



MassChroQ library documentation

version: 3.0

Olivier Langella

Contents

Introduction	3
Input format	3
Output format	5
MSrun retention time alignment	8
Peak quality code	8
Match between run difference with legacy MassChroQ	8

Introduction

MassChroQ is designed to extract ion current of a list of peptide and measure the area under the curve on a single run.

Input format

input file format

```
{
  "project_parameters": {
    "i2MassChroQ_VERSION": {
      "category": 1,
      "value": "1.0.18"
    },
    "cvparam_SpectraDataFileFormat": {
      "category": 5,
      "value": "MS:1000544"
    },
    "cvparam_SpectraDataSpectrumIDFormat": {
      "category": 5,
      "value": "MS:1000824"
    },
    "AnalysisSoftware_name": {
      "category": 2,
      "value": "X!Tandem"
    },
  },
  "masschroq_methods": {
    "alignment_method": {
      "ms2_tendency": 10,
      "ms1_smoothing": 3,
      "ms2_smoothing": 5,
    },
    "quantification_method": {
      "match_between_run": true,
      "isotope_minimum_ratio": 0.9,
      "extraction": {
        "rt_range": 300,
        "integration": "max",
        "precision": {
          "unit": "ppm",
          "up": 10,
          "down": 10,
        }
      },
    },
    "prefilter": {},
    "detection": {
      "type": "zivy",
      "meanfilter": 1,
      "minmax": 3,
      "maxmin": 2,
      "threshold_on_max": 5000,
    },
  },
}
```

```
        "threshold_on_min": 3000,
      }
    },

    "identification_data": {
      "msrun_list": {
        "msrunb19": {
          "file": "/gorgone/pappso/data_extraction_pappso/
mzXML/20120906_balliau_extract_1_B08_teal-5.mzXML"
        }
      },
      "protein_list": {
        "protala1": {
          "description": "P02769|ALBU_BOVIN SERUM ALBUMIN PRECURSOR",
          "sequence": "SLTNDWEDHLAVK",
        },
      },
      "peptide_list": {
        "pepala1": {
          "proforma": "SLTNDWEDHLAVK",
          "proteins": ["protala1"],
          "mods": "free text",
          "label_list": { "light": {"proforma": "SLTNDWEDHLAVK"} }
        },
      },
    },

    "msrunpeptide_list": {"msruna1": {
      "peptide_obs": { "pepala1":
        [{
          "scan_index": 2345,
          "label": "light",
          "precursor": {
            "charge": 2,
            "mz": 2456.45,
            "intensity": 4580,
            "rt": 345.67,
          }
        }
      ],
    },
  },
},

"actions": {
  "group_list": { "g1": ["msruna1", "msruna2", "msruna3"]
  },
  "align_group": {"g1": {
```

```
    "alignment_reference": "msruna1"}
  },
  "quantify_all": true
}
}
```

CSV input file format

```
peptide proformat, charge, retention time
SLTNDWEDHLAVK, 2, 853.78
SLTNDWEDHLAVK, 3, 853.78
```

Output format

Wouldn't be nice to try CBOR ? CBOR is supported in QT6

It could be something liket that :

output file format

```
{
  "alignment_data" : [
    {
      "alignment_id": "a1",
      "group_id": "g1",
      "timestamp1": "ISODate"
      "alignment": {
        "msrun_ref": "msruna2",
        "corrections": {
          "msruna1": {
            "original": [1,2,3,4.5],
            "aligned": [1,2,3,4.5]
          },
          "msruna2": {
            "original": [1,2,3,4.5],
            "aligned": [1,2,3,4.5]
          }
        }
      },
      "timestamp2": "ISODate"
    }
  ],
  "quantification_data" : [
    {
      "timestamp1": "ISODate",
      "quantify_id": "q1",
      "group_id": "g1",
      "first_pass": {"msruna1": "qr_data_block", "msruna2": "qr_data_block"},
      "timestamp2": "ISODate",
      "second_pass": {"msruna1": "qr_data_block", "msruna2": "qr_data_block"},
      "timestamp3": "ISODate"
    }
  ],
}
```

```
"end":{"timestamp":"2025-01-27T15:30:11"}
}
```

QrDataBlock element

QrDataBlock stands for **quantification run data block** it handles quantification data for a single MSrun.

qr_data_block element format

```
{ "msrunal" : {
  "msrun" : { "id": "msrunal", "filename": "/gorgone/truc", "sample":
"échantillon" },
  "peptide_measurements": { "pepalal":
  {
    "proforma": "SLTNDWEDHLAVK",
    "mods": "free text",
    "rt_target": 853.78,
    "xics": [
      {
        "mz": 764.3755373,
        "xic_coord": { "mz_range": [764.344, 765.56666], "ims_im_index":
[150, 175] },
        "charge": 2,
        "isotope": 0,
        "rank": 1,
        "th_ratio": 0.568912,
        "quality": "a",
        "label": "light",
        "trace": {
          "x": [851.78, 852.78, 853.78, 854.78],
          "y": [851.78, 852.78, 853.78, 854.78]
        },
        "peak_shape": {
          "trace": {
            "x": [851.78, 852.78, 853.78, 854.78],
            "y": [851.78, 852.78, 853.78, 854.78]
          }
        },
        "peak": {
          "area": 450245623,
          "max_intensity": 2345.456,
          "rt": [851.78, 852.78, 853.78],
          "aligned_rt": [851.78, 852.78, 853.78]
        }
      }
    ],
    "mz": 764.8769,
    "charge": 2,
    "isotope": 1,
    "quality": "a",
    "trace": {
      "x": [851.78, 852.78, 853.78, 854.78],
      "y": [851.78, 852.78, 853.78, 854.78]
    }
  }
}
}
```

```
        },
        "peak": {
            "area": 550245623,
        }
    }
]
},
"pepala2": {
},
}
}
}
```

output file format

```
{
"alignment_data" : [
    {
        "alignment_id": "a1",
        "group_id": "g1",
        "timestamp1": "ISODate"
        "alignment": {
            "msrun_ref": "msruna2",
            "corrections": {
                "msruna1": {
                    "original": [1,2,3,4.5],
                    "aligned": [1,2,3,4.5]
                },
                "msruna2": {
                    "original": [1,2,3,4.5],
                    "aligned": [1,2,3,4.5]
                }
            }
        },
        "timestamp2": "ISODate"
    },
],
"quantification_data" : [
    {
        "timestamp1": "ISODate",
        "quantify_id": "q1",
        "group_id": "g1",
        "first_pass": {"msruna1": "qr_data_block", "msruna2": "qr_data_block"},
        "timestamp2": "ISODate",
        "second_pass": {"msruna1": "qr_data_block", "msruna2": "qr_data_block"},
        "timestamp3": "ISODate"
    }
],
"end": {"timestamp": "2025-01-27T15:30:11"}
}
```

MSrun retention time alignment

MS run retention time function

```
std::shared_ptr<pappso::MsRunRetentionTime<QString>> &  
mcql::MsRunPeptideList::buildMsRunRetentionTimeSp(const mcql::AlignmentMethodSp  
&alignment_method)
```

Peak quality code

- aa** best quality : many MS2 fragmentation event, only one peak directly detected
- zaa** same as aa, but this charge state was not directly observed in MS2 fragmentation events in this MSrun
- a** good quality, single MS2 fragmentation event, one peak detected
- za** same as a, but this charge state was not directly observed in MS2 fragmentation events in this MSrun
- ab** many MS2 fragmentation event, but more than one peak detected, the greater peak (area) is chosen, it is obviously fragmented... perhaps a hint to check for peak detection parameters
- zab** same as ab, but this charge state was not directly observed in MS2 fragmentation events in this MSrun
- b** peak obtained by “match between run” on the mean aligned observed retention times in MS2 fragmentation events **and** also matching with the retention time given by other detected and quantified MS1 apex peaks in other MS runs
- c** peak obtained by “match between run” only on the mean of aligned observed retention times in MS2 fragmentation events
- d** peak obtained by “match between run” only matching with the retention time given by other detected and quantified MS1 apex peaks in other MS runs
- missed** no peak detected, no quantification

Match between run difference with legacy

MassChroQ

The match between run process behaviour is slightly different between legacy MassChroQ and MassChroQlite. The new process is more conservative as it will not try to find different peptide charge state if the peptide was observed for an MSrun in an other charge state.

An important difference is that MassChroQlite reports missed peaks : this lead to a natural data cleaning when used with MCQR. Indeed, if the most theoretical abundant isotope is not found in **any** msruns, then it will be discarded. This lead to less quantified peptides but with a better quality.