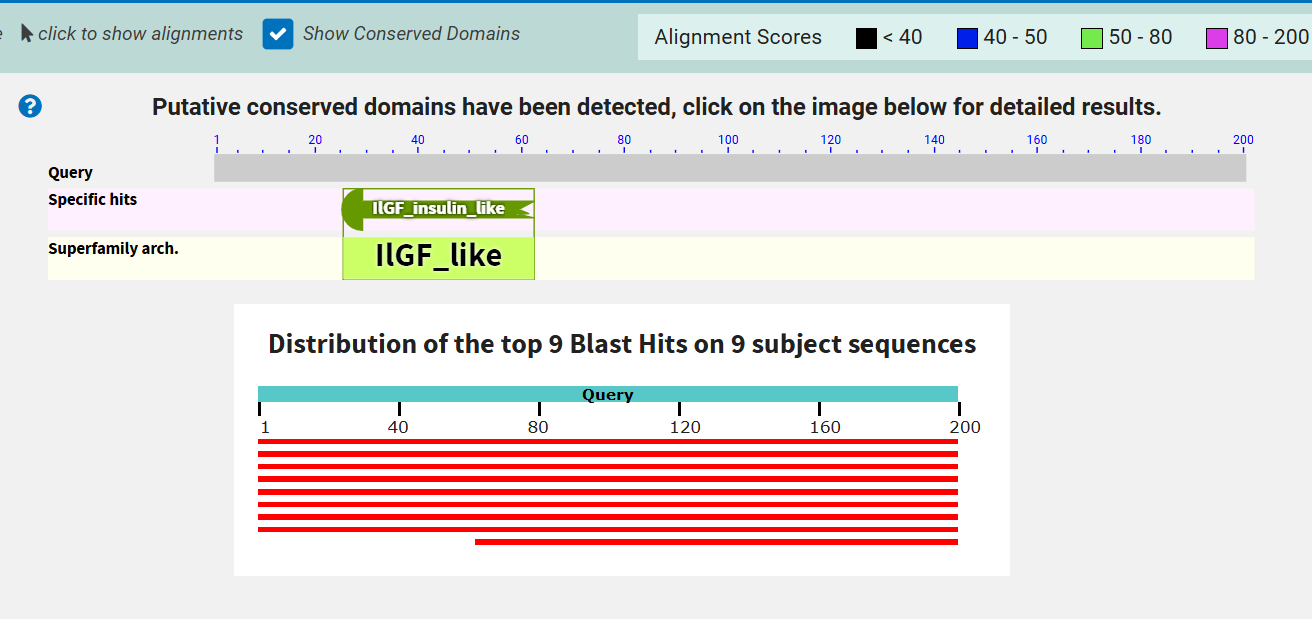
## ****4. Results****

After performing BLASTp analysis, homologous sequences were retrieved. The top hits included insulin proteins from closely related primates and mammals.

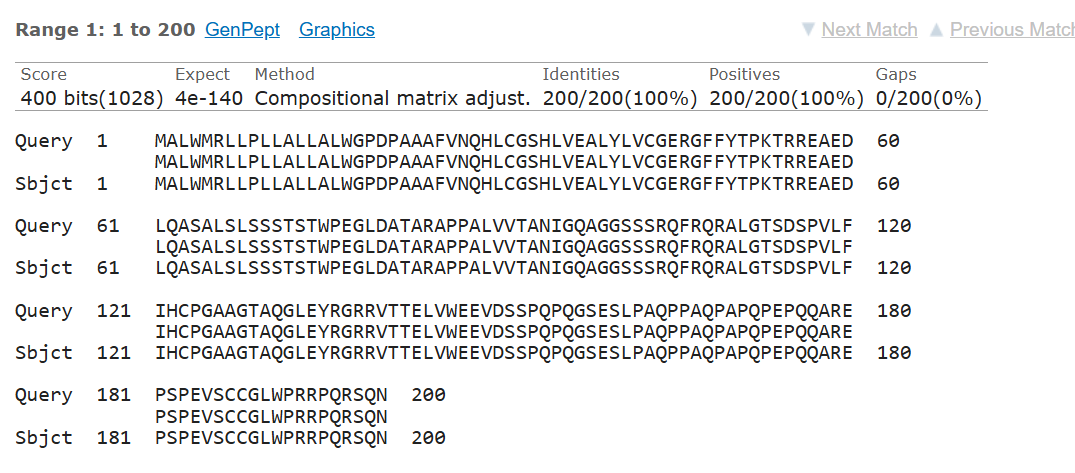


High sequence identity indicates strong evolutionary conservation of insulin’s structure and function.The **query coverage = 100 %** and **percent identity ≈ 100 %** for the top hits, which strengthens the conservation point.  
An E-value 4e-140 signifies that these matches are **highly significant**, meaning the similarity is not due to chance.

### Figure 2: BLAST Top Hits (from NCBI Results Page)



### 🔹 Figure 3: Alignment View of Query vs. Subject Sequence

100% query coverage and percentage identity and E-value 4e-140.

## ****Conclusion****

The BLASTp analysis of the human insulin protein sequence shows extremely high similarity and identity among mammalian insulin proteins, with 100% query coverage and nearly 100% sequence identity for top primate hits. The very low E-values (as low as 4e-140) confirm that these similarities are highly significant.