Introduction to Parameters Settings

GPQuest2 employs a JSON (JavaScript Object Notation) file for storing parameter information, where parameters are structured in a key-value format. The following section provides an explanation for each parameter within the file.

1. “Mass” : all mass values used in GPQuest
    a. “AA”: residue mass values of amino acids
    b. “GPF”: glycan fragment mass added on Y ions (peptide + glycan fragment)
        i. “mass”: added mass values
        ii. “charge”: related charge states
        iii. “names”: names of the Y ions
    c. “MIC”: all other mass values not classified
    d. “MOD”: mass values of modifications
    e. “GLYCAN”: residue mass values of monosaccharides
    f. “LABEL”: mass values of reporter ions
        i. “TMT10plex”/”TMT11plex”/”iTRAQ4plex”/”iTRAQ8plex”/
            1. “channel”: channel names of TMT labels
            2. “ions”: m/z values of reporter ions
            3. “FIXMOD”: additional mass value add on N-term or K
    g. “OXONIUM_IONS”: m/z values of oxonium ions
2. “WORKFLOW”: define steps used in GPQuest
    a. “PREPROCESS”: a list of steps used in preprocess
    b. “SEARCH”: search mode, now only support “ION INDEX SEARCH”
    c. “REFINE”: "Spectral Library Searching" (developing)
    d. “VALIDATION”: use "FDR" to estimate the false discovery rate
    e. “MSCORRECTION”: regenerate isotope peaks of precursor ions by considering the closest +/-N MS scans, default N=5
3. “SEARCH”: define searching settings
    a. “CORES”: use N thread if applicable, default N = 8
    b. “LABEL”: “None” for label free, ”TMT11plex”/”iTRAQ4plex” for TMT or iTRAQ labels defined in “MASS” section
    c. "MAX_RANK": Consider N best candidates in the final comparison, default N=5
    d. "MODE": "SPEG" (Deprecated option, will be removed in future, please do not change)
    e. “SCANS”: only search the list of target scans, default [] to search all scans
    f. “KEEP”: keep all glycan information, default “ALL”
    g. “SAVE_SPLIB”: save a spectral library for all the identified PSMs, default “YES”
    h. “FORECE_RENEW”: overwrite everything if rerun the code, default 0: not overwrite
    i. “MODIFICATIONS”: define peptide modifications
        i. “MAX_OPT_MOD”: max number of variable modifications, default: 2
        ii. “mods”: modifications added to the peptides: (definition same as MSGF+)
            1. "C2H3N1O1” : composition/composition mass
            2. “C”: amino acid
            3. “fix”: fixed or optional modification, options: fix, opt

4. “any”: positions of modification, options: any (anywhere), N-term (peptide N-term), C-term (peptide C-term), Prot-N-term (protein N-term), Prot-C-term (protein C-term)
5. “Carbamidomethyl”, modification name
6. “c”: abbreviation for modified amino acid

iii. "glycan_mods": (developing)

j. “MASS_TOLERANCE”:
   i. "MS1(PPM)": mass tolerance of MS peak, default: 10,
   ii. "MS2(PPM)": mass tolerance of MS/MS peak, default: 20,
   iii. "MIN": min value of mass tolerance, default: 0.005 Da

k. “CHARGE”: precursor charge range
   i. “max”: default: 5
   ii. “min”: default: 2

l. “MASS_SHIFT”: precursor mass shift, default [-2,-1,0]

m. “FRAGMENTS”: fragment types used in the search
   i. “CHARGE”: max charge state considered in searching, default:2
   ii. “IONS”: default [“b”,”y”]

n. “INTACT_GLYCOPEPTIDE”: Y ions required in the qualified PSMs
   i. “NUMBER”: number of Y ions required according to peptide length
   ii. “ESSENTIAL_IONS”: Y ions must be observed,default: []
   iii. “NONESSENTIAL_IONS”: name of Y ions and a list of charge states to search

o. “PEPTIDE_FRAGMENT”:
   i. “MIN_NUMBER”: min value of peptide fragments required for a qualified PSM
   ii. “MIN_RATIO”: min value of the intensity of peptide fragments required for a qualified PSM

p. “FDR”: false discovery rate

4. “PREPROCESS”: define settings for preprocess work
   a. “ISOTOPE_WINDOW”: (Deprecated option)
   b. “PEAK_SELECTION”: how many peaks used in search
      i. “TOPN”: top highest peaks, default: 100
   c. “OXONIUM_IONS”:
      i. “TOPN”: essential ions should be in the top N peaks, default N: 10
      ii. “ESS_IONS”: a list of essential oxonium ion names
      iii. “NON_ESS_IONS”: a list of non-essential oxonium ion names
5. “DATABASE”:
   a. “PATH”:
      i. “PROTEIN”: absolute path to the fasta file (.fasta) or glycopeptide file (.xlsx)
      ii. “NGLYCAN”: absolute path to the N-linked glycan file (.xlsx)
      iii. “OGLYCAN”: absolute path to the O-linked glycan file (.xlsx)
   b. “PROTEIN_DIGEST”:
      i. “PROTEASES”: a list of proteases for digestion, now only support single enzyme, enzyme name can be found in the documentation of pyteomics or https://web.expasy.org/peptide_cutter/
      ii. "MIN_PEP_LEN": min value of peptide length, default: 6,

iii. "MAX_PEP_LEN": max value of peptide length, default: 50,
iv. "MISSED_CLEAVAGES": number of missed cleavages, default: 2

c. “"ION_INDEX": ion index is the dictionary of fragments
i. "FACTOR": bin size of m/z value of fragments, default: 10,
ii. "MIN_IONS": min value of required matches, default: 6

NOTE: The red segments represent critical parameters for initiating a new search, encompassing databases, modifications, and sample labeling strategies. The remaining parameters can either remain at their default values or be adjusted for advanced configurations.