

README

Usage example

```
from get_kinase_group import get_kinase_group
from get_substrates import get_substrates
from get_windows import get_windows
from force_motifs import forcemotif
from get_relevant_db import get_relevant_db
import pandas as pd
from Bio import SeqIO
from Bio.Alphabet import generic_protein
from Bio.Alphabet import IUPAC
from Bio.Seq import Seq
from Bio import motifs

my_kinases = get_kinase_group("./regPhos/RegPhos_kinase_human.txt", "CMGC")
my_kinases[0:5]

my_substrates = get_substrates("./regPhos/RegPhos_Phos_human.txt", my_kinases)
my_substrates.head()

my_substrates['kinase'][0:5]
my_substrates['substrates'][0]

fasta_db = SeqIO.parse("./ModelOrganisms/UP000005640_9606.fasta",
                        "fasta", IUPAC.extended_protein)

my_windows = []

for i in (my_substrates['substrates'].tolist()):
    fasta_db = SeqIO.parse("./ModelOrganisms/UP000005640_9606.fasta",
                            "fasta", IUPAC.extended_protein)
    relevant_db = get_relevant_db(fasta_db, i['AC'])
    my_windows.append(
        get_windows(
            relevant_db,
            i['AC'],
            i['position']))

my_windows[2].head()

my_motifs = [[] if len(window['window'])==0 else
              motifs.create(window['window']) for
              window in my_windows]

my_motifs[2].counts[0:2]
```

```

# PWM is acronym for position weight matrices

my_pwm = [[] if (len(m) == 0) else
            m.counts.normalize(pseudocounts=1) for
            m in my_motifs]

my_pwm[2]

# pssm is acronym for position specific scoring matrices

my_pssm = [[] if (len(pwm) == 0) else
            pwm.log_odds() for
            pwm in my_pwm]

my_pssm[2]

```

TODO list

- Check whether it is relevant to change the pseudocounts to remove the ambiguous aminoacid codes to the addition.