QuaMeter "IDFree" Manual

David L. Tabb, September 10, 2012

Background

QuaMeter [1] was created to compute the NIST quality metrics [2] for LC-MS/MS experiments in an open-source, reliable framework that was compatible, via ProteoWizard [3], with data from multiple instrument vendors. In its initial implementation, the software was intended for use in data generated from a common sample throughout a long time series. To meet the needs of the NCI Clinical Proteomic Tumor Analysis for Cancer (CPTAC), the Data Integration and Bioinformatics Working Group sought a tool that could generate metrics from data representing arbitrary complex samples, rather than the simple defined protein mixtures in common use for quality assessment. The "IDFree" mode of QuaMeter was produced in the Tabb Laboratory in response to this need.

Installation and Execution

QuaMeter is available from the Tabb Laboratory website: http://fenchurch.mc.vanderbilt.edu/. The link from the Software page will take users to the "TeamCity" server, where they should log in as guest users. At present, QuaMeter is available for Linux and Windows platforms. To take advantage of the software's support for instrument-native formats, users must deploy the QuaMeter build for "Windows x86" because these libraries are inaccessible under Linux. Libraries to read native formats are included in the QuaMeter installation. The software is made available in a .tar.bz2 format. Many users have found that 7-Zip (http://www.7-zip.org/) is helpful for decompressing the "bz2" file and expanding the resulting "tar" file to a directory. This directory can then be relocated to its proper location, such as "C:\Program Files (x86)."

The QuaMeter software features only a command-line interface, at present. One suggested structure for the command line looks like the following:

```
"C:\Program Files (x86)\quameter-bin-windows-x86-vc100-release-1_0_74\quameter.exe" *.raw -MetricsType idfree -OutputFilepath metrics.tsv -cpus 1
```

The "`*.raw`" parameter specifies that all Thermo RAW files in the current directory should be analyzed by the software; a wide variety of formats are supported, such as mzML, mzXML, and native formats from Bruker and Agilent. Conversion to mzML format by the AB Sciex converter is currently recommended for WIFF files from QSTAR instruments to read precursor charges, though QTRAP instruments can be handled natively.

The "`-MetricsType idfree`" flag causes the software to operate in identification-independent mode rather than the published identification-dependent mode.

By default, QuaMeter will direct its generated metrics to standard output. The "`-OutputFilepath metrics.tsv`" parameter will cause these values to be written to a tab-delimited file of this name, with metrics for each input file listed in a row.

QuaMeter reports its progress more verbosely when it is limited to processing a single file at a time. The "`-cpus 1`" flag ensures this to be the case. Processing multiple input files at once is possible, but it places a higher load on I/O performance (CPU power is less a concern in generating these metrics).

For best performance, you should create a quameter.cfg file in the directory containing the input files. We recommend that you employ a file like one the following:

Quadrupole mass analyzer example

```
ChromatogramMzLowerOffset = 0.5mz
ChromatogramMzUpperOffset = 0.5mz
```

FT mass analyzer mass example

```
ChromatogramMzLowerOffset = 10ppm
ChromatogramMzUpperOffset = 10ppm
```

The lower and upper offsets are used to define how wide an m/z region is summed to produce chromatograms for each precursor. Users should *not* include a space character between the number and the unit for these offsets.

Software run-time will typically be less than two minutes per LC-MS/MS experiment. This can be considerably longer for profile data where peaklisting must be conducted, for extremely long LC gradients, or for slow disk transfers or CPU performance.

Metric Definitions

The metrics computed by QuaMeter in IDFree mode diverge significantly from the set described in the NIST metrics paper [2]. The full set of IDFree metrics are enumerated in the table on the next page. The NIST metrics frequently give greater attention to precursor ions that give rise to identified spectra; QuaMeter stratifies MS features into three categories:

- Ions that are observed in MS scans only, contributing to the TIC of an MS spectrum.
- Ions that are selected for tandem mass spectrometry.
- Ions that are selected for tandem mass spectrometry and that have sustained extracted ion chromatograms (XICs).

The definition of this final category needs further description. XICs are generated for all MS/MS precursors, and the peaks are sorted in descending order of peak width. The set of peaks that account for the top half of peak width are emphasized in peak height and peak width metrics (XIC-FWHM-* and XIC-Height-* metrics).

Filename	What is the name of the file from which the metrics were computed?
StartTimeStamp	At what time did acquisition begin for this experiment?
XIC-WideFrac	What fraction of precursor ions account for the top half of all peak width?
XIC-FWHM-Q1	What is the 25%ile of peak widths for the wide XICs?
XIC-FWHM-Q2	What is the 50%ile of peak widths for the wide XICs?
XIC-FWHM-Q3	What is the 75%ile of peak widths for the wide XICs?
XIC-Height-Q2	The log ratio for 50%ile of wide XIC heights over 25%ile of heights.
XIC-Height-Q3	The log ratio for 75%ile of wide XIC heights over 50%ile of heights.
XIC-Height-Q4	The log ratio for maximum of wide XIC heights over 75%ile of heights.
RT-Duration	What is the highest scan time observed minus the lowest scan time observed?
RT-TIC-Q1	The interval when the first 25% of TIC accumulates divided by RT-Duration
RT-TIC-Q2	The interval when the second 25% of TIC accumulates divided by RT-Duration
RT-TIC-Q3	The interval when the third 25% of TIC accumulates divided by RT-Duration
RT-TIC-Q4	The interval when the fourth 25% of TIC accumulates divided by RT-Duration
RT-MS-Q1	The interval for the first 25% of all MS events divided by RT-Duration
RT-MS-Q2	The interval for the second 25% of all MS events divided by RT-Duration
RT-MS-Q3	The interval for the third 25% of all MS events divided by RT-Duration
RT-MS-Q4	The interval for the fourth 25% of all MS events divided by RT-Duration
RT-MSMS-Q1	The interval for the first 25% of all MS/MS events divided by RT-Duration
RT-MSMS-Q2	The interval for the second 25% of all MS/MS events divided by RT-Duration
RT-MSMS-Q3	The interval for the third 25% of all MS/MS events divided by RT-Duration
RT-MSMS-Q4	The interval for the fourth 25% of all MS/MS events divided by RT-Duration
MS1-TIC-Change-Q2	The log ratio for 50%ile of TIC changes over 25%ile of TIC changes
MS1-TIC-Change-Q3	The log ratio for 75%ile of TIC changes over 50%ile of TIC changes
MS1-TIC-Change-Q4	The log ratio for largest TIC change over 75%ile of TIC changes
MS1-TIC-Q2	The log ratio for 50%ile of TIC over 25%ile of TIC
MS1-TIC-Q3	The log ratio for 75%ile of TIC over 50%ile of TIC
MS1-TIC-Q4	The log ratio for largest TIC over 75%ile TIC
MS1-Count	How many MS scans were collected?
MS1-Freq-Max	What was the fastest frequency for MS collection in any minute? (Hz)
MS1-Density-Q1	What was the 25%ile of MS scan peak counts?
MS1-Density-Q2	What was the 50%ile of MS scan peak counts?
MS1-Density-Q3	What was the 75%ile of MS scan peak counts?
MS2-Count	How many MS/MS scans were collected?
MS2-Freq-Max	What was the fastest frequency for MS/MS collection in any minute? (Hz)
MS2-Density-Q1	What was the 25%ile of MS/MS scan peak counts?
MS2-Density-Q2	What was the 50%ile of MS/MS scan peak counts?
MS2-Density-Q3	What was the 75%ile of MS/MS scan peak counts?
MS2-PrecZ-1	What fraction of MS/MS precursors is singly charged?
MS2-PrecZ-2	What fraction of MS/MS precursors is doubly charged?
MS2-PrecZ-3	What fraction of MS/MS precursors is triply charged?
MS2-PrecZ-4	What fraction of MS/MS precursors is quadruply charged?
MS2-PrecZ-5	What fraction of MS/MS precursors is quintuply charged?
MS2-PrecZ-more	What fraction of MS/MS precursors is charged higher than +5?
MS2-PrecZ-likely-1	What fraction of MS/MS precursors lack known charge but look like +1s?
MS2-PrecZ-likely-multi	What fraction of MS/MS precursors lack known charge but look like >+1s?

Most of the QuaMeter IDFree metrics form sets. For example, RT-MSMS-Q1, RT-MSMS-Q2, RT-MSMS-Q3, and RT-MSMS-Q4 all describe the distribution of retention times for MS/MS scans throughout an LC-MS/MS experiment. RT-MSMS-Q1 describes the duration during which the first 25% of MS/MS events occur, divided by the total duration during which MS/MS events took place. The sum of these four metrics is always one. The same can be said for many of the other sets of metrics.

MS2-PrecZ-likely-1 and MS2-PrecZ-likely-multi may not, at first, seem familiar to researchers. In fact, they represent a common strategy for identifying peptides in data sets where precursor charges are unknown (such as those from quadrupole ion traps). Each MS/MS spectrum may have 90% of its fragment intensity below a precursor, in which case most software assumes that it represents a singly-charged peptide. An MS/MS spectrum in which more than 10% of observed fragment intensity appears above the precursor is generally assumed to be either doubly- or triply-charged. For the full set of precursors, QuaMeter subdivides those for which charge is unknown into these two categories.

While these metrics are all subject to change as the software continues to evolve, the authors anticipate that this set will remain essentially stable while greater attention focuses on the interpretation of these values.

Interpretation

The ability to generate metrics for LC-MS/MS experiments without resorting to identification is valuable. Making decisions on the basis of these data, however, is a considerable research challenge. A univariate approach might generate a Shewhart process control chart to report performance over time for a given instrument. By asking when metrics fall outside upper and lower control limits (typically two standard deviations), the researcher identifies outliers. For this set of 45 metrics, however, one would expect 2 metrics to be "out of control" by random chance, if they were independently variant. That assumption, of course, is also flawed; metrics such as MS2-PrecZ-3 and MS2-PrecZ-4 tend to be correlated. Multivariate techniques are necessary to subdivide these metrics to uncorrelated components (as in principal components analysis) or to generate metrics that otherwise summarize these values while accounting for the correlation.

The following visualizations are frequently useful for evaluating metrics across several files:

- How is the total ion current (TIC) distributed among MS scans? Create a stacked bar chart for the following fields: RT-TIC-Q1, RT-TIC-Q2, RT-TIC-Q3, RT-TIC-Q4. Typically, Q1 and Q4 bars will be longest, suggesting that the average TIC is lower for early and late MS scans than it is for MS scans in the second and third quartiles.
- How do MS/MS scans distribute over the course of LC-MS/MS? Create a stacked bar chart for the following fields: RT-MSMS-Q1, RT-MSMS-Q2, RT-MSMS-Q3, RT-MSMS-Q4. If the instrument only produces MS/MS scans when it finds charge states above +1, these scans will likely be more frequent when many peptides are eluting from chromatography.

- How stable is MS TIC from scan to scan? Create a plot of the MS1-TIC-Change-Q4 field for a set of files. This metric evaluates the TIC changes from one scan to the next. A high log ratio means that a file has a really huge jump in intensity between a pair of successive MS scans; this may mean electrospray instability or a chemical contaminant.
- How variable was MS/MS acquisition? It may be useful to produce an X-Y plot of these metrics: MS2-Count, MS2-Freq-Max. The former shows how many MS/MS scans were found throughout the file, while the second asks the fastest rate of MS/MS acquisition for a minute-long interval. By dividing MS2-Count by RT-Duration, you can compute the MS2-Freq-Avg to compare with MS2-Freq-Max.

Good luck with QuaMeter, and we would be glad to hear about your experiences with it!

References

1. Ma Z-Q, Polzin KO, Dasari S, Chambers MC, Schilling B, Gibson BW, et al. QuaMeter: Multivendor Performance Metrics for LC-MS/MS Proteomics Instrumentation. Anal. Chem. 2012 Jul 17;84(14):5845–5850. PMID: 22697456

2. Rudnick PA, Clauser KR, Kilpatrick LE, Tchekhovskoi DV, Neta P, Blonder N, et al. Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses. Mol. Cell Proteomics. 2010 Feb;9(2):225–241. PMID: 19837981

3. Kessner D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics. 2008 Nov 1;24(21):2534–2536. PMID: 18606607