## Background

Many species produce venoms, which are primarily used for either predation or defense purposes. Different venom proteins can have different mechanisms of action, and in some cases these proteins have evolved specifically to mimic a hormone or other ligand that activates the pain receptors in its target species. Today, you will be looking at one such venom protein (Mg1a) that was isolated from the Australian red bull ant (*Myrmica gulosa*).

## Questions

### Part 1: Find Similar Sequences with BLAST [4 pts total]

To get the sequence for the ant venom protein in question, go to the [UniProt](https://www.uniprot.org/) database and find the protein associated with accession #P0DSL4 (<https://www.uniprot.org/uniprotkb/P0DSL4/entry>). From within UniProt, BLAST this protein against the rest of the database using default parameters.

#### 1A. What is the best-scoring BLAST result (not including the protein matching against itself)? What is the accession/Uniprot ID #, and what species does this best-scoring protein come from? Is the species with the best hit one that you would expect to be closely-related to an ant species? *(1 pt*)

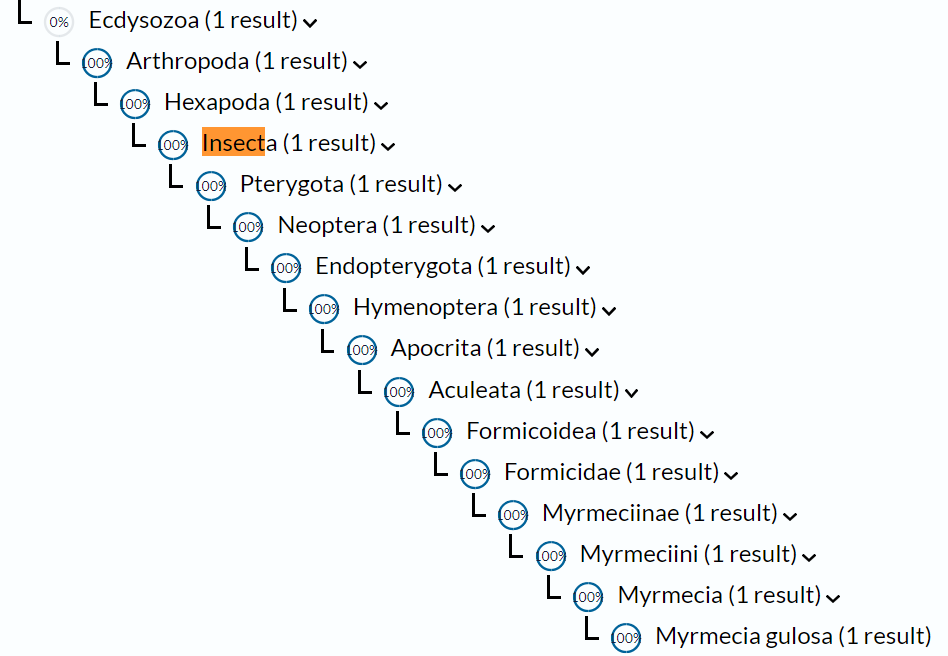
[A0A8D2Q4D8](https://www.uniprot.org/uniprotkb/A0A8D2Q4D8/entry)

Heparin binding EGF like growth factor

Varanus komodoensis (Komodo dragon)

This is not expected to be closely-related to an ant species.

#### 1B. How many proteins from other insect species are in the top 250 BLAST results? *(1 pt*)



0 proteins are from other insect species.

Next, either download the sequence or copy and paste it into a file, and go to the NCBI website to run PSI-BLAST against the standard databases with default parameters.

#### 1C. Does using the PSI-BLAST method, which should be better for finding more distantly related proteins, change the results you got earlier in any way? *(2 pts*)

Using PSI-BLAST should be better for finding more distantly related proteins, changed the results, and gave more insect/ant species.

### Part 2: Compare Multiple Sequence Alignment Results from Different Methods [4 pts total]

Download the [Fasta](https://uncc.instructure.com/courses/192992/files/20292864?wrap=1) file containing the sequences of 4 ant and bee venom proteins from the Spitz family (this is the family that is typical of most ant venoms). Go to the [EBI page](https://www.ebi.ac.uk/Tools/msa/) for Multiple Sequence Alignment tools, and perform an alignment of these proteins using 2 different methods: MUSCLE and T-COFFEE. Save each result in the Pearson/FASTA format, and then use the [VerAlign](https://www.ibi.vu.nl/programs/veralignwww/)

[Links to an external site.](https://www.ibi.vu.nl/programs/veralignwww/)

webtool to compare each result to a [reference alignment file.](https://uncc.instructure.com/courses/192992/files/20292875?wrap=1) Use the Results Summary tab from the alignment results page to find details about scores and percent similarity among pairs.

#### 2A. How do the MUSCLE alignment and T-COFFEE alignment results each compare to the reference alignment? *(2 pts)*

They are both really close to the reference alignment. MUSCLE: 1.00 SPDist, T-COFFEE: 0.99 SPDist.

#### 2B. Using the best alignment from 2A, what is the average % identity among the Spitz proteins? *(2 pts*)

87.77% is the average % identity among the Spitz proteins.

### Part 3: Multiple Sequence Alignments with "Mystery" Venom Protein [3 pts total]

Using the same tool that produced the best alignment in Part 2, perform another multiple sequence alignment of the Spitz proteins, but this time add the P0DSL4 protein sequence from Part 1 into the set, to see how well it aligns to the other family members.

#### 3A. On average, what is the percent similarity of the mystery protein from part 1 with the other insect venom proteins? *(1 pt*)

30.07% is the average percent similarity of P0DSL4 to the other family members.

Now compare the protein from Part 1 with a set of [human EGF-domain containing proteins](https://uncc.instructure.com/courses/192992/files/20293163?wrap=1). Use the same alignment method as before.

#### 3B. What is the average percent similarity of the mystery protein with the human EGF proteins? Is this higher or lower than the average percent identity among the different human proteins? *(2 pts*)

34.2325% is the average percent similarity of the mystery protein with the human EDF proteins. This is higher than the average percent similarity of humans amongst themselves.

### Part 4: Find Closest Match to Mystery Protein Among Local Marsupial Species [4 pts total]

For the final part of the exercise, compare the mystery ant protein to [a set o EGF-domain contain sequences from Australian mammal species](https://uncc.instructure.com/courses/192992/files/20293411?wrap=1)that would be found in this ant's natural habitat.

#### 4A. How well does the ant protein align to these marsupial EGF sequences compared to the alignments with the other insects and the other mammals? Which species does it share the closest match with? *(3 pts*)

The average percent similarity is 40.88% and is higher than with the other insects and other mammals. It is most closely related to the wombat.

Use the [Emboss Cons](https://www.ebi.ac.uk/Tools/msa/emboss_cons/) tool to get a consensus sequence from the multiple sequence alignment from part 4A.

#### 4B. What is the consensus sequence for this alignment? *(1 pt)*?

>EMBOSS0001

MKLSxFILKLLLAAVFSALVSGESLERAQSRLSENRGTENPDSSTGxLNQMLHSxVSRTE

VLDLQDTEHDLVRVAFSSKPQALVTPSKExNGQKKRKxKGMGRKRDPCLRKYKDYCIHGS

CKYLKELRMPSCICQTGYHGERCHGLSLPVENPLYGYDHTTILAVVSVVLSSVCLLIIAG

LLMFRYxxxxxxxxESxEKxKxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

xxxxxxxxxxxxx