General outline of the dataframe:  
Meta data:

|  |  |  |  |
| --- | --- | --- | --- |
| Gene\_name | Protein\_name | Family | Cell\_location |
| MAPK1 | Mitogen-activated protein kinase 1 | CMGC | Nucleus |
|  |  |  |  |
|  |  |  |  |

Phosphosite:

|  |  |  |  |
| --- | --- | --- | --- |
| Gene\_name | phosphorylation\_sites | Genomic\_location | Neighbouring\_sequences |
| MAPK1 |  |  |  |
|  |  |  |  |

Inhibitors:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Inhibitors | Gene\_name | Structure (in jpg/png format) | Molecular\_weight | Empirical\_formula |
| [(5Z)-7-Oxozeaenol](http://www.kinase-screen.mrc.ac.uk/screening-compounds/349381) | TAK1 | oxozeaenol.jpg | 362.37 | C19H22O7 |
|  |  |  |  |  |

General rules and stuff:

1. For unavailable data, fill in ‘NULL’ .
2. Export the dataframe into csv and upload into the ‘Data\_mining’ folder
3. Inhibitor will need to attach the picture in jpg/png, the folder will possibly be very large, so might need to zip it.
4. For any functions created, please include a small annotation in the function:

-def get\_protein(input):

“””

This function does this

“””

return eg