

Identifications Folder: Raw Files Created in MS/MS NIST26 Video/Handout

James Little

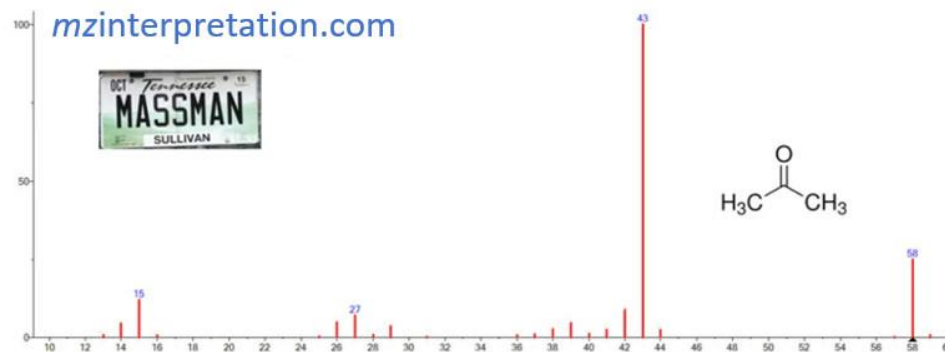
Mass Spec Interpretation Services

April 24, 2026

mzinterpretation.com

See **Full Course** on NIST26 with new **Integrated** Deconvolution/Library Searching for
EI GC-MS and **LC-MS/MS**!

Mass Spec (m/z) Interpretation Services Organic Mass Spectrometry



Identifications Folder in LC-MS/MS Chromatogram Window Processing

- When a file is processed in NIST26 Chromatogram Window
- Group of Files Created with the raw results for the MS/MS Deconvolution/Library Searching
- The information then formally presented in the Chromatogram Window
- Could they be used for PCA, good vs. bad, *etc.*?
- Used ChatGPT to Examine the Files and explain their relationship and possible utility

Local Disk (C:) > NIST26 > MSSEARCH >

Sort View ...

Name	Date modified	Type	Size
apci_msms_nist	4/22/2026 10:40 AM	File folder	
hr_msms_nist	4/22/2026 10:40 AM	File folder	
hr_msms_nist#2	4/22/2026 10:40 AM	File folder	
hr_msms_nist#3	4/22/2026 10:40 AM	File folder	
identifications	4/24/2026 8:55 AM	File folder	
Checkout_TestMix_AMSMS.log	4/24/2026 8:55 AM	Text Document	19 KB
Checkout_TestMix_AMSMS.mgf	4/24/2026 8:55 AM	MGF File	9,321 KB
Checkout_TestMix_AMSMS.run	4/24/2026 8:55 AM	RUN File	1 KB
Checkout_TestMix_AMSMS.tic	4/24/2026 8:55 AM	TIC File	64 KB
Checkout_TestMix_AMSMS.tsv	4/24/2026 8:55 AM	TSV File	2,595 KB

ChatGPT's Conclusions, See the Rest of the *Handout* for More Detailed Information

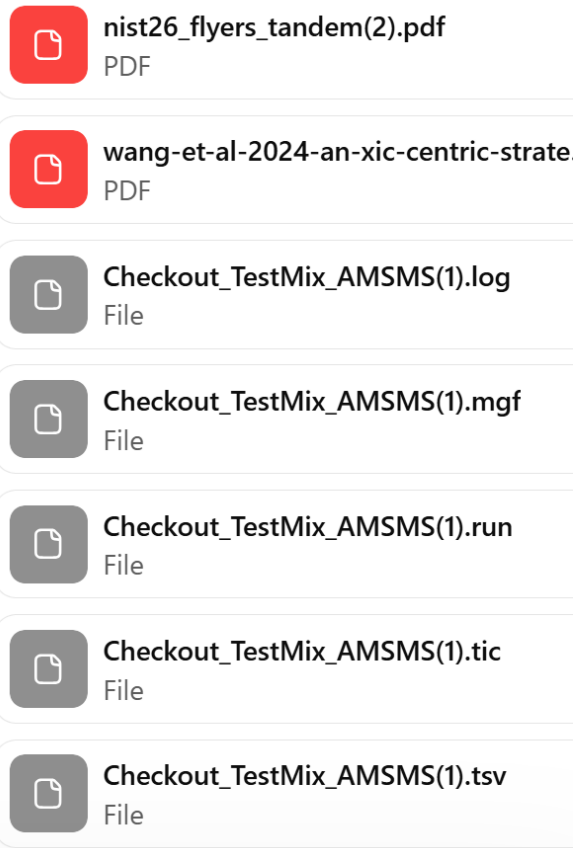
The `.tsv` output generally contains one row per detected chromatographic feature, including information such as:

- Retention time
- Precursor m/z
- Peak area or intensity
- Library match score
- Compound name
- Formula
- Adduct information (when available)

That makes it an excellent starting point for creating a feature matrix for:

- Principal Component Analysis (PCA)
- Hierarchical clustering
- Good vs. bad sample comparisons
- Outlier detection
- Marker discovery

ChatGPT Prompt



the new nist software uses xics to process the data, i have attached a flyer and research papers to show the processing in MS/MS, tandem, mode. It creates a groups of files in a folder when completed. What is all the stuff in the folder? they are all the files beginning with check

Information on MS2 (MS/MS) Spectra Processed

The program processed **3,380 total spectra** in the mzML file. It split them into 4 chunks: 1–845, 846–1690, 1691–2535, and 2536–3380. [Checkout_TestMix_AMSMS](#)

But only **216 tandem/MS2 spectra** were extracted for MS/MS library searching:

$8 + 39 + 83 + 86 = 216$ MS2 spectra [Checkout_TestMix_AMSMS](#)

So your **158 final listed MS2 searches** likely means **158 spectra survived into the final displayed/reportable results**, while **216 MS2 spectra were actually searched** from a file containing **3,380 total scans/spectra**.

James Little Note: took 11 seconds to process 158 spectra, fast!

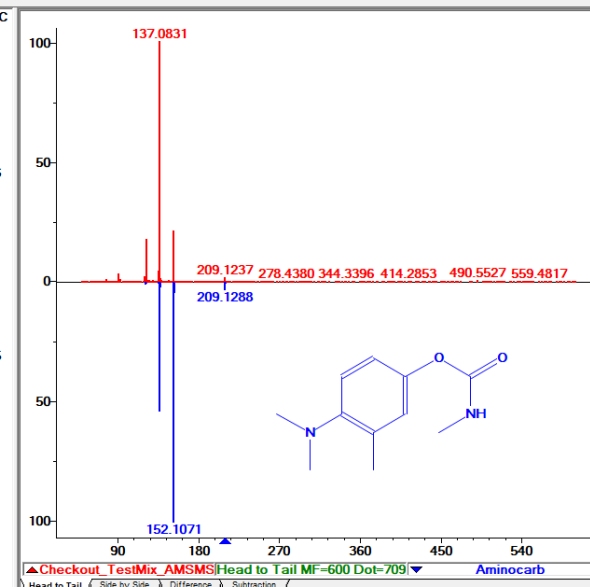
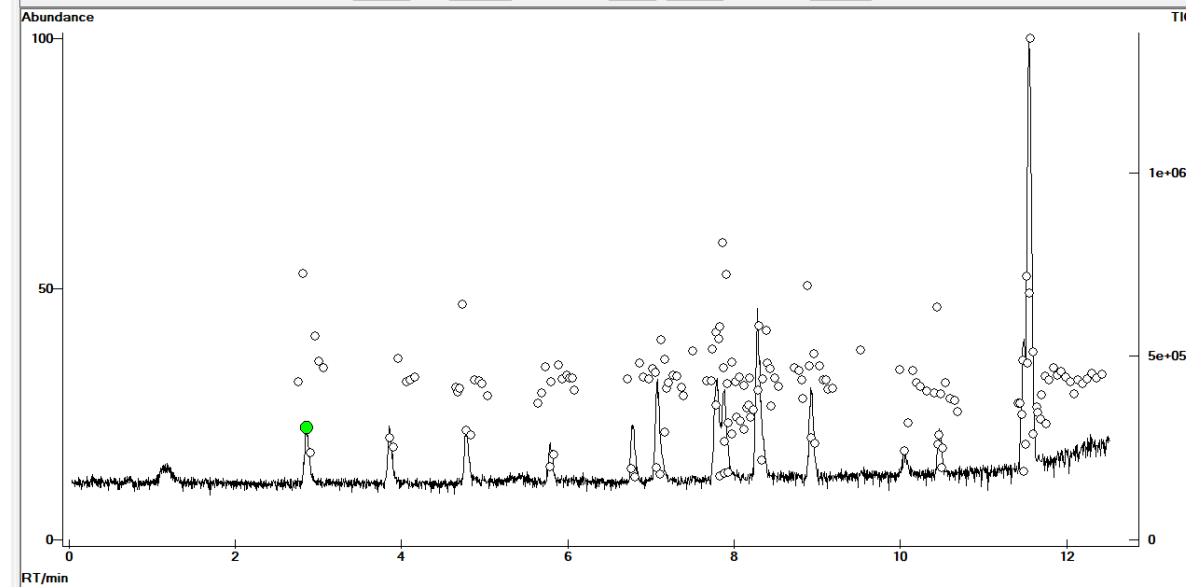
Chromatogram Results for File

PRE-RELEASE NIST MS Search 4.0 (Any mode) - [Checkout_TestMix_AMSMS, MS/MS, nHits=158]

File View Window Help



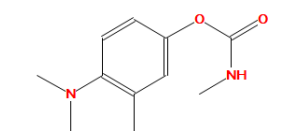
Reload on Startup
 Best Hits
 Only IDs
 Highlight Score: None
 Bkgd: <=75%
 Filter Scores: < 100
 Average:
 Ab. Threshold: Auto



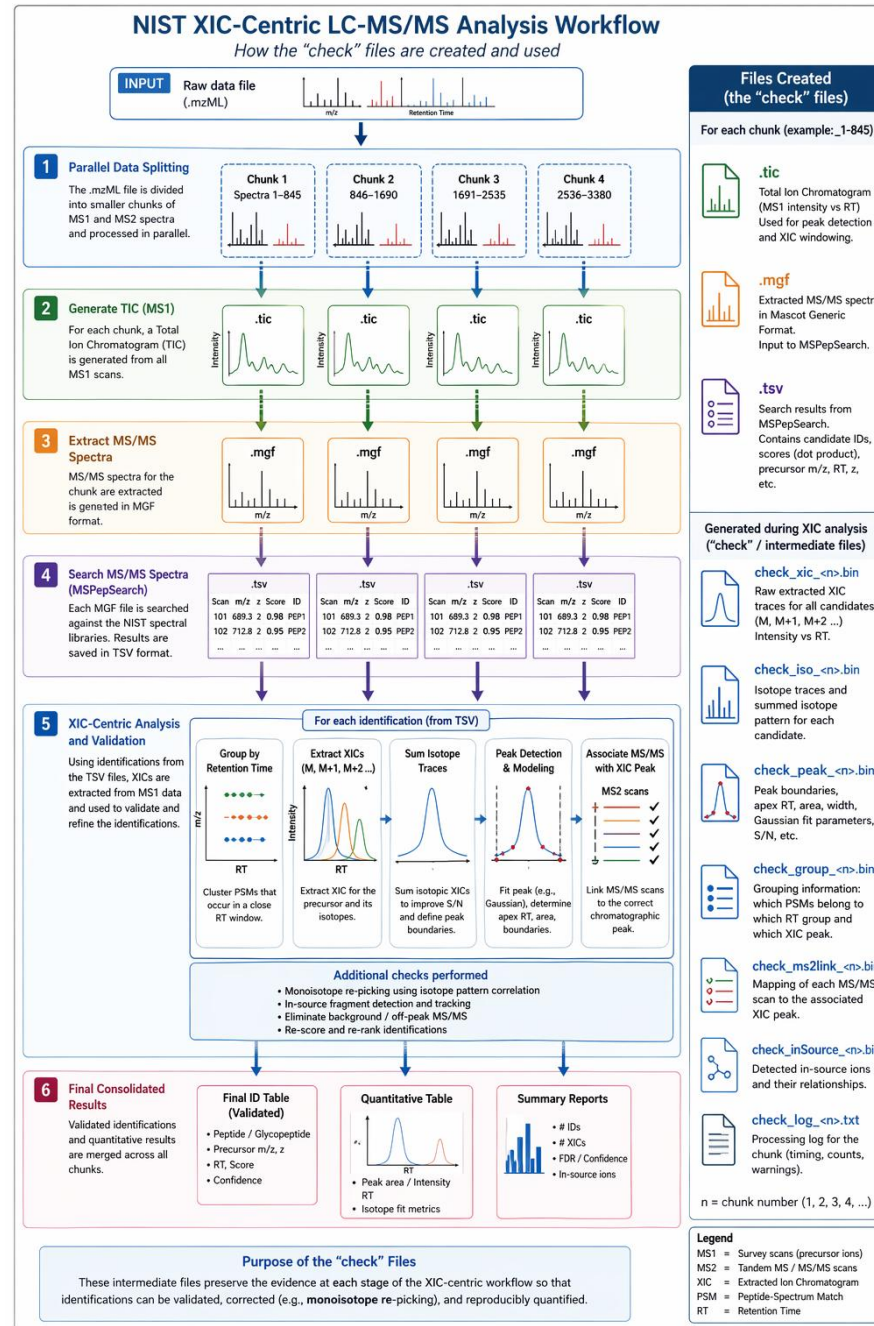
Name: Checkout_TestMix_AMSMS.834.834. File:'Checkout_TestMix_AMSMS.mzML' purity: MW: N/A ID#: 192 DB: Text File
Comment: RTINSECONDS=171.708000 Parent=209.1287 Charge=1
10 largest peaks:
 137.0831 999.0 | 152.1067 209.74 | 122.0607 175.63 | 136.0784 44.91 | 137.1871 34.1
 91.0523 32.09 | 120.0710 22.33 | 152.1802 19.59 | 209.1237 16.50 | 138.0869 13.1
Synonyms:
 no synonyms.

Name: Aminocarb
Precursor type: [M+H]⁺
Instrument type: HCD
Collision energy: NCE=20% 18eV
Precursor m/z: 209.1285
Formula: C₁₁H₁₆N₂O₂
MW: 208 **Exact Mass:** 208.121178 **CAS#:** 2032-59-9 **NIST#:** 4095041 **ID#:** 285177 **DB:** hr_m
Comment: NIST Mass Spectrometry Data Center
Ion mode: P
Instrument: Orbitrap Fusion Lumos
Ionization: ESI
Collision gas: N₂
Sample inlet: direct flow injection
Spectrum type: MS2
Notes: micromol/L in water/acetonitrile/formic acid (50/50/0.1); Spec=Consensus Nreps=
InChIKey: IMIDOCRTMDIGLJUHFFFAOYSA-N **Non-stereo**
9 largest peaks:
 152.1071 999.0 | 137.0836 532.7 | 153.1105 39.9 | 209.1288 28.8 | 138.0870 18.0 |

#	Scan	RT	Score	Abund.Rel.	Prec. m/z	Prob	dPPM	XIC Num.	nSpec	Iso. Profile	Width	Prec. Type	Lib	Dbs	Lib ID
1	808	2.7668	1	31.3	209.1257	6	13.4	9	7	1.0000	4.1	[M+H] ⁺	hr_msms_nist#2	17	Aminocarb
2	821	2.8166	389	53.0	209.1280	98	2.4	9	7	1.0000	4.1	[M+H] ⁺	hr_msms_nist#2	17	Aminocarb
3	834	2.8618	600	22.1	209.1287	100	-1.0	9	7	1.0000	4.1	[M+H] ⁺	hr_msms_nist#2	17	Aminocarb
4	848	2.9112	219	17.2	209.1289	76	-1.9	9	7	1.0000	4.1	[M+H] ⁺	hr_msms_nist#2	17	Aminocarb
5	862	2.9644	74	40.5	209.1285	9	0.0	9	7	1.0000	4.1	[M+H] ⁺	hr_msms_nist#2	17	Aminocarb
6	875	3.0141	43	35.4	209.1296	3	-5.3	9	7	1.0000	4.1	[M+H] ⁺	hr_msms_nist#2	17	Aminocarb
7	888	3.0640	40	34.1	209.1289	6	-1.9	9	7	1.0000	4.1	[M+H] ⁺	hr_msms_nist#2	17	Aminocarb
8	1117	3.8574	318	20.1	202.0433	54	-0.0	7	4	1.0000	3.9	[M+H] ⁺	hr_msms_nist#2	36	Thiabenzazole
9	1131	3.9067	150	18.2	202.0436	25	-1.5	7	4	1.0000	3.9	[M+H] ⁺	hr_msms_nist#2	36	Thiabenzazole
10	1145	3.9599	27	36.0	202.0437	5	-2.0	7	4	1.0000	3.9	[M+H] ⁺	hr_msms_nist#2	36	Thiabenzazole
11	1171	4.0595	18	31.3	202.0434	4	-0.5	7	4	1.0000	3.9	[M+H] ⁺	hr_msms_nist#2	36	Thiabenzazole
12	1184	4.1093	9	31.7	202.0428	4	2.5	8	2	0.9925	1.0	[M+H] ⁺	hr_msms_nist#2	36	Thiabenzazole
13	1198	4.1624	10	32.2	202.0432	4	0.5	8	2	0.9925	1.0	[M+H] ⁺	hr_msms_nist#2	36	Thiabenzazole
14	1338	4.6563	1	30.1	262.1241	5	-5.7	12	9	0.9983	3.5	[M+H-2H] ⁺	hr_msms_nist#2	18	Benzenepropanamide, .alpha..beta.-dihydroxy-N-methyl-N-[(1Z)-2-phenylethenyl]-
15	1344	4.6826	0	29.3	261.1762	0	18.0	18	1	0.9906	1.2	[M+H] ⁺	hr_msms_nist	28	Carisoprodol
16	1350	4.7090	8	30.1	262.1238	4	-19.8	12	9	1.0000	3.5	[M+H] ⁺	hr_msms_nist	18	Imazapyr
17	1357	4.7387	233	46.8	262.1173	83	5.0	12	9	1.0000	3.5	[M+H] ⁺	hr_msms_nist	18	Imazapyr
18	1370	4.7839	522	21.7	262.1194	99	-3.1	12	9	1.0000	3.5	[M+H] ⁺	hr_msms_nist	18	Imazapyr
19	1384	4.8338	177	20.7	262.1192	52	-2.3	12	9	1.0000	3.5	[M+H] ⁺	hr_msms_nist	18	Imazapyr
20	1398	4.8870	45	31.8	262.1187	8	-0.4	12	9	1.0000	3.5	[M+H] ⁺	hr_msms_nist	18	Imazapyr
21	1411	4.9368	1	31.6	262.1211	4	10.3	12	9	0.9957	3.5	[M+NH4] ⁺	hr_msms_nist#2	2	2-(2'-Fluoro[1,1'-biphenyl]-4-yl)propanoic acid
22	1419	4.9699	1	30.9	262.1158	6	10.7	12	9	1.0000	3.5	[M+H] ⁺	hr_msms_nist	5	Gly-Trp
23	1435	5.0360	0	28.6	262.1182	0	16.8	12	9	0.9983	3.5	[M+H] ⁺	hr_msms_nist	4	4-([1,1'-Biphenyl]-4-yloxy)aniline
24	1613	5.6508	0	27.0	230.0059	0	4.3	21	11	1.0000	3.9	[M+H] ⁺	hr_msms_nist	26	Dimethoate
25	1623	5.6905	0	29.1	230.0072	0	-1.3	21	11	1.0000	3.9	[M+H] ⁺	hr_msms_nist	26	Dimethoate




Workflow Diagram for Process Created by Chat GPT, also included in Resources Section of My Web Page



Big picture first

In the new NIST LC-MS/MS workflow, everything is built around **XIC-centric processing**:

- Extract XICs from MS1
- Group spectra by RT
- Associate MS/MS with XIC peaks
- Validate/adjust identifications using XIC information  wang-et-al-2024-an-xic-centric-...

To make that work efficiently, the software creates a **whole set of intermediate (“check”) files** — essentially a working scratchpad of the analysis.

What those “check” files actually are

Even though your folder shows them as “check...”, they are really **pipeline stage outputs**. Based on your log:

 Checkout_TestMix_AMSMS

1. Raw data split into chunks

Task #1: spectra 1-845
Task #2: spectra 846-1690
...



 NIST divides your mzML into chunks for **parallel processing**

Each chunk gets its own set of files.

2. `.tic` files → Total Ion Chromatograms

Saving ..._1-845.tic

What it is:

- MS1 intensity vs retention time
- Used for:
 - Peak detection
 - RT alignment
 - XIC windowing

👉 Think of this as the **global chromatogram backbone**

3. `.mgf` files → MS/MS spectra extraction

Convert to mgf
Saving ..._1-845.mgf

What it is:

- Peak-picked MS/MS spectra
- Standard Mascot Generic Format

Used for:

- Library searching (MSPepSearch)

👉 This is the **search input layer**

4. `.tsv` files → search results

```
OUTTAB ..._1-845.tsv
```

What it is:

- Tabular results from MSPepSearch
- Contains:
 - Candidate IDs
 - Scores (dot product, etc.)
 - precursor m/z, RT, charge

👉 This is the **identification layer**

5. XIC grouping + analysis (your key “check” stage)

```
Grouping...  
# groups: 63  
XICing...  
# XICs: 311
```

This is where the **XIC-centric magic** happens:

From the paper/workflow:

- PSMs grouped by RT
- XICs extracted for each candidate
- Isotope traces summed
- Peaks detected
- MS/MS linked back to XIC peak [wang-et-al-2024-an-xic-centric-...](#)

👉 These steps generate internal **“check” data structures** like:

- XIC traces per precursor
- isotope envelopes
- peak boundaries
- RT associations

These often get written out as temporary/check files so:

- the GUI can reload them instantly
- you can inspect intermediate results
- reprocessing is faster

6. Additional internal “check” content (not always obvious)

From your log + workflow, these files likely contain:

✓ XIC traces

- intensity vs RT for each m/z

✓ Isotope traces

- multiple XICs (M, M+1, M+2...)

✓ Peak models

- Gaussian fitting (you saw this in log)

✓ In-source ion tracking

Determining in-source ions ...



✓ Grouping maps

- which MS/MS spectra belong to which chromatographic peak

What “check” really means conceptually

These files are essentially:

👉 “Verification layers” between MS/MS search and final ID

Because NIST is doing something new:

Traditional workflow:

MS/MS → ID → done

NIST XIC workflow:

MS/MS → ID → XIC → validate → refine → re-rank

So the “check” files are:

- ✓ intermediate XIC-derived evidence
- ✓ used to confirm or reject IDs
- ✓ used to improve monoisotopic assignment
- ✓ used to remove background MS2




💡 Why so many files?

Because NIST is doing:

- Parallel processing (multiple chunks)
- Multi-level analysis:
 - MS1 (XIC)
 - MS2 (search)
 - Cross-validation

From flyer:

- "XIC based analysis"
- "MS1 & MS2 analysis"
- "Eliminate background MS2"  nist26_flyers_tandem

👉 Each level produces its own data products.


The most important takeaway (for your understanding)

Those “check” files are not junk — they are:

▶ The core evidence layer of the new NIST approach

They contain:

- Chromatographic truth (XIC)
- Isotope behavior
- RT consistency
- In-source relationships

 This is exactly what allows NIST to:

- fix wrong precursor assignments
- identify in-source fragments
- improve MS/MS confidence

