**Algorithm idea**

The hardest issue of the problem is probably to auto-download and parse response from remote host. If we do it (e.g. using urllib2 in python), the rest is trivial. Iterate over each letter in protein representation and check if it the following 4 letters are correct. You don't need any regexp nor other tools.

**Pseudocode**

**for protein in proteins:**

**for i=0 to length(protein)-3:**

**if protein[i] = 'N' and**

**protein[i+1] ≠ 'P' and**

**protein[i+2] = 'S' or protein[i+2] = 'T' and**

**protein[i+3] ≠ 'P' then:**

**print i**

**Complexity analyses and correctness**

The algorithm requires O(n)O(n) time where nn is summarize length of all proteins.