Introduction

Protein kinases are a large family of enzymes that catalyze the phosphorylation of other proteins.[1] Phosphorylated proteins do specific functions. (Figure 1).

Aberrant kinase function is associated with cancer, immune system diseases and degenerative diseases.[2] Protein kinases are major drug targets [3].

The advances in enable the identification of phosphosites at the proteome level, most of the phosphoproteome is in the dark: more than 95% of all reported human phosphosites have no known kinase or associated biological function [4] (Figure 2).

A large fraction of the kinase is understudied[4]. For most of the kinases there are less than 10 known phosphosites (Figure 3).

DeepKinZero is a program that makes predictions for rare kinases, it first learn the association between the phosphosite and kinase embeddings. This idea is shown in Figure 4.

To learn phosphosite embeddings, Bi-directional Recurrent Neural Network (BRNN) [5] model is used with an attention mechanism over the training data. Figure 5 illustrates the DeepKinZero model.

Objectives

- Understanding how DeepKinZero works is essential to improve it.
- Understanding Gene Ontology to describe cellular location and obtaining kinase substrate annotations.
- Understanding and running GO semantic similarity tools leads to calculate similarity between kinases and proteins.
- Running DeepKinZero with the addition feature which is location information of kinases and proteins.
- Hyper parameter tuning makes input appropriate to use in DeepKinZero.