A basic Introduction to Computational Biology

Here are some basic key points:

Biology has many levels (molecule small, DNA, RNA, Proteins; protein networks, cellular processes, cells, organisms, ecosystems and clinic/patient)

Biology does fast and close-to-optimal decisions, but is not mathematically accurate. Also computational biology algorithms have these properties. This is called a heuristic.
An example is BLAST (basic local alignment search tool) which uses a double index search of the database (like looking for "Thomas Müller" in a phone book): This is fast but not accurate. Only if two such index hits are found close by in a database entry, the accurate alignment is calculated.

BLAST (on NCBI portal) is the basic local alignment search tool. It does only a local alignment of the sequence, so check the other parts of the protein (if any) by a second BLAST search, only then the analysis is complete.

For validation use Motif- and profile searches by PROSITE (on ExPaSy Portal), these check not only the local domain match, but whether there are no detrimental mutations.

We also used a neuronal network when we used the software PREDICTPROTEIN. This uses the proteinsequence and three neuronal networks to predict helices, strands and coil regions. Prof. Rost used a fourth network as arbiter-network in case the three networks predicted ambiguous results (e.g. all three secondary structures at the same place). The neuronal network makes 70% accurate predictions, with the use of multiple alignment it is now 80% accurate, better is not possible without 3D structure information.

Heuristic, easy tests: (i) If you use Wikipedia, please use several languages (English and your native language for instance) and several related key words to study in depth a subject (example shown: Turing Machine).

(ii) for bioinformatics or, in fact, any computer calculation, simply check the result by a related software or source. So for instance the prediction by BLAST by FASTA or Smith and Waterman search. The domain hit by another domain databank search. The motif hit should match the protein and the motif hit can be from PROSITE, but also from conserved domain server (at NIH on BLAST metaserver of NIH).

(iii) Always anchor statements by a documented, trusted reference (principle of Wikipedia, but also of scientific research). Similarly, when a BLAST hit hits a protein, the final similarity should be to a protein with a documented, experimental verified function.

Please check-out the Portals (offering lots of software and databases) at NCBI (PubMed, OMIM, Genbank) and at the Swiss Bioinformatics Institute (SBI, the ExPaSy tool box to analyze proteins; Swissmodel allows to establish homology models for protein structure if a related protein "template" can be found in the database)

Please use several browsers for your search, Google can be replaced by other browsers such as startpage or ecosia

Remember, a browser uses an Index (all websites available) and searches this by an intelligent method. The Index for all three browsers above is provided by Google.

Computers including artificial intelligence can only do calculations (Turing calculable numbers), but by construction they cannot do any ethical decisions, no feelings, no moral, no fundamental decisions:
So please be aware all this you as human beings have to provide, do not delegate this to the web or a machine. Only artificial Life could do this, too, but all Life (man, mice, ants, plants, bacteria) is equal.

We have clearly not yet the maturity as society that we should try in the foreseeable future (next 30 years) to create artificial life. An indicator for the maturity would be that we treat fair every human being (including foreigners, women, ill or handicapped people), or every species in every ecosystem (rainforest is burned and not protected etc.), still a long way to go.

Task: Please find all the Links for the points mentioned in this text in the Web ☺.