For this assignment, BLASTs were performed for queries with several sensitivity modifications. The default parameters excluded all computer generated sequences and /or environmentally produced samples (contamination possibility). The first sensitivity parameter modification was the query’s word size. The word size is the query sequence seed argument length. To initiate pairwise alignment comparison, the database entry’s sequence must contain an exact match of the query sequence seed to initiate pairwise alignment. A very long seed word produces far fewer hits than a smaller seed word. The probability of a match/hit due to chance alone is where n is the word size (our query 🡪 ). Increasing the word size increases the sensitivity of the search but it may also exclude many homologs due to excessive stringency.

The next sensitivity modification was the data source and type under analysis. The database was changed from RefSeq RNA to RefSeq Protein. This reduced the sensitivity and yielded more hits. The degeneracy of the genetic code allows several different, triplet codon sequences to code for the same amino acid. As a result of this degeneracy, two proteins with identical domains could have sequences that differ at every third nucleotide position (67% identity similarity). By searching the protein database, many more true positives, which may have been excluded from a RNA database search, can be identified as hits.

The last sensitivity modification was the substitution matrix. For the last query, the substitution matrix was changed from Blosum45 to Blosum80. These programs build substitution matrices based on a minimum percent identity. The number appended to the end of the Blosum program corresponds to the minimum percent identity. The program builds a matrix of e values based on the probability that sequences of similar percent identities (45% or 80%) are the result of chance alone. The Blosum80 matrix is more sensitive and useful for identifying more recently diverged species. The Blosum45 program is more useful for determining ancestral homology.

To determine the accuracy of the sensitivity modifications, the x globin protein from the Xenopus tropicalis (frog) that was identified as a match from Blast #4 (Blosum45) was reviewed. The x globin protein hit corresponded to RefSeq accession # NP\_001011196.1. The e value for this amino acid sequence was 1.3. This value suggests that hits with amino acid similarity of this magnitude can be expected slightly more often than one time for each query of similar parameters. The magnitude of the e-value suggests ambiguity regarding certainty of homology. After further review of the sequences, only 18% of the domains are conserved. The sequences share similar classes for only 45% of non-matching amino acids. These matching or class sharing amino acids occur intermittently throughout the sequence. The maximum number of contiguously matching amino acids between the sequences is three. The 27 amino acid matches are sprinkled throughout the 150 amino acid length sequence alignment. After initial review, convergent evolution rather than homology was assumed due to the low percent identity and lack of contiguous amino acid matches between the sequences. However, a PubMed article connected to X globin accession number NP\_001011196.1 describes the protein as homologous to the neuro-globin proteins of amniotes. The article describes an intron sliding event as the possible cause of the large gaps in amino acid domain matches (“A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals,” [Roesner A](http://www.ncbi.nlm.nih.gov/pubmed/?term=Roesner%20A%5BAuthor%5D&cauthor=true&cauthor_uid=15356282), [Fuchs C](http://www.ncbi.nlm.nih.gov/pubmed/?term=Fuchs%20C%5BAuthor%5D&cauthor=true&cauthor_uid=15356282), Hankeln T, Burmester T, Institute of Zoology, Johannes Gutenberg University, Mainz, Germany). Therefore, the data supports a distant homologous relationship between the proteins separated by several speciation and duplication events. The speciation and duplication events accounts for the observed differentiated forms of globin protein present in higher level deuterostomes and the single form of globin protein found in “more ancient” protostomes.