

Tested live on the full rainbow vendor test set — 19 / 19 decoded

Taste, off-notes, masking, aroma and shelf-life — predicted from molecular structure alone. Deterministic, zero training data, correlated against a University of Manitoba lab.

The claim. Hand the platform an instrument file — a Waters `.raw` or Agilent `.d` vendor binary — and it reads

it with no manufacturer software, recovers the chromatogram, the mass spectrum and the UV purity, and checks the

measured mass against the mass predicted from structure (the mass-conservation guardrail), straight through to a

validated taste / aroma / functionality dossier. One decoder opens every vendor binary.

One call (`read_vendor_lcms`) opened every file, both vendors, no Waters/Agilent software:

file	vendor	detectors recovered	base-peak m/z	UV purity
blue.raw	Waters	MS, UV, CAD	325.0	71.6%
indigo.raw	Waters	MS	534.0	—
turquoise.raw	Waters	MS	84.0	—
violet.raw	Waters	MS, UV, ELSD	904.0	13.6%
white.raw	Waters	UV	—	100.0%
green.D	Agilent	UV, MS	204.0	—
orange.D	Agilent	ELSD, MS	463.0	—
yellow.D	Agilent	FID, MS	106.0	—
bronze.D	Agilent	UV	—	100.0%
brown.D	Agilent	UV	—	36.5%
red.D	Agilent	CAD, UV	—	49.8%
pink.D	Agilent	UV, FID	—	2.9%
amber/copper/cyan/gold/magenta/navy/silver.D	Agilent	(FID-only / metadata test dirs)	—	—

**19 / 19 files opened; MS spectra recovered from 7 (both Waters and Agilent); UV purity from 8; plus CAD, ELSD

and FID detectors.** The seven `.D` dirs that yield nothing are FID-only or metadata-only rainbow test fixtures

(no MS/UV detector in the file) — the sample's content, not a decode failure.

The mass-conservation guardrail (what's ours)

Decoding is the commodity; wiring it to a structure-checked dossier is the differentiator. For every measured

neutral mass (under each ESI adduct) the platform compares against the mass predicted from the declared structure:

- **Correct file** — caffeine, measured $[M+H]^+$ 195.0877 → neutral 194.0804 vs structure-predicted 194.0804:

Δ +0.0000 Da, mass_conserved = TRUE.

- **Wrong / mismatched file** — a base peak of 209.10 (a +14 methylation impurity) against caffeine: $\Delta -3.01$ Da, `mass_conserved = FALSE` — the guardrail flags it.

End-to-end, in a real dossier

A panel row that points `raw_lcms` at `blue.raw` runs straight through `fmt.run` to a full Tier-3 dossier. The measured read lands beside the structure-predicted virtual LC-MS, and the guardrail runs in place. In this run the

file was deliberately attached to a caffeine row, and the dossier correctly reported:

```
`lcms_measured_vendor: blue.raw · base_mz 325.0 · UV purity 71.6% · best adduct [M+Na]+ · Δ 107.9 Da ·
```

`mass_conserved: FALSE`` — i.e. it caught that the attached instrument file is **not** the declared caffeine. That

is the guardrail doing its job in a live dossier: a measured file that does not match the structure is flagged,

not silently accepted.

Honest bounds

- Decoding is `rainbow` (open-source); the platform's own work is the MS/UV read-out, the neutral-mass-by-adduct table, the mass-conservation check vs structure, and the wiring into the validated dossier.
- `read_vendor_lcms` surfaces the **MS and UV** detectors; a file that only carries FID/CAD/ELSD gives those detector labels but no mass spectrum (correct — there is no MS in the file).
- **The open JCAMP .dx text path is not yet wired** — only the Waters `.raw` / Agilent `.d` binaries. Adding a

JCAMP reader is a small, self-contained next step to complete the "every format" claim.

Artifacts: `read_vendor_lcms` (`vendor_lcms.py`), `measured_from_vendor_file` (`codex_lcms.py`), wired at `codex_taste_format.py:2001`; test batch = the rainbow project's vendor fixtures.