

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.

The screenshot shows the UniProt entry page for P68871 (HBB_HUMAN). The top navigation bar includes links for BLAST, Align, Peptide search, ID mapping, SPARQL, UniProtKB, Advanced, List, Search, and Help. The main content area is titled "P68871 · HBB_HUMAN". On the left, there's a sidebar with categories: Function, Names & Taxonomy, Subcellular Location, Disease & Variants, PTM/Processing, Expression, Interaction, Structure, Family & Domains, Sequence, and Similar Proteins. The "Names & Taxonomy" section shows "Protein" as "Hemoglobin subunit beta", "Gene" as "HBB", "Status" as "UniProtKB reviewed (Swiss-Prot)", and "Organism" as "Homo sapiens (Human)". The "Function" section describes the protein as involved in oxygen transport and mentions "Spinorphin". The "Amino acids" section indicates 147 amino acids. The "Protein existence" section shows evidence at the protein level. The "Annotation score" is 5/5. Below the main content, there are tabs for Entry, Variant viewer (810), Feature viewer, Genomic coordinates, Publications, External links, Tools, Download, Add, Community curated (3), Add a publication, and Entry feedback. A "Help" button is located in the bottom right corner.

Fri Jan 2 21:45

46% 74% 1.6 kB 100%

WhatsApp (3190) Crime Pa HBB - Hemoglo rest.uniprot.org/un GitHub NCBI Blast:sp|P 1AOU(MMDB) in +

https://rest.uniprot.org/uniprotkb/P68871.fasta

For quick access, place your bookmarks here on the bookmarks toolbar. [Manage bookmarks...](#)

```
>sp|P68871|HBB HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2
MVLTPPEEKSAVTALWGKVNVDEVGGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK
VKAHGKKVLAGFSGDLAHLNDNLKGTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPVQAYOKVAGVANALAHKYH
```

Fri Jan 2 21:46

42% 79% 110 kB 100%

49 (39) WhatsApp (3190) Crime Pa HBB - Hemoglo rest.uniprot.org/un GitHub NCBI Blast:sp|P 1AOU(MMDB) in +

https://blast.ncbi.nlm.nih.gov/Blast.cgi

For quick access, place your bookmarks here on the bookmarks toolbar. [Manage bookmarks...](#)

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](#) [▲ Previous](#) [◀ Descriptions](#)

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 MVHLTPPEEKSAVTALWGKVNVDEVGGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60
Sbjct 1 MVHLTPPEEKSAVTALWGKVNVDEVGGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60

Query 61 VKAHGKKVLAGFSGDLAHLNDNLKGTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120
Sbjct 61 VKAHGKKVLAGFSGDLAHLNDNLKGTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG 120

Query 121 KEFTPVQAYOKVAGVANALAHKYH 147
Sbjct 121 KEFTPVQAYOKVAGVANALAHKYH 147

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

Range 1: 1 to 147 [GenPept](#) [Graphics](#) [▼ Next](#) [▲ Previous](#) [◀ Descriptions](#)

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

Feedback

Fri Jan 2 21:46

55% 81% 14 kB 100%

WhatsApp Crime Po HBB - Hemoglo rest.uniprot.org/un GitHub NCBI Blast:sp| 1A0U(MMDB) |

https://blast.ncbi.nlm.nih.gov/Blast.cgi

For quick access, place your bookmarks here on the bookmarks toolbar. [Manage bookmarks...](#)

Descriptions Graphic Summary Alignments Taxonomy

Alignment view Pairwise ? Restore defaults Download

67 sequences selected

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 MVHLTPEEKSAVTALGKVNIDEVGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60
Sbjct 1 MVHLTPEEKSAVTALGKVNIDEVGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60

Query 61 VKAHGGKVKLGAFSGDLAHLNKGTFATLSELHCDKLHVDPENFRLGNVLVCVLAHFG 120
Sbjct 61 VKAHGGKVKLGAFSGDLAHLNKGTFATLSELHCDKLHVDPENFRLGNVLVCVLAHFG 120

Query 121 KEFTPPVQAYQKVAVANALAHKYH 147
Sbjct 121 KEFTPPVQAYQKVAVANALAHKYH 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

Feedback

Fri Jan 2 21:46

55% 81% 14 kB 100%

WhatsApp Crime Po HBB - Hemoglo rest.uniprot.org/un GitHub NCBI Blast:sp| 1A0U(MMDB) |

https://blast.ncbi.nlm.nih.gov/Blast.cgi

For quick access, place your bookmarks here on the bookmarks toolbar. [Manage bookmarks...](#)

Descriptions Graphic Summary Alignments Taxonomy

Alignment view Pairwise ? Restore defaults Download

67 sequences selected

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 MVHLTPEEKSAVTALGKVNIDEVGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60
Sbjct 1 MVHLTPEEKSAVTALGKVNIDEVGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60

Query 61 VKAHGGKVKLGAFSGDLAHLNKGTFATLSELHCDKLHVDPENFRLGNVLVCVLAHFG 120
Sbjct 61 VKAHGGKVKLGAFSGDLAHLNKGTFATLSELHCDKLHVDPENFRLGNVLVCVLAHFG 120

Query 121 KEFTPPVQAYQKVAVANALAHKYH 147
Sbjct 121 KEFTPPVQAYQKVAVANALAHKYH 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

Feedback

Fri Jan 2 21:46

55% 80% 0.6 kB 100%

WhatsApp Crime Po HBB - Hemoglo rest.uniprot.org/un GitHub NCBI Blast:sp| 1A0U(MMDB) |

https://blast.ncbi.nlm.nih.gov/Blast.cgi

For quick access, place your bookmarks here on the bookmarks toolbar. [Manage bookmarks...](#)

Descriptions Graphic Summary Alignments Taxonomy

Alignment view Pairwise ? Restore defaults Download

67 sequences selected

beta globin chain variant [Homo sapiens]
Sequence ID: [AN84548.1](#) 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 MVHLTPEEKSAVTALGKVNIDEVGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60
Sbjct 1 MVHLTPEEKSAVTALGKVNIDEVGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60

Query 61 VWAHCKKVKLGAFSGDLAHLNKGTFATLSELHCDKLHVDPENFRLGNVLVCVLAHFG 120
Sbjct 61 VWAHCKKVKLGAFSGDLAHLNKGTFATLSELHCDKLHVDPENFRLGNVLVCVLAHFG 120

Query 121 KEFTPPVQAYQKVAVANALAHKYH 147
Sbjct 121 KEFTPPVQAYQKVAVANALAHKYH 147

Related Information
Gene - associated gene details

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%)

Query 1 MVHLTPEEKSAVTALGKVNIDEVGEALGRLLVVPWTORFFESFGDLSTPDAVMGNPK 60

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

Feedback

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.