Data mining

To retrieve the data, we had four primary sources:

* Uniprot: a database of proteins, containing a variety of functional and structural information
* European Bioinformatics Institute which has also contains a variety of information on many proteins in an easily parsable XML format
* Ensembl which provides genomic information on a number of proteins and their transcripts
* PhosphoSitePlus: a database of post-translational modifications in human, mouse and rat proteins

The datamining was completed via python scripts and the various APIs available on the respective sources. The data was incorporated into a number of csv files, which could then be used to populate our database.

Kinase characteristics

For our database we required a variety of information on human kinases. To begin