CodeAlpha Bioinformatics Internship –

Task 1 Report & Result

Title

DNA / Protein Sequence Analysis using BLAST

Student Name: **Gulam Sarwer**

Internship Platform: **CodeAlpha**

Domain: **Bioinformatics**

Task Number: **Task 1**

Introduction

Bioinformatics is an interdisciplinary field that combines biology and computer science to analyze biological data. Sequence analysis is one of the fundamental tasks in bioinformatics. helping in identify similarities between DNA or protein sequences and understand their biological significance. This task focuses on performing **DNA/Protein sequence analysis using the BLAST tool**.

Objective

     To perform protein sequence analysis using BLAST

     To identify homologous protein sequences

     To analyze sequence similarity, identity, query coverage, and E-value

Tools and Databases Used

     NCBI BLAST (BLASTp)

     UniProt Database

     Web based bioinformatics tools

Selected Sequence Details

Protein Name: Hemoglobin Subunit Beta

Organism: Homo sapiens

UniProt ID: P68871

Sequence Format: FASTA

Methodology

Protein sequence was downloaded from UniProt in FASTA format.

The FASTA sequence was submitted to NCBI BLASTp.

Default parameters were used for analysis.

BLAST search was performed against the non-redundant protein database.

Results were analyzed based on identity, query coverage, and E-value.

Results

High similarity with known hemoglobin proteins was observed.

Percent Identity: 100%

Query Coverage: 100%

E-value: 0.0

Alignment score indicated strong homology.

Screenshots of BLAST results and alignments were recorded for documentation.

```
>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH
```

⬇ Download ∨   GenPept Graphics   ▼ Next ▲ Previous ◄ Descriptions

beta globin [Homo sapiens]
Sequence ID: **AAZ39780.1** Length: **147** Number of Matches: **1**
See 4 more title(s) ∨   See all Identical Proteins(IPG)

Range 1: 1 to 147 GenPept Graphics   ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
299 bits(766)	4e-102	Compositional matrix adjust.	146/147(99%)	147/147(100%)	0/147(0%)

```
Query  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60
            MVHLTP+EKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
Sbjct  1    MVHLTPKEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60

Query  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120
            VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
Sbjct  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120

Query  121  KEFTPPVQAAYQKVVAGVANALAHKYH  147
            KEFTPPVQAAYQKVVAGVANALAHKYH
Sbjct  121  KEFTPPVQAAYQKVVAGVANALAHKYH  147
```

⬇ Download ∨   GenPept Graphics   ▼ Next ▲ Previous ◄ Descriptions

hemoglobin beta chain [Homo sapiens]
Sequence ID: **AAD19696.1** Length: **147** Number of Matches: **1**

Range 1: 1 to 147 GenPept Graphics   ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
299 bits(766)	6e-102	Compositional matrix adjust.	146/147(99%)	146/147(99%)	0/147(0%)

Related Information
Gene - associated gene details
Identical Proteins - Identical proteins to AAZ39780.1

Related Information
Gene - associated gene details

Feedback

Descriptions | Graphic Summary | **Alignments** | Taxonomy

Alignment view [Pairwise ▾]   ? **Restore defaults**   Download ▾

67 sequences selected ?

⬇ Download ▾   GenPept   Graphics   ▾ Next ▲ Previous ◄Descriptions

hemoglobin subunit beta [Homo sapiens]
Sequence ID: NP_000509.1 Length: 147 Number of Matches: 1

See 51 more title(s) ▾ See all Identical Proteins(IPG)

Range 1: 1 to 147 GenPept Graphics   ▾ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
301 bits(770)	1e-102	Compositional matrix adjust.	147/147(100%)	147/147(100%)	0/147(0%)

```
Query  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60
            MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
Sbjct  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60

Query  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120
            VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
Sbjct  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120

Query  121  KEFTPPVQAAYQKVVAGVANALAHKYH  147
            KEFTPPVQAAYQKVVAGVANALAHKYH
Sbjct  121  KEFTPPVQAAYQKVVAGVANALAHKYH  147
```

Related Information

Gene - associated gene details
AlphaFold Structure - 3D structure displays
Genome Data Viewer - aligned genomic context
Identical Proteins - Identical proteins to NP_000509.1

⬇ Download ▾   GenPept   Graphics   ▾ Next ▲ Previous ◄Descriptions

Screenshot 3:

⬇ Download ▾   GenPept   Graphics   ▾ Next ▲ Previous ◄Descriptions

beta globin chain variant [Homo sapiens]
Sequence ID: AAN84548.1 Length: 147 Number of Matches: 1

Range 1: 1 to 147 GenPept Graphics   ▾ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
299 bits(766)	4e-102	Compositional matrix adjust.	146/147(99%)	147/147(100%)	0/147(0%)

```
Query  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60
            MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
Sbjct  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60

Query  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120
            VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
Sbjct  61   VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  120

Query  121  KEFTPPVQAAYQKVVAGVANALAHKYH  147
            K+FTPPVQAAYQKVVAGVANALAHKYH
Sbjct  121  KKFTPPVQAAYQKVVAGVANALAHKYH  147
```

Related Information

Gene - associated gene details

⬇ Download ▾   GenPept   Graphics   ▾ Next ▲ Previous ◄Descriptions

beta-globin [Homo sapiens]
Sequence ID: ACU56984.1 Length: 147 Number of Matches: 1

Range 1: 1 to 147 GenPept Graphics   ▾ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
299 bits(766)	4e-102	Compositional matrix adjust.	146/147(99%)	147/147(100%)	0/147(0%)

Related Information

Gene - associated gene details
AlphaFold Structure - 3D structure displays

```
Query  1    MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK  60
```

Discussion

The BLAST analysis confirmed that the selected sequence belongs to Hemoglobin Subunit Beta protein. The high identity and low E-value indicate strong evolutionary conservation and biological relevance of the sequence.

Conclusion

This task successfully demonstrated the use of BLAST for protein sequence analysis. It enhanced understanding of sequence alignment and interpretation of BLAST outputs in bioinformatics research.

Learning Outcome

- Understanding FASTA format
- Hands-on experience with BLAST
- Ability to interpret alignment results

Acknowledgment

I would like to thank **CodeAlpha** for providing this opportunity to learn and practice bioinformatics concepts through hands-on tasks.