**About**

Kinase Kinpendium is a functioning prototype of a web-based software tool for handling molecular biology data. This tool allows the user to explore background information about human protein kinases and the sites they phosphorylate and is capable of calculating the level of activity of each human kinase from experimental uploaded datasets. The web application was developed by a group of 5 students under the supervision of Professor Conrad Bessant and Dr. Fabrizio Smeraldi as part of the MSc Bioinformatics Software Development Group Project at Queen Mary University of London.

The Kinase Kinpendium website search tools are based on a curated database that holds general information about Protein Kinases ( ID, Uniprot Accession ID, Name, Symbol, Group, Family, Sub Family, Synonym, Function, Genomic Location, Subcellular Location), Inhibitors of Protein Kinases ( Name, Chembl ID, Smiles, Kinase Target) and Phosphosites (phosphorylation sites) for each protein kinase ( Substrate ID, Kinase Accession ID, Kinase Gene, Kinase Symbol, Substrate, Substrate Accession ID, Substrate Gene, Site, Domain, Field).

An important feature of Kinase Kinpendium is the Data Analysis tool which allows the user to upload quantitative phosphoproteomics data of tsv format. The web application summarises this data graphically and provides an estimate of the relative activity of every human kinase in the sample represented by the data.

Contents

**About**………………………………………………………………………………………………………………………………….

# 1. **Software outline**…………………………….………………………………………………………………………………

1.1 Software Schematics………………………………………………………………………………………………...

1.2 Software Architecture……………………………………………………………………………………………….

1.3 Information on how to run Kinase Kinpendium…………………………………………………………

1.4 Packages Used……………………………………………………………………………………………………….....

1.5 CSS/HTML………………………………………………………………………………………………………………….

2. **Data collection**………………………………………………………………………………………………………………..

2.1 Protein Kinase Data…………………………………………………………………………………………………..

2.2 Phosphosites Data…………………………………………………………………………………………………….

2.3 Inhibitor Data…………………………………………………………………………………………………………....

3. **Database Schema**…………………………………………………………………………………………………………….

4**. Kinase Kinpendium Features**……………………………………………………………………..........................

4.1 Kinases search function……………………………………………………………………...........................

.

4.2 Substrates search function……………………………………………………………………......................

4.3 Inhibitors search function……………………………………………………………………........................

4.4 Data Analysis tool: Uploading Phosphoproteomic data…………………………………………......

5. **Deploying Kinase Kinpendium**…………………………………………………………………….........................

5.1 Local Deployment…………………………………………………………………….....................................

6. **Limitations**………………………………………………..................………………………………………………...........

6.1 Web feature limitations………………………………………………..................................................

6.2 Database limitations………………………………………………..................…………………………………..

7. **Technical Solutions / Optimization**………………………………………………..................……………………

8. **Further Development**………………………………………………..................…………………………………….....

9. **References**………………………………………………..................………………………………………………...........

**2. Data collection**

2.1 Protein Kinase Data

The raw data for Protein Kinases was obtained from the KinHub website ( <http://www.kinhub.org/kinases.html> ). This version of the csv file contained information such as the Kinase Symbol, Name, Group, Family, Subfamily and Uniprot Accession Number. Furthermore, to approach the requirements of the group project, additional information such as the Synonym, Function, Genomic location, Subcellular location and a PDB image link were added and the data was curated. The gathering of the additional information was obtained by using the following websites: UniProt ( <https://www.uniprot.org> ), Ensembl ( <https://www.ensembl.org/index.html> ) and Protein Data Bank ( <http://www.rcsb.org> ). This final version of the file is available in both excel and csv data formats. Protein Kinase data was used for the Kinases search function to browse general information about protein kinases.

2.2 Phosphosites Data

Available Phosphosites data ( Kinase\_Substrate\_Dataset.gz and Phosphorylation\_site\_dataset.gz ) were downloaded from the PhosphoSitePlus website ( <https://www.phosphosite.org/homeAction.action> ). In order to obtain the relevant information from the raw downloaded files (human data and protein kinase post-translational modifications only), the datasets were filtered. Phosphosites data was mainly used for the Data Analysis tool as well as for the Substrates search function therefore allowing the user to browse for substrates phosphorylated by a protein kinase.

2.3 Inhibitor Data

The inhibitor data was gathered from PKIDB – A Curated, Annotated and Updated Database of Protein Kinase Inhibitors in Clinical Trials ( <http://www.icoa.fr/pkidb/index.html> ). This curated data contains information such as the Inhibitor Name, Chembl ID, Smiles and Kinase Target. Inhibitor data was used for the Inhibitors search function thus allowing the user to browse inhibitors of protein kinases.