SCOPE:

1)

* NCBI is the National Center for Biotechnology Information. It provides access to databases of biomedical and genomic information.
* Entrez is a cross-database search system that allows users to search the databases of the NCBI
* Gene Expression Omnibus (GEO) is a microarray database from NCBI on gene expression. It is a peer reviewed public repository.

2)

* FASTA is a text-based format for representing nucleotide or peptide sequences using a single letter.
* BLAST finds regions of similarity between biological sequences. It compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

3)

* FASTA uses a Smith-Waterman algorithm
* BLAST is its own algorithm
* The search for local regions for similarity
* Blast is much faster and more accurate than FASTA.

4)

* PubMed is a search engine for a compilation of biomedical literature from MEDLINE, life science journals, and online books provided by NCBI.
* Alternatives include eTBLAST 3.0, GoPubMed, Medline, PubGet, and Quertle

**PART C:**

# Times ELVIS appears = 174

# Times LIVES appears = 126

# Times ELVIS and TIMES appear consecutively = 0

# Times ELVIS and DEAD appear together = 0

# Times PERL appears = 3963

# Times PERLISGREAT appears = 0

If any of the larger statements appeared in the protein database, this would mean that this segment was favored throughout evolution or there was some miracle that caused these characters to occur consecutively. Usually, strings of 10 consecutive characters do not occur regularly. There might also be inaccurate data.

**PART D:**

- A B as the quality score means that the Phred score is 2. If the data ends in a long line of B’s, this could indicate that the read ended with a segment of mostly low quality and therefore were replaced with all B’s. These B’s do specify a particular error rate, but rather the uselessness of the data.

- Assuming that the B is actually characterizing the quality and not the uselessness of data, A Phred score of 2 (B) would result in a probability of .63096