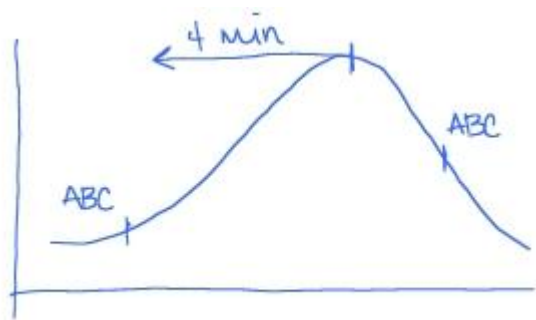


CMathematical Metric Definitions

The table below shows the definition of the metrics given by ‘Rudnick and Stein, NIST Performance Metrics, MCP 2010’, as well as the mathematical definition we defined based off of those given definition.

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-1A	Chromatography	Fraction of repeat peptide IDs with divergent RT	-4 min	Fraction	↓	Estimates very early peak broadening	Fraction of all peptides identified at least 4 min earlier than max MS1 for ID

Mathematical Definition:



for each peptide:

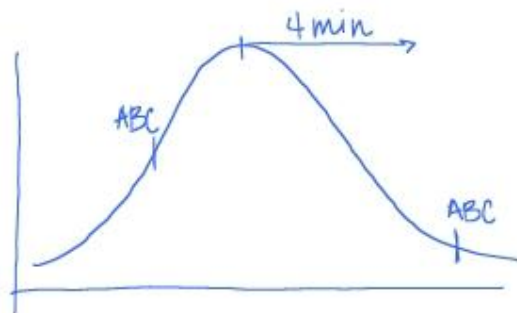
sort by peptide IDs
Peptide_Sequence found in xt

$$\Rightarrow \frac{\text{\# of peptides identified 4 min earlier than the max MS1}}{\text{\# of all peptides}}$$

link ScanNumber (found in xt) to FragScanNumber (found in SICstats)
link OptimalPeakApexScanNumber (found in SICstats) with ScanTime (found in ScanStats)
link FragScanNumber (found in SICstats) to ScanTime (found in SICstats)

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-1B	Chromatography	Fraction of repeat peptide IDs with divergent RT	+4 min	Fraction	↓	Estimates very late peak broadening	Fraction of all peptides identified at least 4 min later than max MS1 for ID

Mathematical Definition:



for each peptide:

sort by peptide IDs
Peptide_Sequence found in xt

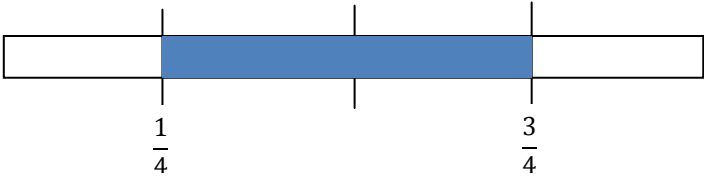
$$\Rightarrow \frac{\text{\# of peptides identified 4 min later than the max MS1 intensity}}{\text{\# of all peptides}}$$

link ScanNumber (found in xt) to FragScanNumber (found in SICstats)
link OptimalPeakApexScanNumber (found in SICstats) with ScanTime (found in ScanStats)
link ScanNumber (found in xt) to FragScanNumber (found in SICstats)

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-2A	Chromatography	Interquartile retention time period	Period (min)	min	↑	Longer times indicate better chromatographic separation	Time period over which 50% of peptides are identified

Mathematical Definition:

filter first: $\text{Peptide_Expected_Value_Log}(e) \leq \text{Max_Log_EValue} = -2$



$$\text{begin} = \frac{1}{4} [\text{total number of scans with identified peptides}]$$

$$\text{end} = \frac{3}{4} [\text{total number of scans with identified peptides}]$$

filter by $\text{Peptide_Expectation_Value_Log}(e)$
 $\text{Peptide_Expectation_Value_Log}(e)$ is found in xt

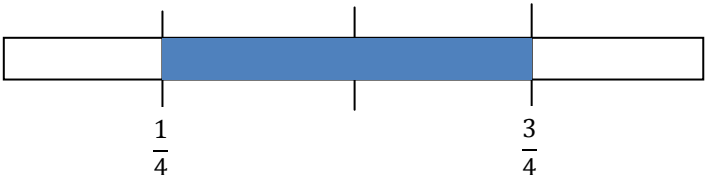
match the scan number to the scan time, and convert the scan time into minutes

$$\Rightarrow \text{end time} - \text{begin time}$$

using ScanNumber in ScanStats
using ScanTime in ScanStats

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-2B	Chromatography	Interquartile retention time period	Pep ID rate	Peps/min	↑	Higher rates indicate efficient sampling and identification	Rate of peptide identification during C-2A

Mathematical Definition:



$$\text{begin} = \frac{1}{4} [\text{total number of scans with identified peptides}]$$

$$\text{end} = \frac{3}{4} [\text{total number of scans with identified peptides}]$$

$$\Rightarrow \frac{\text{begin scan} \leq \# \text{ peptides identified} \leq \text{end scan}}{\text{Period}(\text{min})}$$

using ScanNumber in ScanStats
using ScanTime in ScanStats

C-3A	Chromatography	Peak width at half-height for IDs	Median value	S	↓	Sharper peak widths indicate better chromatographic separation	Median peak widths for all identified unique peptides (s)
------	----------------	-----------------------------------	--------------	---	---	--	---

Mathematical Definition:

create an array of peak widths
for non-unique peptides, use the peptide with the lowest Peptide_Expectation_Value_Log(e)
peak width - FWHMInScans is found in SICstats
Peptide_Expectation_Value_Log(e) is found in xt

convert the peak width from minutes to seconds

$$\Rightarrow \text{median peak width}$$

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-3B	Chromatography	Peak width at half-height for IDs	Interquartile distance	S	↓	Tighter distributions indicate more peak width uniformity	Measure of the distribution of the peak widths; small values indicate consistency

Mathematical Definition:

$$\text{begin} = \frac{1}{4} [\text{total number of scans with identified peptides}]$$

$$\text{end} = \frac{3}{4} [\text{total number of scans with identified peptides}]$$

using ScanNumber in ScanStats
link ScanNumber (in ScanStats) to FragScanNumber (in SICstats)

create an array of peak widths
for non-unique peptides, use the peptide with the lowest Peptide_Expectation_Value_Log(e)
peak width – FWHMInScans is found in SICstats
Peptide_Expectation_Value_Log(e) is found in xt

$$\text{begin scan} \leq \text{peak width} \leq \text{end scan}$$

convert the peak width from minutes to seconds

⇒ median peak width

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-4A	Chromatography	Peak widths at half-max over RT deciles for IDs	First decile	s	↓	Estimates peak widths at the beginning of the gradient	Median peak width for identified peptides in last RT decile (late)

Mathematical Definition:

begin scan = first identified peptide

end scan = last identified peptide

total = end scan – begin scan
total number of scans

using ScanNumber in ScanStats
link ScanNumber (in ScanStats) to FragScanNumber (in SICstats)

last decile = end scan – $\frac{1}{10}$ [total]
subtract $\frac{1}{10}$ from the end scan to find the last 10th in the retention spread

last decile ≤ peak widths ≤ end scan
put peak widths in array and sort array
peak width – FWHMInScans is found in SICstats

convert the peak width from minutes to seconds

⇒ median peak width

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-4B	Chromatography	Peak widths at half-max over RT deciles for IDs	Last decile	s	↓	Estimates peak widths at the end of the gradient	Median peak widths for identified peptides in first RT decile (early)

Mathematical Definition:

begin scan = first identified peptide

end scan = last identified peptide

total = end scan – begin scan
total number of scans

using ScanNumber in ScanStats
link ScanNumber (in ScanStats) to FragScanNumber (in SICstats)

first decile = $\text{begin scan} - \frac{1}{10}[\text{total}]$
add $\frac{1}{10}$ to the begin scan to find the first 10th

begin scan ≤ peak widths ≤ first decile
put peak widths in array and sort array
peak width – FWHMInScans is found in SICstats

convert the peak width from minutes to seconds

⇒ median peak width

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-4C	Chromatography	Peak widths at half-max over RT deciles for IDs	Median value	s	↓	Estimates peak widths in the middle of the gradient	Median peak width for identified peptides in median RT decile (middle)

Mathematical Definition:

begin scan = first identified peptide

end scan = last identified peptide

total = end scan – begin scan
total number of scans

using ScanNumber in ScanStats
link ScanNumber (in ScanStats) to FragScanNumber (in SICstats)

middle_b decile = begin scan + $\frac{4.5}{10}$ [total]
add $\frac{4.5}{10}$ to begin scan to find the beginning of the middle 10th

middle_e decile = begin scan + $\frac{5.5}{10}$ [total]
subtract $\frac{5.5}{10}$ from end scan to find the end of the middle 10th

middle_b decile ≤ peak widths ≤ middle_e decile
put peak widths in array and sort array
peak width – FWHMInScans is found in SICstats

convert the peak width from minutes to seconds

⇒ median peak width

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
C-5A	Chromatography	Average elution order differences	Between	Percent	↓	Estimates peptide elution similarity run to run	Average elution rank order difference for identified peptides between series
C-5B	Chromatography	Average elution order differences	Betw/in	Ratio	↓	Estimates peptide elution similarity between series	Ratio of average rank order difference between series to average rank order differences within a series (low values indicate similarity between series)
C-6A	Chromatography	Fraction of extra early eluting peptides in row series (– = fewer)	Between	Fraction	↓	Used to detect differences in the numbers of early peptides	Estimates relative frequency of early eluting peptides
C-6B	Chromatography	Fraction of extra late eluting peptides in row series (– = fewer)	Between	Fraction	↓	Used to detect differences in the numbers of late peptides	Estimates relative frequency of late eluting peptides

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
DS-1A	Dynamic Sampling	Ratio of peptide ions IDed by different numbers of spectra	Once/twice	Ratio	↑	Estimates oversampling	Ratio of peptides identified by 1 spectrum divided by number identified by 2 spectra

Mathematical Definition:

a.

$$\Rightarrow \frac{\text{\# of peptides identified by 1 spectra}}{\text{\# of peptides identified by 2 spectra}}$$

b.

$$\Rightarrow \frac{\text{\# of peptides identified by 1 spectra}}{\text{total \# of peptides identified}} \\ \text{spectra - ?}$$

DS-1B	Dynamic Sampling	Ratio of peptide ions IDed by different numbers of spectra	Twice/thrice	Ratio	↑	Estimates oversampling	Ratio of peptides identified by 2 spectra divided by number identified by 3 spectra
-------	------------------	--	--------------	-------	---	------------------------	---

Mathematical Definition:

a.

$$\Rightarrow \frac{\text{\# of peptides identified by 2 spectra}}{\text{\# of peptides identified by 3 spectra}}$$

b.

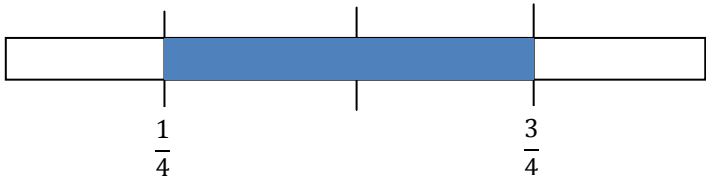
$$\Rightarrow \frac{\text{\# of peptides identified by 2 spectra}}{\text{total \# of peptides identified}}$$

c.

$$\Rightarrow \frac{\text{\# of peptides identified by 3 spectra}}{\text{total \# of peptides identified}} \quad (?) \\ \text{spectra - ?}$$

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
DS-2A	Dynamic Sampling	Spectrum counts	MS1 scans/full	Count	↓	Fewer MS1 scans indicates more sampling	Number of MS1 scans taken over C-2A

Mathematical Definition:



$$\text{begin} = \frac{1}{4} [\text{total number of scans with identified peptides}]$$

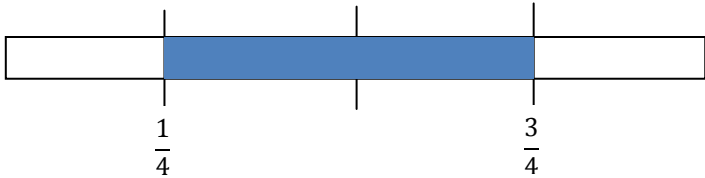
$$\text{end} = \frac{3}{4} [\text{total number of scans with identified peptides}]$$

filter by Peptide_Expectation_Value_Log(e)
 Peptide_Expectation_Value_Log(e) is found in xt

⇒ begin scan ≤ total # MS1 scans ≤ end scan
 use scan type to determine if its MS1 or MS2
 use ScanType found in ScanStats

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
DS-2B	Dynamic Sampling	Spectrum counts	MS2 scans	Count	↑	More MS2 scans indicates more sampling	Number of MS2 scans taken over C-2A

Mathematical Definition:



$$\text{begin} = \frac{1}{4} [\text{total number of scans with identified peptides}]$$

$$\text{end} = \frac{3}{4} [\text{total number of scans with identified peptides}]$$

filter by Peptide_Expectation_Value_Log(e)
 Peptide_Expectation_Value_Log(e) is found in xt

⇒ begin scan ≤ total # MS2 scans ≤ end scan
 use scan type to determine if its MS1 or MS2
 scan type 2 or higher
 use ScanType found in ScanStats

DS-3A	Dynamic Sampling	MS1 max / MS1 sampled abundance ratio IDs	Median all IDs	Ratio	↓	Estimates position on peak where sampled for peptides of all abundances	Ratio of MS1 maximum to MS1 value at sampling for median decile of peptides by MS1 maximum intensity (1 = sampled at peak maxima)
-------	------------------	---	----------------	-------	---	---	---

Mathematical Definition:

for all peptides calculate:

$$\frac{\text{PeakMaxIntensity}}{\text{ParentIonIntensity}}$$

use PeakMaxIntensity found in SICstats
 use ParentIonIntensity found in SICstats

⇒ median value

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
DS-3B	Dynamic Sampling	MS1 max/ MS1 sampled abundance ratio IDs	Med bottom 1/2	Ratio	↓	Estimates position on peak where sampled for least abundant 50% of peptides	Ratio of MS1 maximum to MS1 value at sampling for bottom 50% of peptides by MS1 maximum intensity (1 = sampled at peak maxima)

Mathematical Definition:

sort by MS1 intensities
take the bottom 50%

for all peptides calculate:

$$\frac{\text{PeakMaxIntensity}}{\text{ParentIonIntensity}}$$

use PeakMaxIntensity found in SICstats
use ParentIonIntensity found in SICstats

⇒ median value

IS-1A	Ion Source	MS1 during middle (and early) peptide retention period	MS1 jumps >10x	Count	↓	Flags ESI instability	Number of times where MS1 signal greatly decreased between adjacent scans more than 10-fold (electrospray instability)
-------	------------	--	----------------	-------	---	-----------------------	--

Mathematical Definition:

compare sequential scans and calculate the intensity difference between them
increment count if the MS1 intensity decreases (falls) more than 10-fold (10x)
done over the entire run

⇒ # MS1 intensity falls > 10-fold between sequential scans
use scan type 1
use ScanType found in ScanStats
use BasePeakIntensity found in ScanStats

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
IS-1B	Ion Source	MS1 during middle (and early) peptide retention period	MS1 falls >10x	Count	↓	Flags ESI instability	Number of times where MS1 signal greatly increases between adjacent scans more than 10-fold (electrospray instability)

Mathematical Definition:

compare sequential scans and calculate the intensity difference between them
increment count if the MS1 intensity increases (jumps) more than 10-fold (10x)
done over the entire run

⇒ # MS1 intensity jumps > 10-fold between sequential scans
use scan type 1
use ScanType found in ScanStats
use BasePeakIntensity found in ScanStats

IS-2	Ion Source	Precursor m/z for IDs	Median	Th	↓	Higher median m/z can correlate with inefficient or partial ionization	Median m/z value for all identified peptides (unique ions)
------	------------	-----------------------	--------	----	---	--	--

Mathematical Definition:

filter first: $\text{Peptide_Expected_Value_Log}(e) \leq \text{Max_Log_EValue} = -2$

create an array of precursor m/z values with only unique ions
use mz column in SICstats

⇒ return the median

Code	Category	Metric Group	Metric	Units	Optimal	Purpose/Use	Description
IS-3A	Ion Source	IDs by char state (relative to 2+)	Charge 1+	Ratio	↓	High ratio of 1+ / 2+ peptides may indicate inefficient ionization	Number 1+ peptides over 2+ peptides

Mathematical Definition:

filter first: $\text{Peptide_Expected_Value_Log}(e) \leq \text{Max_Log_EValue} = -2$

$$\Rightarrow \frac{\text{total \# of peptides with 1+ charge}}{\text{total \# of peptides with 2+ charge}}$$

use Charge found in xt

IS-3B	Ion Source	IDs by char state (relative to 2+)	Charge 3+	Ratio	↓	High ratio of 3+ / 2+ peptides may indicate inefficient ionization	Number 3+ peptides over 2+ peptides
-------	------------	---------------------------------------	-----------	-------	---	--	-------------------------------------

Mathematical Definition:

filter first: $\text{Peptide_Expected_Value_Log}(e) \leq \text{Max_Log_EValue} = -2$

$$\Rightarrow \frac{\text{total \# of peptides with 3+ charge}}{\text{total \# of peptides with 2+ charge}}$$

use Charge found in xt

IS-3C	Ion Source	IDs by char state (relative to 2+)	Charge 4+	Ratio	↓	High ratio of 4+ / 2+ peptides may indicate inefficient ionization	Number 4+ peptides over 2+ peptides
-------	------------	---------------------------------------	-----------	-------	---	--	-------------------------------------

Mathematical Definition:

filter first: $\text{Peptide_Expected_Value_Log}(e) \leq \text{Max_Log_EValue} = -2$

$$\Rightarrow \frac{\text{total \# of peptides with 4+ charge}}{\text{total \# of peptides with 2+ charge}}$$

use Charge found in xt

