import re

# Defining the LRRK2 sequence as provided

lrrk2\_sequence = (

    "MASGSCQGCEEDEETLKKLIVRLNNVQEGKQIETLVQILEDLLVFTYSERASKLFQGKNI"

    "HVPLLIVLDSYMRVASVQQVGWSLLCKLIEVCPGTMQSLMGPQDVGNDWEVLGVHQLILK"

    "MLTVHNASVNLSVIGLKTLDLLLTSGKITLLILDEESDIFMLIFDAMHSFPANDEVQKLG"

    "CKALHVLFERVSEEQLTEFVENKDYMILLSALTNFKDEEEIVLHVLHCLHSLAIPCNNVE"

    "VLMSGNVRCYNIVVEAMKAFPMSERIQEVSCCLLHRLTLGNFFNILVLNEVHEFVVKAVQ"

    "QYPENAALQISALSCLALLTETIFLNQDLEEKNENQENDDEGEEDKLFWLEACYKALTWH"

    "RKNKHVQEAACWALNNLLMYQNSLHEKIGDEDGHFPAHREVMLSMLMHSSSKEVFQASAN"

    "ALSTLLEQNVNFRKILLSKGIHLNVLELMQKHIHSPEVAESGCKMLNHLFEGSNTSLDIM"

    "AAVVPKILTVMKRHETSLPVQLEALRAILHFIVPGMPEESREDTEFHHKLNMVKKQCFKN"

    "DIHKLVLAALNRFIGNPGIQKCGLKVISSIVHFPDALEMLSLEGAMDSVLHTLQMYPDDQ"

    "EIQCLGLSLIGYLITKKNVFIGTGHLLAKILVSSLYRFKDVAEIQTKGFQTILAILKLSA"

    "SFSKLLVHHSFDLVIFHQMSSNIMEQKDQQFLNLCCKCFAKVAMDDYLKNVMLERACDQN"

    "NSIMVECLLLLGADANQAKEGSSLICQVCEKESSPKLVELLLNSGSREQDVRKALTISIG"

    "KGDSQIISLLLRRLALDVANNSICLGGFCIGKVEPSWLGPLFPDKTSNLRKQTNIASTLA"

    "RMVIRYQMKSAVEEGTASGSDGNFSEDVLSKFDEWTFIPDSSMDSVFAQSDDLDSEGSEG"

    "SFLVKKKSNSISVGEFYRDAVLQRCSPNLQRHSNSLGPIFDHEDLLKRKRKILSSDDSLR"

    "SSKLQSHMRHSDSISSLASEREYITSLDLSANELRDIDALSQKCCISVHLEHLEKLELHQ"

    "NALTSFPQQLCETLKSLTHLDLHSNKFTSFPSYLLKMSCIANLDVSRNDIGPSVVLDPTV"

    "KCPTLKQFNLSYNQLSFVPENLTDVVEKLEQLILEGNKISGICSPLRLKELKILNLSKNH"

    "ISSLSENFLEACPKVESFSARMNFLAAMPFLPPSMTILKLSQNKFSCIPEAILNLPHLRS"

    "LDMSSNDIQYLPGPAHWKSLNLRELLFSHNQISILDLSEKAYLWSRVEKLHLSHNKLKEI"

    "PPEIGCLENLTSLDVSYNLELRSFPNEMGKLSKIWDLPLDELHLNFDFKHIGCKAKDIIR"

    "FLQQRLKKAVPYNRMKLMIVGNTGSGKTTLLQQLMKTKKSDLGMQSATVGIDVKDWPIQI"

    "RDKRKRDLVLNVWDFAGREEFYSTHPHFMTQRALYLAVYDLSKGQAEVDAMKPWLFNIKA"

    "RASSSPVILVGTHLDVSDEKQRKACMSKITKELLNKRGFPAIRDYHFVNATEESDALAKL"

    "RKTIINESLNFKIRDQLVVGQLIPDCYVELEKIILSERKNVPIEFPVIDRKRLLQLVREN"

    "QLQLDENELPHAVHFLNESGVLLHFQDPALQLSDLYFVEPKWLCKIMAQILTVKVEGCPK"

    "HPKGIISRRDVEKFLSKKRKFPKNYMSQYFKLLEKFQIALPIGEEYLLVPSSLSDHRPVI"

    "ELPHCENSEIIIRLYEMPYFPMGFWSRLINRLLEISPYMLSGRERALRPNRMYWRQGIYL"

    "NWSPEAYCLVGSEVLDNHPESFLKITVPSCRKGCILLGQVVDHIDSLMEEWFPGLLEIDI"

    "CGEGETLLKKWALYSFNDGEEHQKILLDDLMKKAEEGDLLVNPDQPRLTIPISQIAPDLI"

    "LADLPRNIMLNNDELEFEQAPEFLLGDGSFGSVYRAAYEGEEVAVKIFNKHTSLRLLRQE"

    "LVVLCHLHHPSLISLLAAGIRPRMLVMELASKGSLDRLLQQDKASLTRTLQHRIALHVAD"

    "GLRYLHSAMIIYRDLKPHNVLLFTLYPNAAIIAKIADYGIAQYCCRMGIKTSEGTPGFRA"

    "PEVARGNVIYNQQADVYSFGLLLYDILTTGGRIVEGLKFPNEFDELEIQGKLPDPVKEYG"

    "CAPWPMVEKLIKQCLKENPQERPTSAQVFDILNSAELVCLTRRILLPKNVIVECMVATHH"

    "NSRNASIWLGCGHTDRGQLSFLDLNTEGYTSEEVADSRILCLALVHLPVEKESWIVSGTQ"

    "SGTLLVINTEDGKKRHTLEKMTDSVTCLYCNSFSKQSKQKNFLLVGTADGKLAIFEDKTV"

    "KLKGAAPLKILNIGNVSTPLMCLSESTNSTERNVMWGGCGTKIFSFSNDFTIQKLIETRT"

    "SQLFSYAAFSDSNIITVVVDTALYIAKQNSPVVEVWDKKTEKLCGLIDCVHFLREVMVKE"

    "NKESKHKMSYSGRVKTLCLQKNTALWIGTGGGHILLLDLSTRRLIRVIYNFCNSVRVMMT"

    "AQLGSLKNVMLVLGYNRKNTEGTQKQKEIQSCLTVWDINLPHEVQNLEKHIEVRKELAEK"

    "MRRTSVE"

)

# Function to search for the motif

def find\_motif(sequence):

    motif = r"[WFY][A-Z]{2}[ILV]"

    matches = re.finditer(motif, sequence)

    results = [(match.start() + 1, match.group()) for match in matches]  # +1 to convert to 1-based index

    return results

# Finding motifs in the LRRK2 sequence

motif\_matches = find\_motif(lrrk2\_sequence)

# Display the results

import pandas as pd

df\_motif\_matches = pd.DataFrame(motif\_matches, columns=["Position", "Motif"])

import ace\_tools as tools; tools.display\_dataframe\_to\_user(name="LRRK2 Motif Matches", dataframe=df\_motif\_matches)

df\_motif\_matches