

IBO-style Bioinformatics Problem: Multiple Sequence Alignment

BoGuide

2026

IBO-style Bioinformatics Problem: Multiple Sequence Alignment

You are studying a family of small metal-binding proteins (putative zinc fingers) from five different species. Their amino acid sequences are given below in FASTA format.

Sequences:

Sp1 (Species A)
MCPECGKSFSQKSDLVRHQRTHTGEK

Sp2 (Species B)
MCPECGKSFSQKSDLVRHQRTHTGEK

Sp3 (Species C)
MCPECGKSFSQKSDLVRHQRTQTGEK

Sp4 (Species D)
MCPECGKSFSQKSDLVRHQRTQTGEK

Sp5 (Species E)
MCPECGKSFSQKSDLVRHQRTHTGEK

The protein is **27 amino acids** long. Assume that the conserved Cys and His residues are part of the zinc-finger motif. The sequences have already been aligned for you in the first line of the alignment (but you should verify manually if needed).

Task 1. Visual inspection (no software required)

- a) Which positions are **100% identical** across all five sequences? List the amino acid(s) and their positions (count from the N-terminus, starting at 1).
- b) Which positions show **conservative substitutions**? (i.e., amino acids with similar physico-chemical properties: acidic, basic, hydrophobic, polar uncharged). Name the positions and the changes.
- c) The motif “**CXXC**” (Cys-X-X-Cys) is characteristic of zinc fingers. Where does this motif occur in these sequences? Write the exact four-residue block.

Task 2. Phylogenetic inference

Based on the sequence differences, draw a simple **cladogram** (branching diagram) that shows the most parsimonious evolutionary relationships among the five species. Indicate which sequences are identical and which differ. Assume no horizontal gene transfer and that each difference is a single substitution event.

Task 3. Alignment quality assessment

If you performed a multiple alignment using **Clustal Omega** with default parameters (gap opening penalty = 10, gap extension = 0.2, substitution matrix = BLOSUM62), would the alignment change? Justify your answer.

Scoring scheme (optional – for manual alignment practice)

- Match = **+1**
- Mismatch = **-1**
- Gap (insertion/deletion) = **-2** per gap residue

Try to align the sequences by hand and calculate the alignment score. Compare with the provided alignment in the solution.

Note: This problem is designed for International Biology Olympiad (IBO) style training on BoGuide.