

# Comparative Evolutionary Analysis of Hemoglobin β Across Vertebrates

**Author:** Kashvi Rai  
**B.Tech Biotechnology**  
**Shoolini University**  
**Independent Project**  
**(2025-26)**

## 1. Abstract

Hemoglobin β is a protein responsible for oxygen transport in vertebrates. Due to its essential physiological role, it is expected to be evolutionarily conserved. In this study, hemoglobin β protein sequences from selected vertebrates *Homo sapiens*, *Pan troglodytes*, *Canis lupus familiaris*, *Rattus norvegicus*, *Xenopus borealis*, and *Ictalurus punctatus* were analyzed to investigate sequence conservation and evolutionary relationships. Pairwise sequence similarity analysis, multiple sequence alignment, functional residue examination, and phylogenetic tree construction with bootstrap support were performed. The results reveal complete conservation between human and chimpanzee hemoglobin β, high conservation within mammals, and progressively increasing divergence in amphibians and fish. Despite this divergence, key functional motifs involved in oxygen binding and structural stability remain conserved across all species, indicating strong purifying selection. Phylogenetic analysis mirrors known vertebrate evolutionary history, confirming that hemoglobin β evolution closely follows phylogenetic distance while maintaining essential biological function.

## 2. Materials and Methods

### 2.1 Sequence Retrieval

Hemoglobin β protein sequences were retrieved from public databases (NCBI and UniProt) for the following species:

- *Homo sapiens*
- *Pan troglodytes*
- *Canis lupus familiaris*
- *Rattus norvegicus*
- *Xenopus borealis*
- *Ictalurus punctatus*

### 2.2 Sequence Similarity Analysis

Pairwise BLAST comparisons were performed to calculate percentage identity and E-values relative to human hemoglobin β.

### 2.3 Multiple Sequence Alignment

Multiple sequence alignment was carried out using Clustal Omega (v1.2.4) to identify conserved regions and functional motifs.

## 2.4 Phylogenetic Analysis

A phylogenetic tree was constructed using MEGA with the Maximum Likelihood method under the JTT amino acid substitution model. Statistical reliability was assessed using standard bootstrap analysis with 1000 replicates.

# 3. Results

## 3.1 Comparative Sequence Similarity Analysis

The screenshot shows the NCBI Blast: Protein Sequence results page. The query descriptor is "unnamed protein product" with a length of 147. The subject ID is lcl|Query\_3427435 (amino acid). There are no subject descriptors or lengths provided. Under "Other reports", there are links for "Multiple alignment", "MSA viewer", and a help icon. The main section is titled "Clusters producing significant alignments". A table lists one cluster: "unnamed protein product" with a Max Score of 142, Total Score of 142, Query Cover of 100%, E value of 3e-49, Per. Ident of 46.26%, Acc. Len of 147, and Accession of Query\_3427435. The "Clusters" tab is selected, while "Graphic Summary", "Alignments", and "Dot Plot" tabs are also present. A "Download" button, "Select columns" dropdown, and a "Show 100" button are at the top right of the table area.

The screenshot shows the Protein BLAST search interface. The search type is set to "blastp". The "Enter Query Sequence" section contains the query sequence: MVHLTPPEEKSAVTALWGKVNVDEVGGGEALGRLLVVYPWTQRYFGSGFNLSTPDAVMGNPKVKAHGKKVVLGAFSDGLAHLDNLKGTFATLSLHSEKLHVDPDNENFRLLGNVLVCLVLAHHFGKEFTPVQQAQVKVAGVANALAHKYH. Below it, there are fields for "From" and "To" subranges, "Choose file" (No file chosen), "Job Title" (empty), and a checked checkbox for "Align two or more sequences". The "Enter Subject Sequence" section contains the subject sequence: MVVWTDFERATIQQDVFSQLIDYESVGHQALSRLCLVVYPWTQRYFGSGFNLNTAAAIIGNPKVAAGHLVVRGLEKAAKNMNDNIKAYADLSLVLHSEKLHVDPDNFLKLADCTIVIVASVLGASFTAEVQAAQLKFLAVVSLGKQYQ. Below it, there are fields for "From" and "To" subranges, "Choose file" (No file chosen), and a checked checkbox for "Align two or more sequences". The "Program Selection" section shows "blastp (protein-protein BLAST)" selected. A "Feedback" button is located in the bottom right corner.

Species	% Identity	E-value	Interpretation
<i>Pan troglodytes</i>	100%	5e-112	Very highly conserved
<i>Canis lupus familiaris</i>	89.73%	1e-100	Highly conserved
<i>Rattus norvegicus</i>	81.63%	2e-92	Moderately conserved
<i>Xenopus borealis</i>	58.09%	2e-58	Marked evolutionary divergence
<i>Ictalurus punctatus</i>	46.26%	3e-49	Strong evolutionary divergence

The progressive decline in hemoglobin β sequence identity from primates to fish mirrors vertebrate evolutionary history, underscoring strong functional conservation coupled with lineage specific divergence.

**Note: Hemoglobin subunit beta-1 sequences were analyzed for *Rattus norvegicus* and *Ictalurus punctatus* due to the absence of curated canonical β-globin entries in the database.**

### 3.2 Multiple Sequence Alignment and Motif Conservation

Multiple sequence alignment revealed strong conservation of functionally critical regions across vertebrates:

- N-terminal region (1–60): VVYPWTQRY .** This region is involved in heme interaction and oxygen binding. Strong conservation shows critical functional region. Fish sequences show substitutions consistent with deep evolutionary divergence. The core region remains conserved with increasing divergence from mammals to amphibians to fish.
- Core functional region (61–120): HVDP(D/E)N(F)**  
Conserved with minor conservative substitutions present in all. This region stabilizes the globin fold. This block is under strong purifying selection across all vertebrates.
- C-terminal region (121–147): AHKYH (or equivalent)**  
Conserved with small length variations conserved at the end. Terminal residues are important for subunit interactions. Even distant species retain structural constraints. C-terminal region tolerates more variation but preserves key residues.

The screenshot shows a multiple sequence alignment interface. At the top, there's a header bar with links for 'Root output', 'Sequence', 'Aligner', 'Output types', 'Residues', and 'Aligner parameters'. Below this is a 'Download' button. The main area displays a sequence alignment with several test entries (test5, test6, test4, test1, test2, test3) and their corresponding aligned sequences. The sequences are color-coded by residue type, and asterisks (\*) indicate positions of conservation across the different species. The alignment shows a highly conserved N-terminal region (residues 1-60), a core functional region (residues 61-120), and a C-terminal region (residues 121-147).

Multiple sequence alignment of hemoglobin subunit beta across vertebrates revealed extensive conservation within mammals, complete identity between human and chimpanzee, and progressively increasing divergence in amphibians and fish. Despite sequence variation, key functional motifs involved in oxygen binding and structural stability were conserved across all species, highlighting strong purifying selection acting on hemoglobin β throughout vertebrate evolution. These results indicate strong purifying selection acting on essential functional regions.

### 3.3 Functional Residue Analysis

Key functional residues in hemoglobin β include:

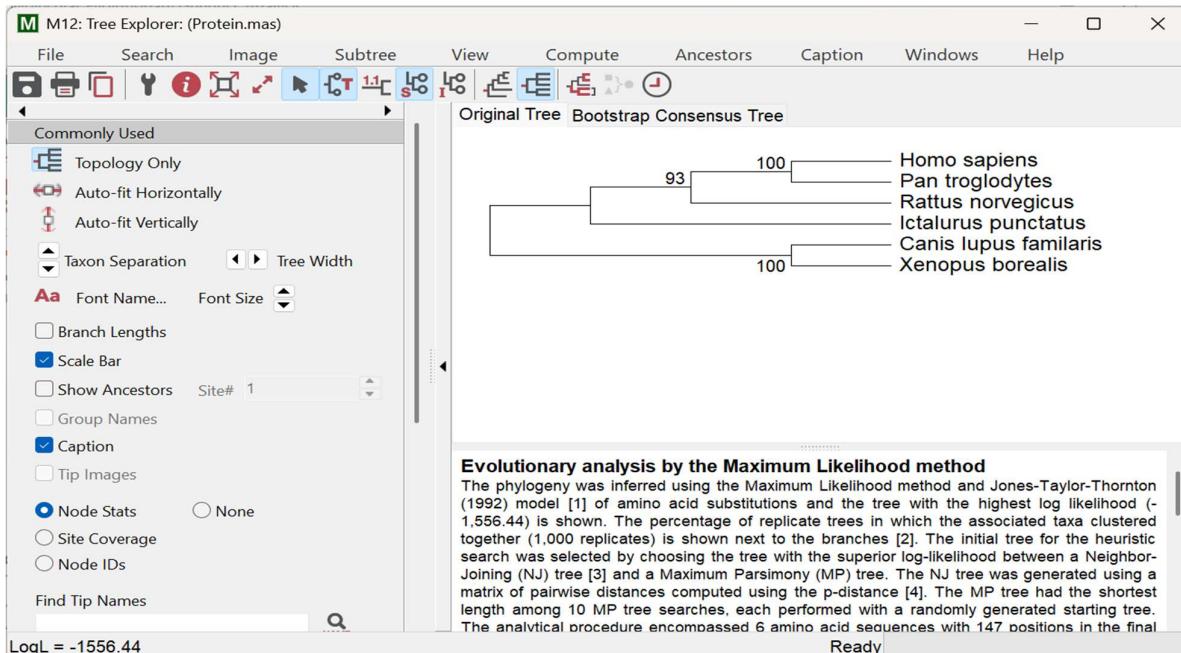
- **Heme/Oxygen-binding residues**
- Proximal histidine (His93β) binds the heme iron ( $\text{Fe}^{2+}$ ), while distal histidine (His63β) stabilizes bound oxygen, enabling reversible oxygen binding.
- **Structural core residue**
- Conserved hydrophobic residues maintain the α-helical globin fold and correct heme positioning.
- **Subunit interface residues**
- Charged and polar residues form salt bridges and hydrogen bonds between α and β subunits, stabilizing the tetramer and enabling cooperative oxygen binding.
- **Motif Conservation Summary**

Motif	Mammals	Frog	Fish	Substitution Type	Selection Pressure
VVYPWTQRY	Yes	Yes	Yes	Mostly conservative	Strong
HVDP(D/E)N(F)	Yes	Yes	Yes	Rare, conservative	Very strong
AHKYH	Yes	Yes	Yes	Conservative ± few non-conservative	Moderate

### 3.4 Phylogenetic Analysis and Bootstrap Support

The Maximum Likelihood phylogenetic tree revealed:

- A strongly supported mammalian clade (bootstrap = 93%)
- *Homo sapiens* and *Pan troglodytes* forming a sister group with 100% support
- *Xenopus borealis* branching separately with 100% support
- *Ictalurus punctatus* occupying a basal position



The tree topology mirrors known vertebrate evolutionary relationships, confirming strong conservation of hemoglobin  $\beta$ .

## 4. Discussion

Hemoglobin  $\beta$  exhibits high evolutionary conservation due to its important role in oxygen transport. While overall sequence divergence increases with phylogenetic distance, essential motifs and functional residues remain conserved across vertebrates. This pattern reflects strong purifying selection, allowing limited variation in non-critical regions while preserving functional integrity. The phylogenetic analysis further supports this conclusion by reproducing established vertebrate evolutionary relationships with high statistical confidence.

## 5. Limitations of the Study

The analysis was limited by the availability of curated  $\beta$ -globin coding sequences in non-model organisms therefore, protein level evolutionary analysis was emphasized over codon based selection metrics.

## 6. Conclusion

This study demonstrates that hemoglobin  $\beta$  is highly conserved across vertebrates despite increasing sequence divergence with evolutionary distance. Preservation of key functional motifs involved in heme binding, structural stability, and subunit interaction highlights the action of strong purifying selection. Phylogenetic analysis with robust bootstrap support confirms that hemoglobin  $\beta$  evolution closely follows vertebrate evolutionary history.

## **7. References**

1. Clustal Omega: Multiple Sequence Alignment Tool (2011)
2. MEGA X: Molecular Evolutionary Genetics Analysis (2018)
3. NCBI Protein Database, National Center for Biotechnology Information
4. UniProt Knowledge base (UniProtKB)