**Project description Anna van Tiel**

Proteins are essential highly specialised components of living cells and organisms. They are usually part of signal transduction pathways and work together with other proteins to regulate cellular processes. There are many different proteins, with many different functions, ranging from effector functions to regulatory functions such as activation or inhibition of specific processes. In signal transduction pathways, each protein is like a key which affects the next key in the pathway, for instance through amplification and inhibition. All proteins together can integrate many different signals, and this way regulate cellular processes very precisely, resulting in a physiological balanced system.

One such pathway is the PI3K-Akt signalling pathway, which regulates cell cycle progression, cell division, and apoptosis (controlled suicide of a cell). At the beginning of this pathway is the protein Akt1, also known as protein kinase B, which is a serine/threonine specific protein kinase which activates several proteins through phosphorylation (attachment of a phosphate group) at their serine and threonine residues (amino acids, the building blocks of proteins). Akt1 itself is also controlled through phosphorylation at its serine and threonine residues, which induces a conformational change, allowing it to bind with its Pleckstrin-Homology (PH) domain to the cell membrane, localising Akt1 to the proteins it needs to activate. This concept is inherent to proteins: form fits function.

The Akt1 protein exist in many species, from mammals, to birds, to insects. Over the course of hundreds of thousands of years of evolution, the proteins across different species may have accumulated mutations and diverted across these very different species. Yet, since form fits function, the proteins cannot endlessly accumulate mutations without affecting its functionality. This is what the current project is about. In this project, **I am investigating whether there is evidence for evolutionary pressure on the Akt1 protein**. Given that Akt1 is essential for regulation of the PI3K-Akt pathway and controlling cell cycle progression, my hypothesis is that there is evolutionary pressure on specific parts of the protein, in particular the sequences that fall within a specific domain with its own function, such as the PH or catalytic domains discussed earlier.

The **data** which I used to answer my research question are:

* DNA sequences from the Akt1 genes of six diverse organisms: human, rhesus monkey, rat, chicken, zebrafish, and fruit fly
* Consensus amino acid sequences of 5 domains matching the amino acid sequences of Akt1 after matching the Akt1 amino acid sequences from these six organisms to the Pfam domain database (see part 4)

**Part 1: aligning Akt1 genes and proteins**

In accordance with the central dogma which describes how genes, which are DNA sequences, encode proteins, which are amino acid sequences that are folded to form a protein, my project started with DNA. I acquired the coding sequences of Akt1 genes from the National Centre for Biotechnology Information (NCBI). I used the Biopython library to translate these sequences into amino acid sequences. I used the command-line ClustalW tool to align both the gene and protein sequences of the different organisms to obtain alignment scores and draw a phylogenetic tree to illustrate the distance between these genes and proteins. Rhesus monkey and human Akt1 were the closest. Fruit fly Akt1was the most divergent from the other species.

**Part 2: investigating single amino acid conservation**

I looked at the conservation of single amino acids using the substitutions attribute of alignment objects in Biopython and concluded that there is evidence of stronger conservation between human, rhesus monkey, and rat compared to all six from single amino acid mutation data. This is in accordance with the alignment and tree I saw earlier.

**Part 3: investigating sequence conservation**

In addition to single amino acids, I looked at conservation of amino acid sequences. Specifically, I looked for motifs of at least 10 amino acids which were conserved across at least 80% of the proteins (5 out of 6) at each position. I obtained 9 sequences which met these criteria. I used the motif module to compute a consensus motif for each of these 9 sequences. Since these motifs stretched across the complete protein with some smaller and some larger gaps in between, they form additional evidence for evolutionary pressure, specifically, for evolutionary pressure on specific parts of the protein.

**Part 4: investigating conservation of protein domains**

To investigate whether these conserved motifs were part of functional domains in the protein, I first needed to know which common domains and corresponding sequences are present in the proteins. I compared the proteins to a common protein domain database (Pfam) and looked for matching domains. I obtained the domain sequences from a different source because of problems with obtaining the specific sequences from the Pfam database as can be read in the process book. Then I aligned the conserved motifs I found with the domain sequences. I found no evidence that the conserved sequences are part of any of these domains.

**Part 5: looking for a common ancestor of the two closest Akt1 genes**

Because the Akt1 DNA sequences from rhesus monkey and human in the initial ClustalW alignment were 96% similar, I was curious to see whether I could identify a common ancestor, which would be ultimate evidence that these are in fact the same genes. I did a blast of the human gene to the NCBI coding sequence database and used the results to draw a phylogenetic tree. I managed to do so, but it was not very insightful, because although both human and rhesus monkey were in the tree, the human gene grew directly from the root of the tree, disabling me from identifying a common ancestor. It probably makes sense that this happened because I used the human gene for blasting, but could not do further research because my time was up at this point.

**Conclusion**

High similarity between the Akt1 genes and proteins of diverse species in sequences as long as 54 amino acids with >80% conservation at each position provide evidence for evolutionary pressure on the Akt1 protein.