UNCERTFP: UNCERTAINTY-AWARE MULTI-VIEW EVIDENCE FUSION FOR MOLECULAR FINGERPRINT PREDICTION FROM TANDEM MASS SPECTRA

Anonymous authors
Paper under double-blind review

ABSTRACT

Molecular fingerprint prediction from tandem mass spectrometry (MS/MS) is hindered by incomplete fragmentation and by the fact that distinct molecular structures can produce similar spectra. Existing methods rely mainly on a single view and thus ignore complementary compositional cues. Moreover, similar bit-presence probabilities can reflect different evidence strengths, obscuring whether a prediction has weak or substantial support. Multiple views provide additional structural information, but shared observations can inflate accumulated evidence and conflicting views can undermine prediction. We propose UncertFP, an uncertainty-aware multi-view evidence fusion framework that infers per-bit positive and negative evidence from fragment-peak, neutral-loss, and peak-difference views. Specifically, UncertFP adjusts the contribution of each view through bit reliability and uses a learnable fusion coefficient to control evidence accumulation by interpolating between direct accumulation and weighted averaging. Disagreement discounting further reduces fused evidence when strongly supported views favor opposing bit states. Subsequently, the fused evidence and the shared prior jointly yield a probabilistic fingerprint and an uncertainty mass that quantifies evidence insufficiency. We leverage the relative ranking of whole-fingerprint prediction errors to supervise spectrum-level uncertainty, thereby enabling selective prediction. Across three datasets, two encoders, and three fingerprint dimensionalities, UncertFP improves Soft Tanimoto by 0.07-0.16 over baselines. The complete prediction and scoring pipeline also improves candidate-structure ranking, while uncertainty-based spectrum selection prioritizes predictions with lower average error.

1 INTRODUCTION

Tandem mass spectrometry (MS/MS) is widely used to annotate small-molecule structures in untargeted metabolomics through fragmentation-pattern analysis (Alseekh et al., 2021). Extended-connectivity fingerprints (ECFPs), which encode local atomic environments as binary bits, serve as a common structural representation for spectrum-based prediction and candidate-structure retrieval (Rogers & Hahn, 2010). In a typical spectrum-to-fingerprint workflow, a model first predicts a molecular fingerprint from an observed spectrum and then ranks candidate structures by comparing the predicted fingerprint with those computed from candidate molecules (Dührkop et al., 2015; Goldman et al., 2023). Here, we focus on predicting molecular fingerprints given an MS/MS spectrum, the precursor molecular formula, and the adduct type. This task is inherently ill-posed because a limited set of fragmentation observations can be compatible with multiple molecular structures, while acquisition conditions and adduct forms can substantially alter the spectrum of the same molecule; consequently, many fingerprint bits cannot be uniquely determined from a single spectrum (Heinonen et al., 2012; Stravs et al., 2022; Hoang et al., 2024; Liu et al., 2025).

Kernel methods, formula-aware encoders, and spectral pretraining have advanced this task by exploiting spectral and fragmentation-tree similarities, incorporating neural representations into kernel-based predictors, extracting structural cues from compositional information, and learning transferable representations from large unlabeled spectral collections (Dührkop et al., 2015; Dührkop, 2022; Goldman et al., 2023; Bushuiev et al., 2026). However, these advances do not resolve a fundamental limitation of bit-presence probabilities: the same probability can reflect markedly different amounts of sup-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

porting evidence, making probability alone insufficient to distinguish a prediction based on scarce information from one supported by substantial observations; see Appendix B.2 for detailed proof.

In fact, the same spectrum can derive views that capture more clues and supporting evidence, yet this is often overlooked by existing methods. For example, fragment peaks represent the compositions of observed fragments and provide direct evidence of retained substructures; neutral losses capture the compositional differences between the precursor and its fragments, indicating what is lost during fragmentation, and they complement the structural similarity information in fragment spectra; inter-fragment differences describe compositional changes that point to related local substructures and provide clues to molecular core structures (Aisporna et al., 2022; Xing et al., 2020). Although these views expose distinct structural signals, they are derived from the same spectrum, share observations, and may provide unequal or conflicting support for the state of an individual fingerprint bit.

As in Figure 1, for the same fingerprint bit, one view supplies weak evidence for one state, while the other two supply strong evidence for opposing states. In such case, equal-weight probability averaging gives weakly supported predictions the same influence as strongly supported ones. Conversely, direct evidence accumulation can count shared information multiple times and inflate the apparent support for the fused prediction without adequately accounting for disagreement among views. Multi-view fusion for molecular fingerprint prediction therefore faces two challenges: how to allocate view contributions while controlling the scale of evidence accumulation under shared observations; and how to retain less evidence when strongly supported views favor opposing states, while falling back to a shared prior when evidence is scarce across all views (Xu et al., 2024; Bezirganyan et al., 2025).

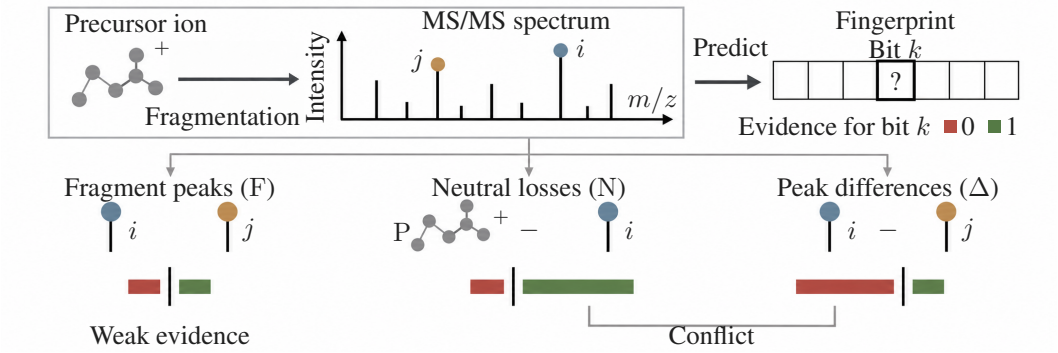

Figure 1: Illustrative example of sparse and conflicting evidence for one fingerprint bit across three views of an MS/MS spectrum. In this example, the fragment-peak (F) view provides weak support, whereas the neutral-loss (N) and peak-difference (Δ) views provide stronger support for opposite states. Bar lengths indicate evidence strength; red and green denote support for bit absence (0) and presence (1), respectively.

To address these challenges, we propose UncertFP, an uncertainty-aware multi-view evidence fusion framework for molecular fingerprint prediction. Specifically, UncertFP learns positive and negative evidence for each fingerprint bit from the fragment-peak, neutral-loss, and peak-difference views, thereby integrating structural information from different views. Since view reliability varies across fingerprint bits, UncertFP introduces bit-wise weights to adjust the contribution of each view and employs learnable interpolation between direct accumulation and weighted averaging to control the scale of fused evidence under shared observations. When views with strong evidence support opposite bit states, disagreement discounting further reduces the evidence retained after aggregation. Combining the fused evidence with a shared prior gives per-bit Beta distributions whose means form the probabilistic fingerprint and whose uncertainty masses quantify evidence insufficiency. The training process jointly optimizes bit-wise classification and whole-fingerprint similarity. Furthermore, the ranking of whole-fingerprint errors provides supervision for spectrum-level uncertainty mass, enabling its use in selecting spectra with lower prediction error. The main contributions are summarized as follows:

1. We propose UncertFP, an uncertainty-aware multi-view evidence fusion framework for molecular fingerprint prediction that integrates bitwise positive and negative evidence from three complementary views with a shared prior to generate a probabilistic fingerprint and uncertainty mass.

2. We develop an evidence fusion mechanism that incorporates bitwise reliability weighting, a learnable fusion coefficient, and disagreement discounting, and provide theoretical analyses of the monotonicity of disagreement discounting and shrinkage toward the prior under low evidence.
3. We jointly optimize bitwise classification, whole-fingerprint similarity, and error-ranking to learn probabilistic fingerprints while encouraging spectrum-level uncertainty mass to reflect whole-fingerprint error ranking.

2 Related work

Spectrum-to-fingerprint prediction. FingerID and CSI:FingerID map spectra to molecular fingerprints through bitwise classification and probability estimation. Subsequent deep kernel and neural models improve the representations used for this task (Heinonen et al., 2012; Dührkop et al., 2015; Dührkop, 2022; Baygi & Barupal, 2024). Metabolite Inference with Spectrum Transformers (MIST) improves fingerprint prediction through formula-aware spectral encoding (Goldman et al., 2023). MultiView Projection (MVP) jointly aligns molecular graphs, fingerprints, individual spectra, and consensus spectra through contrastive learning, then ranks candidates in the shared embedding space (Zhou Chen & Hassoun, 2026). These fingerprint prediction methods focus on improving the mapping from spectral representations to fingerprints, whereas MVP aligns different representations for candidate retrieval. UncertFP constructs fragment-peak, neutral-loss, and peak-difference views from the same spectrum. It estimates each view's positive and negative evidence for every fingerprint bit and incorporates evidence strength into fusion. Appendix A provides further comparisons of spectral representations and downstream applications.

Evidential prediction and risk estimation. Evidential Deep Learning (EDL) uses nonnegative evidence to construct a Dirichlet distribution over class probabilities (Sensoy et al., 2018). Multi-label evidential learning extends this formulation with per-label Beta distributions for co-occurring classes (Zhao et al., 2023). The prior weight controls the relative influence of evidence and the prior, allowing the model to represent evidence strength alongside the predictive mean (Chen et al., 2024). The identifiability and risk interpretation of uncertainty measures from second-order distributions depend on the training objective and evaluation protocol (Bengs et al., 2022; Juergens et al., 2024; Shen et al., 2024). Failure prediction and selective prediction provide ways to use confidence for assessing prediction quality and selecting samples (Corbière et al., 2019; Geifman & El-Yaniv, 2017). In contrast to failure prediction methods that learn true class probabilities, UncertFP supervises a spectrum-level metric using the relative ordering of whole-fingerprint prediction errors. This metric aggregates per-bit uncertainty mass, linking uncertainty mass to whole-fingerprint prediction quality for spectrum selection.

Multi-view evidence fusion. Trusted multi-view classification combines predictions and reliability across views by aggregating evidential opinions (Han et al., 2023; Liu et al., 2022; 2023). Evidential Conflictive Multi-view Learning (ECML) aggregates opinions through an operation equivalent to evidence averaging to handle differences in view reliability (Xu et al., 2024). Discounted Belief Fusion (DBF) measures conflict using probability distances and joint certainty, discounts individual opinions, and then applies generalized averaging (Bezirganyan et al., 2025). These methods account for conflict, but compositional views of the same spectrum also share observations. This requires assigning view contributions per binary fingerprint bit and controlling the accumulation of shared information. In contrast to ECML's evidence averaging, UncertFP assigns view contributions through bitwise reliability weights. A global learnable fusion coefficient controls the combination of direct accumulation and weighted averaging. Whereas DBF discounts individual opinions before fusion, UncertFP discounts aggregated positive and negative evidence according to disagreement at each fingerprint bit. The shared prior is then added once to the discounted evidence.

3 Problem formulation

Given an MS/MS spectrum and its precursor information, we aim to predict the molecular fingerprint and quantify evidence insufficiency for each bit. The input spectrum $S = \{(m_i, I_i)\}_{i=1}^{L}$ contains L peaks, where m_i and I_i denote the mass-to-charge ratio and relative intensity of peak i. The precursor information is $\mathcal{M} = (\mathbf{c}_{\text{prec}}, a_{\text{adduct}})$, where $\mathbf{c}_{\text{prec}} \in \mathbb{N}_0^D$ records atom counts in a fixed order of

D elements and a_{adduct} denotes the adduct type. For spectrum n, the prediction target is a K-bit molecular fingerprint $\mathbf{y}_n \in \{0,1\}^K$, with $y_{n,k} = 1$ indicating an active bit.

Similar bit-presence probabilities can reflect different evidence strengths. The model therefore outputs a probabilistic fingerprint $\boldsymbol{\mu}_n \in [0,1]^K$ together with per-bit uncertainty masses $\mathbf{u}_n \in [0,1]^K$ that quantify evidence insufficiency. These outputs support spectrum selection and candidate-structure ranking. Bold symbols denote vectors; we omit the sample index n when considering a single spectrum.

4 Uncertainty-aware multi-view evidence fusion and fingerprint prediction

Complementary compositional views provide evidence for fingerprint prediction, but shared observations and conflicting predictions require control over evidence accumulation and retention. UncertFP therefore estimates positive and negative evidence from each view, combines them through bitwise reliability weights and a learnable fusion coefficient, and discounts disagreement (Figure 2). The retained evidence and a shared prior define per-bit Beta distributions through a conjugate Beta–Bernoulli update parameterization (Sensoy et al., 2018; Chen et al., 2024). Their means form the probabilistic fingerprint, while uncertainty mass quantifies evidence insufficiency. To use this uncertainty for spectrum selection, error-ranking supervision aligns its spectrum-level ranking with whole-fingerprint prediction errors.

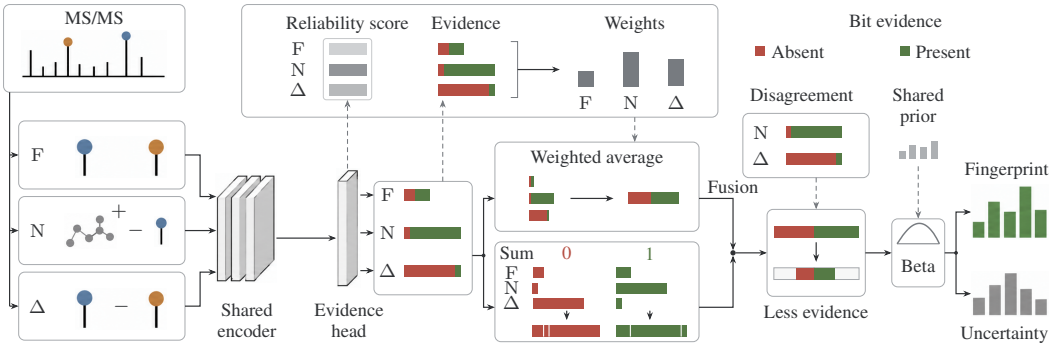

Figure 2: UncertFP evidence fusion and fingerprint prediction. The fragment-peak (F), neutral-loss (N), and peak-difference (Δ) views share an encoder to estimate per-bit positive and negative evidence. Bitwise reliability weighting and a learnable coefficient η interpolate between direct accumulation and reliability-weighted averaging; disagreement discounting by δ_k reduces retained evidence. The retained evidence and a shared prior define per-bit Beta distributions whose means $\boldsymbol{\mu}$ and uncertainty masses $\mathbf{u}$ form the outputs. Pale outlines mark pre-discount evidence; bars and shades are illustrative.

4.1 Spectral views and view-specific evidence

Complementary compositional views capture observed fragments and their relationships to the precursor and to one another. Neutral losses and pairwise fragment mass differences support molecular similarity analysis (Aisporna et al., 2022; Xing et al., 2020). Using MIST preprocessing and subformula representations (Goldman et al., 2023), we construct three views. For annotated peaks, the fragment-peak view uses formula vectors $\mathbf{c}_i$ and transformed intensities $\widetilde{I}_i$. The neutral-loss and peak-difference views use $\mathbf{c}_{\text{prec}} - \mathbf{c}_i$ and $\mathbf{c}_i - \mathbf{c}_j$, respectively, with peak differences directed from higher to lower mass. We retain componentwise nonnegative composition differences and append a token encoding the precursor formula and adduct type to each sequence (Appendix C.2).

We denote the fragment-peak, neutral-loss, and peak-difference views by F, N, and Δ, respectively. For each view $v \in \mathcal{V} = \{\text{F}, \text{N}, \Delta\}$, a shared encoder produces a representation $\mathbf{h}_v$ and bit-presence probabilities $\boldsymbol{\pi}_v$. To distinguish evidence strength from bit-presence probability, an evidence head estimates how much support each view provides and how it should contribute to fusion. It projects $\mathbf{h}_v$ to 256 dimensions, followed by layer normalization and a sigmoid linear unit (SiLU). Two linear

maps produce evidence preactivations $\boldsymbol{\ell}_v \in \mathbb{R}^K$ and reliability scores $\mathbf{q}_v \in \mathbb{R}^K$. A capped softplus converts $\ell_{v,k}$ into positive, bounded evidence strength $\kappa_{v,k}$. We then use $\pi_{v,k}$ to divide this evidence into support for bit presence and absence, as defined below. Appendix C.3 specifies the transformation and its bounds.

Definition 1 (View-specific evidence mapping)**.** *For each view v and fingerprint bit k, given $\pi_{v,k} \in (0,1)$, $\kappa_{v,k} > 0$, prior bit-presence probability $a_k \in (0,1)$, and prior strength $W > 0$, define $r_{v,k} = \kappa_{v,k}\pi_{v,k}$ and $s_{v,k} = \kappa_{v,k}(1-\pi_{v,k})$ as positive and negative evidence. The view-specific uncertainty mass and predictive mean are*

$$u_{v,k} = \frac{W}{\kappa_{v,k} + W}, \qquad \mu_{v,k} = (1 - u_{v,k})\pi_{v,k} + u_{v,k}a_k. \tag{1}$$

Less evidence raises $u_{v,k}$ and pulls $\mu_{v,k}$ toward a_k. These quantities feed view weighting and evidence fusion in Eqs. 2–3; after fusion, U_n supports spectrum selection and $\mathbf{1} - \mathbf{u}$ supports candidate ranking.

4.2 Conflict-aware fusion and per-bit prediction

Because the views share spectral observations, direct summation can lead to excessive evidence accumulation. Their reliability also varies across fingerprint bits, so fusion must control both view contributions and the amount of retained evidence. Subjective logic uses cumulative fusion for independent evidence and averaging fusion for dependent evidence (Jøsang et al., 2010; van der Heijden et al., 2018). Accordingly, UncertFP interpolates between direct accumulation and bitwise reliability-weighted averaging using a global learnable scalar $\eta \in [0,1]$. For each fingerprint bit k, the reliability scores $q_{v,k}$ and the available-evidence fractions $1 - u_{v,k}$ jointly determine the view weights:

$$\lambda_{v,k} = \mathrm{softmax}_{v\in\mathcal{V}}[q_{v,k} + \gamma \log \mathrm{sg}(\max\{1 - u_{v,k}, \varepsilon_u\})], \tag{2}$$

where softmax normalizes across views, $\gamma \geq 0$ controls the influence of the evidence fraction, and $\varepsilon_u > 0$ lower-bounds the logarithm's argument. The stop-gradient operator $\mathrm{sg}(\cdot)$ preserves input values while blocking gradients through this term. Thus, for each bit k, the aggregated positive and negative evidence components are given by the following expressions, with $\sum_{v\in\mathcal{V}} \lambda_{v,k} = 1$:

$$\begin{aligned} r_k^{(0)} &= (1-\eta)\sum_{v\in\mathcal{V}} r_{v,k} + \eta \sum_{v\in\mathcal{V}} \lambda_{v,k} r_{v,k}, \\ s_k^{(0)} &= (1-\eta)\sum_{v\in\mathcal{V}} s_{v,k} + \eta \sum_{v\in\mathcal{V}} \lambda_{v,k} s_{v,k}. \end{aligned} \tag{3}$$

where $\lambda_{v,k}$ controls contributions within the weighted term, while adjustment coefficient η controls the total evidence and the influence of learned weights. Appendix C.3 gives their parameterizations and settings.

Even after controlling accumulation, strong views may disagree on bit-presence probabilities. To determine how much of their combined evidence to retain, we quantify per-bit disagreement using probability distance and joint certainty (Xu et al., 2024; Bezirganyan et al., 2025):

$$\chi_k = \frac{2}{|\mathcal{V}|(|\mathcal{V}|-1)} \sum_{v<v'} |\mu_{v,k} - \mu_{v',k}|\,(1 - u_{v,k})(1 - u_{v',k}), \tag{4}$$

where $v < v'$ counts each unordered pair once, and the joint evidence fraction reduces contributions from weak-evidence views. Then, we retain evidence using $\delta_k = \exp(-\tau\chi_k)$, where strength $\tau > 0$ is learnable.

$$r_k = \delta_k r_k^{(0)}, \quad s_k = \delta_k s_k^{(0)}. \tag{5}$$

With τ and undiscounted evidence fixed, greater disagreement reduces total evidence while preserving its positive-to-negative ratio. For $|\mathcal{V}| = 1$, we define $\chi_k = 0$ because no pair of views is available.

To provide a reference when retained evidence is weak, we use the clipped training frequency a_k and fixed strength W to define a shared prior $\mathrm{Beta}(Wa_k, W(1-a_k))$. Since all views share this reference, its contributions are added only once after fusion and discounting. For a query $(S, \mathcal{M})$, let ρ_k denote the random bit-presence probability, while a_k remains its fixed prior mean. The retained evidence determines the Beta shape parameters $\alpha_k = r_k + Wa_k$ and $\beta_k = s_k + W(1-a_k)$:

$$\rho_k \mid S, \mathcal{M} \sim \mathrm{Beta}(\alpha_k, \beta_k), \qquad \mu_k = \frac{\alpha_k}{\alpha_k + \beta_k}, \qquad u_k = \frac{W}{r_k + s_k + W}. \tag{6}$$

The means form the probabilistic fingerprint $\boldsymbol{\mu} = (\mu_k)_{k=1}^K$, while $\mathbf{u} = (u_k)_{k=1}^K$ quantifies evidence insufficiency for its bits.

To characterize how discounting changes prediction, consider one bit and write $E = r + s > 0$ and $p = r/E$, omitting k. Equation 6 gives $u = W/(E + W)$ and $\mu = (1 - u)p + ua$, hence $|\mu - a| = (1 - u)|p - a|$. Discounting scales both evidence components equally, changing E while preserving p.

Proposition 1 (Monotonicity of disagreement discounting)**.** *Fix the prior parameters, $\tau > 0$, and the undiscounted evidence with total $E_0 > 0$. As χ increases, $E = e^{-\tau\chi}E_0$ strictly decreases, u strictly increases, and $|\mu - a|$ is nonincreasing.*

Proposition 2 (Shrinkage toward the prior under low evidence)**.** *Fix $p \in [0, 1]$, $a \in (0, 1)$, and $W > 0$. As $E \to 0^+$, $u \to 1$ and $\mu \to a$. If $p = a$, then $\mu = a$ for every $E > 0$, although u varies with E.*

Proofs are provided in Appendices B.1 and B.2. Together, the propositions justify the downstream use of uncertainty: conflict increases u and shrinks predictions toward the prior, whereas weak evidence yields prior fallback. These properties motivate Eq. 5, U_n-based spectrum retention, and $\mathbf{1} - \mathbf{u}$-weighted candidate ranking, evaluated in Secs.5.1, 5.2, and 5.3.

4.3 Training objective and candidate-structure ranking

We jointly optimize bitwise prediction, whole-fingerprint overlap, and spectrum-level uncertainty ranking with class-weighted BCE, Soft Tanimoto, and ranking losses.

Soft Tanimoto measures active-bit overlap without rewarding joint absence; its weighted form also supports evidential candidate scoring. For any soft fingerprint $\mathbf{z} \in [0, 1]^K$, binary fingerprint $\mathbf{b} \in \{0, 1\}^K$, and nonnegative bit weights $\mathbf{w} \in \mathbb{R}^K_{\geq 0}$, the training errors and candidate scores share a weighted Tanimoto function, defined as

$$\mathrm{Tan}(\mathbf{z}, \mathbf{b}; \mathbf{w}) = \frac{\mathbf{w}^\mathsf{T}(\mathbf{z} \odot \mathbf{b})}{\max\{\mathbf{w}^\mathsf{T}(\mathbf{z} + \mathbf{b} - \mathbf{z} \odot \mathbf{b}), \varepsilon_T\}}, \tag{7}$$

where $\odot$ denotes elementwise multiplication, T denotes transposition, and $\varepsilon_T > 0$ lower-bounds the denominator. For spectrum selection, Eq. 8 summarizes each spectrum by whole-fingerprint error e_n and mean uncertainty mass U_n:

$$e_n = 1 - \mathrm{Tan}(\boldsymbol{\mu}_n, \mathbf{y}_n; \mathbf{1}), \qquad U_n = \frac{1}{K}\sum_{k=1}^K u_{n,k}, \tag{8}$$

where $\mathbf{1}$ is the K-dimensional all-ones vector. Within minibatch $\mathcal{B}$, the Soft Tanimoto loss $\mathcal{L}_{\mathrm{sim}} = \frac{1}{|\mathcal{B}|}\sum_{n\in\mathcal{B}} e_n$ measures mismatch, while pairs in $\mathcal{P}_{\mathrm{rank}} = \{(n, n') \in \mathcal{B}^2 : e_n > e_{n'} + \Delta_e\}$ train U_n to follow the error ranking; the ranking loss is

$$\mathcal{L}_{\mathrm{rank}} = \frac{1}{|\mathcal{P}_{\mathrm{rank}}|}\sum_{(n,n')\in\mathcal{P}_{\mathrm{rank}}} \max\{0, m_{\mathrm{rank}} - U_n + U_{n'}\}, \tag{9}$$

where $m_{\mathrm{rank}} > 0$ is the target margin; the loss is zero when the pair set is empty.

Ultimately, the overall objective of UncertFP is

$$\mathcal{L} = \lambda_{\mathrm{BCE}}\mathcal{L}_{\mathrm{BCE}} + \lambda_{\mathrm{sim}}\mathcal{L}_{\mathrm{sim}} + \lambda_{\mathrm{rank}}\mathcal{L}_{\mathrm{rank}}. \tag{10}$$

where class-weighted BCE λ_{BCE} supervises bits, while Soft Tanimoto and ranking losses supervise fingerprint overlap and uncertainty ranking; Appendix C.5 gives formula details and settings.

At inference, lower U_n prioritizes spectra. Candidate ranking uses $T_u(x) = \mathrm{Tan}(\boldsymbol{\mu}, \mathbf{g}(x); \mathbf{1} - \mathbf{u})$ and $R(x) = T_u(x) - \lambda_V\widehat{\sigma}_B(x)$; Appendices B.3 and B.4 give sampling and variance details.

5 EXPERIMENTS

We evaluate UncertFP on three datasets: CANOPUS (Goldman et al., 2023), MassSpecGym (Bushuiev et al., 2024), and SpectraVerse-positive (Gupta et al., 2026), against matched MIST (Goldman et al., 2023) and DiffMS (Bohde et al., 2025) baselines under each dataset, encoder, and fingerprint-dimensionality setting. Soft and Binary Tanimoto measure probabilistic and thresholded fingerprint overlap with the ground truth, respectively; candidate ranking uses $\mathrm{Acc@}k$ and MRR, and selective risk is mean error $1 - \mathrm{Soft}$ at matched coverage. Dataset and protocol details are provided in Appendices C.1 and C.6.

5.1 MOLECULAR FINGERPRINT PREDICTION AND CANDIDATE-STRUCTURE RANKING

Fingerprint prediction. Across two spectral encoders and three fingerprint dimensionalities, three datasets with 1024-, 2048-, and 4096-bit fingerprints yield 18 paired comparisons using weights trained in our experiments. We additionally evaluate four pairs initialized from released 4096-bit encoder weights on CANOPUS and MassSpecGym; within each pair, UncertFP starts from the corresponding baseline weights. As shown in Table 1, UncertFP consistently improves both fingerprint metrics across all trained-weight pairs, with absolute gains of 0.07–0.16 in Soft Tanimoto and 0.06–0.19 in Binary Tanimoto. For example, on CANOPUS, UncertFP improves DiffMS-4096 from 0.31 to 0.47 in Soft Tanimoto and from 0.40 to 0.51 in Binary Tanimoto. Both metrics also improve in all four released-weight comparisons, so the gains are consistent across datasets, encoders, dimensionalities, and initialization sources.

Table 1: Molecular fingerprint prediction on three datasets. Rows marked "released" use encoder weights released by the original authors; all other rows use weights trained in our experiments. UncertFP initializes from the corresponding baseline weights. Slashes indicate unavailable released encoder weights. **Bold** values mark the higher score in each pair; ↑ indicates that higher is better.

	CANOPUS		**MassSpecGym**		**SpectraVerse-positive**	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-1024	0.28	0.35	0.21	0.26	0.24	0.19
UncertFP-MIST-1024	**0.39**	**0.48**	**0.33**	**0.45**	**0.38**	**0.31**
MIST-2048	0.31	0.37	0.22	0.31	0.23	0.20
UncertFP-MIST-2048	**0.45**	**0.50**	**0.36**	**0.43**	**0.35**	**0.37**
MIST-4096	0.34	0.43	0.24	0.35	0.27	0.32
UncertFP-MIST-4096	**0.48**	**0.52**	**0.38**	**0.49**	**0.39**	**0.43**
MIST-4096 (released)	0.33	0.44	0.23	0.28	/	/
UncertFP-MIST-4096 (released)	**0.47**	**0.51**	**0.37**	**0.46**	/	/
DiffMS-1024	0.28	0.39	0.29	0.37	0.26	0.30
UncertFP-DiffMS-1024	**0.41**	**0.46**	**0.38**	**0.48**	**0.41**	**0.43**
DiffMS-2048	0.29	0.38	0.31	0.42	0.25	0.27
UncertFP-DiffMS-2048	**0.43**	**0.49**	**0.39**	**0.50**	**0.40**	**0.45**
DiffMS-4096	0.31	0.40	0.35	0.46	0.29	0.37
UncertFP-DiffMS-4096	**0.47**	**0.51**	**0.42**	**0.52**	**0.44**	**0.48**
DiffMS-4096 (released)	0.33	0.45	0.33	0.44	/	/
UncertFP-DiffMS-4096 (released)	**0.47**	**0.53**	**0.41**	**0.53**	/	/

Candidate ranking. Because fingerprint overlap does not directly test whether predicted fingerprints identify the true molecule among structures sharing a formula, we evaluate trained 4096-bit models on identical candidate sets within each dataset. UncertFP improves $\mathrm{Acc@}1$ by 0.92–10.37 percentage points and raises DiffMS-4096 MRR from 0.53 to 0.60 on MassSpecGym; complete candidate-set details and retrieval results are provided in Appendix D.

5.2 ABLATION STUDIES

Complementarity of spectral views. We ablate the Fragment-peak (F), Neutral-loss (N), and Peak-difference (Δ) views, together with view fusion, across all 18 configurations, holding data splits,

Table 2: CANOPUS spectral-view ablation using trained 4096-bit encoders. Checkmarks ✓ denote included views; **bold** marks column maxima.

	Spectral views			MIST-4096		DiffMS-4096	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-4096 / DiffMS-4096	✓			0.34	0.43	0.31	0.40
UncertFP-F	✓			0.36	0.45	0.32	0.41
UncertFP-N		✓		0.35	0.43	0.31	0.40
UncertFP-Δ			✓	0.39	0.40	0.29	0.40
UncertFP-FN	✓	✓		0.43	0.50	0.42	0.50
UncertFP-FΔ	✓		✓	0.41	0.50	0.41	0.50
UncertFP-NΔ		✓	✓	0.41	0.49	0.41	0.49
UncertFP-FNΔ	✓	✓	✓	**0.48**	**0.52**	**0.47**	**0.51**

initialization, training budgets, and checkpoint selection fixed within each comparison. We report representative results for MIST-4096 and DiffMS-4096 on CANOPUS in Table 2, with complete results provided in Appendix E. Fusing views consistently outperforms individual views for both encoders, and a view's standalone performance does not fully reflect its contribution to fusion. For DiffMS-4096 on the CANOPUS dataset, the UncertFP-Δ achieves lower Soft Tanimoto than UncertFP-F (0.29 vs. 0.32), yet adding Δ to UncertFP-FN improves Soft Tanimoto from 0.42 to 0.47. The full UncertFP-FNΔ achieves the highest Soft and Binary Tanimoto for both encoders, and this advantage holds across all the configurations. These results support the complementary value and necessity of using three spectral views and validate the effectiveness of our UncertFP.

Contribution of ranking supervision. We ablate each loss by setting its weight to zero. Adding ranking supervision to class-weighted BCE and Soft Tanimoto improves both fingerprint metrics by 0.01–0.03 across all settings; complete configurations and results are provided in Appendix F.

Effectiveness of evidence fusion. Using the same three views, we compare uniform probability averaging, deterministic bitwise weighting, and UncertFP to determine whether the gains come from evidence-aware fusion rather than from adding views alone. EqualAvg $<$ DetFusion $<$ UncertFP holds for both metrics across all evaluated settings, indicating that the gains arise from reliability- and conflict-aware fusion; complete comparisons are provided in Appendix G.

Contributions of disagreement discounting and the learnable fusion coefficient. As the learnable fusion coefficient η controls evidence accumulation and disagreement discounting controls evidence retained under conflict, we cross fixed versus learned η with disabled versus enabled discounting. For CANOPUS DiffMS-4096, discounting raises Soft from 0.41 to 0.45 at $\eta = 0.5$, and learning η with discounting retained raises it to 0.47; across all 18 configurations, each component improves both metrics and their combination performs best, as detailed in Appendix H.

5.3 Spectrum selection using uncertainty mass

Predictive entropy measures ambiguity in predicted probabilities, whereas uncertainty mass measures evidence insufficiency. To compare their ranking utility with predictions held fixed, we rank the spectra of eight CANOPUS configurations in ascending order of mean uncertainty mass or mean bitwise predictive entropy. We evaluate selective risk, defined as the mean error $1 - \mathrm{Soft}$ among spectra retained at the same coverage, against random ranking and the infeasible error-oracle lower bound. As shown in Figure 4 (protocol in Appendix C.7), at 25% and 50% coverage, uncertainty-mass ranking has lower observed selective risk than entropy ranking and random ranking in all eight configurations. Selective risk generally decreases as coverage decreases across both encoders, all three self-trained dimensionalities, and both released configurations, so uncertainty mass yields the lower observed selective risk within these CANOPUS settings.

5.4 Case studies

Two structural patterns recur among the spectra that baseline predictors misread. The first is a large ring, whose cleavages can follow a chain-like fragmentation route and thus produce a spectrum resembling that of an open-chain molecule. The second is a repeated substructure, such as a nitrogen-

or oxygen-containing heterocycle, whose local fragmentation pattern appears several times within one molecule so that different molecules carrying it produce mutually similar spectra. In both patterns the peaks alone do not separate the molecules, and the baseline responds with low probabilities for bits whose ground-truth states differ. Both patterns follow from how these molecules fragment, so the look-alikes recur rather than arise from measurement noise.

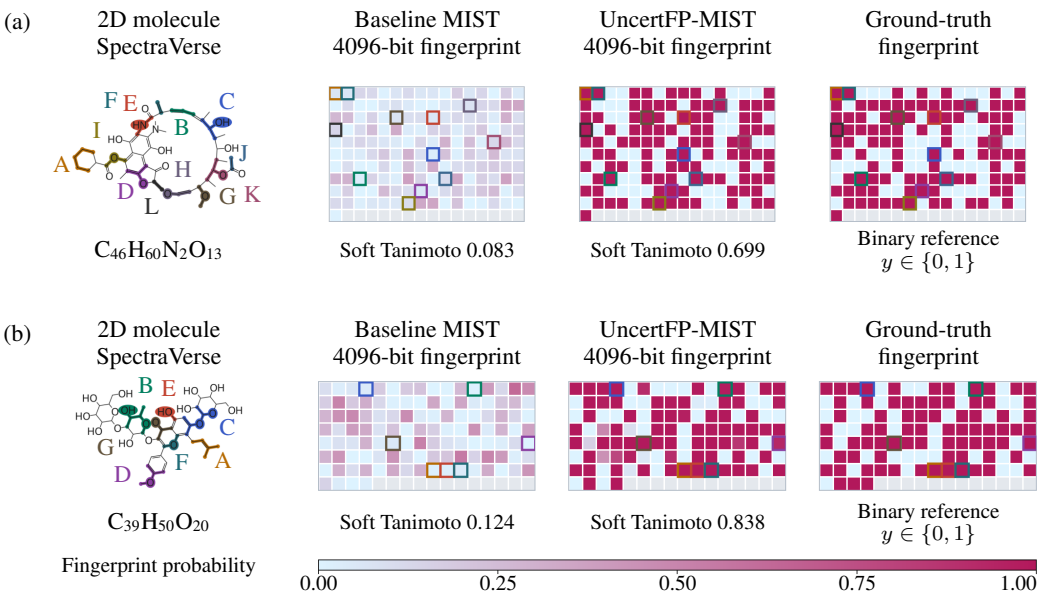

Figure 3: Recovery of annotated fingerprint bits in (a) a larger ring and (b) repeated oxygen-containing heterocycles. Each panel shows the molecular structure, baseline and UncertFP probabilistic fingerprints, and ground truth. Letters link local environments to hashed bits; one bit may represent multiple environments. Grids show a subset in matching bit order, with gray padding; Soft uses all 4096 bits.

Figure 3 examines two SpectraVerse-positive MIST-4096 cases, one per pattern. For the larger ring (a), the baseline assigns probabilities of 0.022–0.136 to the 12 annotated positive bits, and for the repeated oxygen-containing heterocycles (b), 0.003–0.051 to the seven annotated bits. UncertFP raises these probabilities to 0.993–0.995 and 0.933–0.994, recovering every annotated bit at the 0.5 threshold, and Soft Tanimoto increases from 0.083 to 0.699 and from 0.124 to 0.838, respectively. The difference lies not in the spectra, which remain similar to those of competing structures, but in the evidence they carry: because positive and negative evidence is represented per bit and per view, an ambiguous fragmentation pattern can support a present bit strongly in one molecule and only weakly in another. Appendix I extends this analysis to further molecules.

6 Discussion and Conclusion

In this paper, we propose UncertFP, an uncertainty-aware multi-view evidence fusion framework for molecular fingerprint prediction from tandem mass spectra that distinguishes evidence strength from predictive probability. UncertFP models bitwise positive and negative evidence across complementary spectral views, integrating reliability weighting, a learnable fusion coefficient, and disagreement discounting under a shared prior to produce probabilistic fingerprints and uncertainty mass. Experiments across three datasets, two spectral encoders, and three fingerprint dimensionalities demonstrate consistent improvements in molecular fingerprint prediction, while the complete prediction and scoring pipeline also improves candidate-structure ranking. Beyond prediction accuracy, spectrum-level uncertainty mass, supervised through the relative ordering of whole-fingerprint errors, supports selective prediction and yields lower observed selective risk than predictive entropy.

AI USE STATEMENT

We used generative AI tools to assist with language polishing. The authors reviewed the AI-assisted content and take responsibility for the final content of this paper.

REFERENCES

Aries Aisporna, H. Paul Benton, Andy Chen, Rico J. E. Derks, Jean Marie Galano, Martin Giera, and Gary Siuzdak. Neutral loss mass spectral data enhances molecular similarity analysis in METLIN. *Journal of the American Society for Mass Spectrometry*, 33(3):530–534, 2022.

Saleh Alseekh, Asaph Aharoni, Yariv Brotman, Kévin Contrepois, John D'Auria, Jan Ewald, Jennifer C. Ewald, Paul D. Fraser, Patrick Giavalisco, Robert D. Hall, Matthias Heinemann, Hannes Link, Jie Luo, Steffen Neumann, Jens Nielsen, Leonardo Perez de Souza, Kazuki Saito, Uwe Sauer, Frank C. Schroeder, Stefan Schuster, Gary Siuzdak, Aleksandra Skirycz, Lloyd W. Sumner, Michael P. Snyder, Huiru Tang, Takayuki Tohge, Yulan Wang, Weiwei Wen, Si Wu, Guowang Xu, Nicola Zamboni, and Alisdair R. Fernie. Mass spectrometry-based metabolomics: A guide for annotation, quantification and best reporting practices. *Nature Methods*, 18(7):747–756, 2021.

Sadjad Fakouri Baygi and Dinesh Kumar Barupal. IDSL_MINT: A deep learning framework to predict molecular fingerprints from mass spectra. *Journal of Cheminformatics*, 16(1):8, 2024.

Viktor Bengs, Eyke Hüllermeier, and Willem Waegeman. Pitfalls of epistemic uncertainty quantification through loss minimisation. In *Advances in Neural Information Processing Systems*, pp. 29205–29216, 2022.

Grigor Bezirganyan, Sana Sellami, Laure Berti-Equille, and Sébastien Fournier. Multimodal learning with uncertainty quantification based on discounted belief fusion. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, pp. 3142–3150, 2025.

Montgomery Bohde, Mrunali Manjrekar, Runzhong Wang, Shuiwang Ji, and Connor W. Coley. DiffMS: Diffusion generation of molecules conditioned on mass spectra. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 4737–4756, 2025.

Montgomery Bohde, Hongxuan Liu, Mrunali Manjrekar, Magdalena Lederbauer, Shuiwang Ji, Runzhong Wang, and Connor W. Coley. FRIGID: Scaling diffusion-based molecular generation from mass spectra at training and inference time. *arXiv preprint arXiv:2604.16648*, 2026.

Roman Bushuiev, Anton Bushuiev, Niek F. de Jonge, Adamo Young, Fleming Kretschmer, Raman Samusevich, Janne Heirman, Fei Wang, Luke Zhang, Kai Dührkop, Marcus Ludwig, Nils A. Haupt, Apurva Kalia, Corinna Brungs, Robin Schmid, Russell Greiner, Bo Wang, David S. Wishart, Li-Ping Liu, Juho Rousu, Wout Bittremieux, Hannes Rost, Tytus D. Mak, Soha Hassoun, Florian Huber, Justin J. J. van der Hooft, Michael A. Stravs, Sebastian Böcker, Josef Sivic, and Tomáš Pluskal. MassSpecGym: A benchmark for the discovery and identification of molecules. In *Advances in Neural Information Processing Systems*, pp. 110010–110027, 2024.

Roman Bushuiev, Anton Bushuiev, Raman Samusevich, Corinna Brungs, Josef Sivic, and Tomáš Pluskal. Self-supervised learning of molecular representations from millions of tandem mass spectra using DreaMS. *Nature Biotechnology*, 44(4):630–640, 2026.

Mengyuan Chen, Junyu Gao, and Changsheng Xu. R-EDL: Relaxing nonessential settings of evidential deep learning. In *International Conference on Learning Representations*, pp. 35563–35588, 2024.

Charles Corbière, Nicolas Thome, Avner Bar-Hen, Matthieu Cord, and Patrick Pérez. Addressing failure prediction by learning model confidence. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Kai Dührkop. Deep kernel learning improves molecular fingerprint prediction from tandem mass spectra. *Bioinformatics*, 38(Supplement 1):i342–i349, 2022.

Kai Dührkop, Huibin Shen, Marvin Meusel, Juho Rousu, and Sebastian Böcker. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences*, 112(41):12580–12585, 2015.

Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*, pp. 4878–4887, 2017.

Samuel Goldman, Jeremy Wohlwend, Martin Stražar, Guy Haroush, Ramnik J. Xavier, and Connor W. Coley. Annotating metabolite mass spectra with domain-inspired chemical formula transformers. *Nature Machine Intelligence*, 5(9):965–979, 2023.

Vishu Gupta, Hantao Qiang, Hsin-Hsiang Chung, Ehud Herbst, and Michael A. Skinnider. Comprehensive curation and harmonization of small-molecule MS/MS libraries in Spectraverse. *Analytical Chemistry*, 98(5):3934–3943, 2026.

Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2551–2566, 2023.

Markus Heinonen, Huibin Shen, Nicola Zamboni, and Juho Rousu. Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics*, 28(18):2333–2341, 2012.

Corey Hoang, Winnie Uritboonthai, Linh Hoang, Elizabeth M. Billings, Aries Aisporna, Farshad A. Nia, Rico J. E. Derks, James R. Williamson, Martin Giera, and Gary Siuzdak. Tandem mass spectrometry across platforms. *Analytical Chemistry*, 96(14):5478–5488, 2024.

Martin A. Hoffmann, Louis-Félix Nothias, Marcus Ludwig, Markus Fleischauer, Emily C. Gentry, Michael Witting, Pieter C. Dorrestein, Kai Dührkop, and Sebastian Böcker. High-confidence structural annotation of metabolites absent from spectral libraries. *Nature Biotechnology*, 40(3): 411–421, 2022.

Audun Jøsang, Javier Díaz, and Maria Rifqi. Cumulative and averaging fusion of beliefs. *Information Fusion*, 11(2):192–200, 2010.

Mira Juergens, Nis Meinert, Viktor Bengs, Eyke Hüllermeier, and Willem Waegeman. Is epistemic uncertainty faithfully represented by evidential deep learning methods? In *Proceedings of the 41st International Conference on Machine Learning*, pp. 22624–22642, 2024.

Botao Liu, Zhifeng Tang, and Tao Huan. Adduct-induced variability in tandem mass spectrometry. *Analytical Chemistry*, 97(31):17058–17066, 2025.

Wei Liu, Xiaodong Yue, Yufei Chen, and Thierry Denoeux. Trusted multi-view deep learning with opinion aggregation. *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 7585–7593, 2022.

Wei Liu, Yufei Chen, Xiaodong Yue, Changqing Zhang, and Shaorong Xie. Safe multi-view deep classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 8870–8878, 2023.

Jueqing Lu, Wray Buntine, Yuanyuan Qi, Joanna Dipnall, Belinda Gabbe, and Lan Du. Navigating conflicting views: Harnessing trust for learning. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 40411–40435, 2025.

Dai Hai Nguyen, Canh Hao Nguyen, and Hiroshi Mamitsuka. SIMPLE: Sparse interaction model over peaks of molecules for fast, interpretable metabolite identification from tandem mass spectra. *Bioinformatics*, 34(13):i323–i332, 2018.

Florian Rasche, Aleš Svatoš, Ravi Kumar Maddula, Christoph Böttcher, and Sebastian Böcker. Computing fragmentation trees from tandem mass spectrometry data. *Analytical Chemistry*, 83(4): 1243–1251, 2011.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems*, pp. 3183–3193, 2018.

Huibin Shen, Kai Dührkop, Sebastian Böcker, and Juho Rousu. Metabolite identification through multiple kernel learning on fragmentation trees. *Bioinformatics*, 30(12):i157–i164, 2014.

Maohao Shen, J. Jon Ryu, Soumya Ghosh, Yuheng Bu, Prasanna Sattigeri, Subhro Das, and Gregory W. Wornell. Are uncertainty quantification capabilities of evidential deep learning a mirage? In *Advances in Neural Information Processing Systems*, pp. 107830–107864, 2024.

Michael A. Stravs, Kai Dührkop, Sebastian Böcker, and Nicola Zamboni. MSNovelist: De novo structure generation from mass spectra. *Nature Methods*, 19(7):865–870, 2022.

Rens W. van der Heijden, Henning Kopp, and Frank Kargl. Multi-source fusion operations in subjective logic. In *2018 21st International Conference on Information Fusion (FUSION)*, pp. 1990–1997, 2018.

Mingxun Wang et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, 34(8):828–837, 2016.

Shipei Xing, Yan Hu, Zixuan Yin, Min Liu, Xiaoyu Tang, Mingliang Fang, and Tao Huan. Retrieving and utilizing hypothetical neutral losses from tandem mass spectra for spectral similarity analysis and unknown metabolite annotation. *Analytical Chemistry*, 92(21):14476–14483, 2020.

Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 16129–16137, 2024.

Qingyang Zhang, Haitao Wu, Changqing Zhang, Qinghua Hu, Huazhu Fu, Joey Tianyi Zhou, and Xi Peng. Provable dynamic fusion for low-quality multimodal data. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 41753–41769, 2023.

Chen Zhao, Dawei Du, Anthony Hoogs, and Christopher Funk. Open set action recognition via multi-label evidential learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22982–22991, 2023.

Yan Zhou Chen and Soha Hassoun. Learning from all views: A multiview contrastive framework for metabolite annotation. *Analytical Chemistry*, 98(9):6598–6606, 2026.

The appendix follows the paper's argument from problem framing to empirical evidence. We first position UncertFP relative to related representation, uncertainty, and fusion methods; then prove how disagreement and insufficient evidence affect the proposed quantities. We next specify the data, training, and evaluation protocols, followed by complete candidate-ranking and ablation results. We close with qualitative fingerprint-recovery cases that connect aggregate metrics to local molecular environments.

A Additional comparisons with related methods

This appendix asks three questions that connect related methods to UncertFP's design: what spectral information is represented, what prediction target is supervised, and at which stage reliability or conflict is handled. These questions distinguish representation learning, cross-representation alignment, risk estimation, and evidence fusion—the four design choices needed to interpret the comparison below.

Spectral representations and structure prediction. Which spectral relationships become available to a fingerprint predictor depends on the representation. Fragmentation trees organize candidate fragmentation relationships, and multiple kernel learning combines them with spectral features (Rasche et al., 2011; Shen et al., 2014). Sparse interactions between peaks and neural spectrum-to-fingerprint mappings provide additional approaches to extracting structural information (Nguyen et al., 2018; Baygi & Barupal, 2024). Large-scale spectral pretraining learns transferable representations (Bushuiev et al., 2026), while MIST explicitly encodes peak formulae and inter-peak formula differences (Goldman et al., 2023). UncertFP constructs three compositional views so that fragment, precursor–fragment, and interfragment information can contribute separate positive and negative evidence to each fingerprint bit.

Alignment across representations and fingerprint applications. Multi-view approaches differ in the information represented by their views and the predictions those views support. MVP aligns molecular graphs, fingerprints, individual spectra, and consensus spectra through pairwise contrastive objectives, then ranks candidates by embedding similarity (Zhou Chen & Hassoun, 2026). Consensus spectra combine multiple measurements of the same molecule. UncertFP instead constructs three views of one spectrum and fuses their support for individual fingerprint bits, producing both bit-presence probabilities and uncertainty mass. Fingerprints also support structure generation in MSNovelist, DiffMS, and FRIGID (Stravs et al., 2022; Bohde et al., 2025; 2026). Our downstream evaluation ranks structures within a supplied candidate set using fingerprint similarity and variability in sampled scores.

Prediction targets for risk estimation. The target of a confidence measure determines which prediction decisions it can support. COSMIC combines candidate scores and spectrum-related features to assess top-ranked CSI:FingerID annotations (Hoffmann et al., 2022). Our selection task concerns the quality of the entire predicted fingerprint. We therefore supervise spectrum-level uncertainty mass using the relative ordering of whole-fingerprint errors and evaluate the mean error of retained spectra. Multi-label evidential learning provides per-label Beta distributions for co-occurring labels (Zhao et al., 2023). UncertFP applies this representation to fingerprint bits, with prior means determined by training-set bit frequencies.

Reliability weights and the stage of discounting. Quality-aware Multimodal Fusion (QMF) uses historical unimodal losses to constrain fusion weights (Zhang et al., 2023), whereas our ranking supervision acts on spectrum-level uncertainty mass derived from fused predictions. Computational-trust methods learn instance-specific, probability-sensitive view discounts (Lu et al., 2025). ECML aggregates opinions through evidence averaging, and DBF discounts individual opinions before generalized averaging (Xu et al., 2024; Bezirganyan et al., 2025). UncertFP assigns contributions at the fingerprint-bit level and controls the balance between direct accumulation and weighted averaging. Inter-view disagreement then discounts the aggregated positive and negative evidence, after which the shared prior is added once. This construction separates the allocation of view contributions from the reduction of retained evidence under conflict.

B THEORETICAL PROOFS

Building on these design choices, the proofs answer four linked questions. First, does the discounting rule respond to inter-view conflict by retaining less evidence while preserving the positive-versus-negative balance? Second, when all views provide little support, does the predictive mean fall back to the shared prior while uncertainty mass records the loss of support? Third, does uncertainty-mass-weighted similarity recover the ranking induced by the true fingerprint in the ideal limit? Finally, how does per-bit posterior variance enter the variability of a sampled candidate score? We retain the main-text notation, omitting sample indices and, in scalar analyses, bit indices. Each result specifies the quantities held fixed. Sampling expectations and variances condition on the query spectrum S, precursor information $\mathcal{M}$, and trained model parameters.

B.1 MONOTONICITY OF EVIDENCE AND UNCERTAINTY MASS WITH INTER-VIEW DISAGREEMENT

First, we ask whether the discounting rule has the intended response to inter-view conflict: as disagreement grows, does it retain less evidence while preserving the positive-versus-negative balance? Fix one fingerprint bit and its undiscounted evidence $r^{(0)}, s^{(0)} \geq 0$, with total evidence $E_0 = r^{(0)} + s^{(0)} > 0$ and positive-evidence fraction $p_0 = r^{(0)}/E_0$. Keep the prior parameters $a \in (0,1)$, $W > 0$, and discount strength $\tau > 0$ fixed. For disagreement $\chi \geq 0$, the discounting rule gives

$$E(\chi) = e^{-\tau\chi} E_0, \qquad u(\chi) = \frac{W}{E(\chi) + W}, \qquad \mu(\chi) = (1 - u(\chi))p_0 + u(\chi)a. \tag{11}$$

Proof. Differentiating with respect to χ yields

$$\frac{\mathrm{d}E(\chi)}{\mathrm{d}\chi} = -\tau E(\chi) < 0, \qquad \frac{\mathrm{d}u(\chi)}{\mathrm{d}\chi} = \tau u(\chi)(1 - u(\chi)) > 0. \tag{12}$$

Both evidence components receive the same discount: $r(\chi) = e^{-\tau\chi} r^{(0)}$ and $s(\chi) = e^{-\tau\chi} s^{(0)}$. Their positive-evidence fraction remains p_0, giving

$$\begin{aligned} |\mu(\chi) - a| &= (1 - u(\chi))|p_0 - a|, \\ \frac{\mathrm{d}|\mu(\chi) - a|}{\mathrm{d}\chi} &= -\tau u(\chi)(1 - u(\chi))|p_0 - a| \leq 0. \end{aligned} \tag{13}$$

Increasing disagreement therefore strictly reduces total evidence and increases uncertainty mass. The predictive mean moves toward the prior mean when $p_0 \neq a$ and remains at a when $p_0 = a$. Disagreement discounting increases prior influence while preserving the balance between positive and negative evidence. □

B.2 SHRINKAGE TOWARD THE PRIOR AS EVIDENCE VANISHES

After conflict is considered, we ask what happens when all views provide little support: does the predictive mean fall back to the shared prior, and does uncertainty mass still record the loss of support? Fix one fingerprint bit, the prior parameters $a \in (0,1)$, $W > 0$, and the positive-evidence fraction $p = r/(r+s) \in [0,1]$. For total evidence $E = r + s > 0$,

$$u = \frac{W}{E + W}, \qquad \mu = (1 - u)p + ua.$$

Proof. Holding p fixed gives

$$\frac{\mathrm{d}u}{\mathrm{d}E} = -\frac{W}{(E+W)^2} < 0, \qquad |\mu - a| = \frac{E}{E+W}|p - a|. \tag{14}$$

As $E \to 0$, uncertainty mass approaches 1, while $|\mu - a| \to 0$ and hence $\mu \to a$. When $p = a$, every $E > 0$ yields $\mu = a$, although $u = W/(E + W)$ varies with E. This explains why uncertainty mass complements the predicted probability: identical means can correspond to different levels of evidential support. □

B.3 Limiting ranking consistency of uncertainty-mass-weighted Tanimoto

With the limiting behavior of the fused prediction established, we next ask whether uncertainty-mass-weighted similarity recovers the ranking induced by the true fingerprint in the ideal limit.

Corollary 1 (Limiting ranking consistency of uncertainty-mass-weighted Tanimoto)**.** *Fix a query spectrum and its precursor information, with a finite nonempty candidate set $\mathcal{C}(S, \mathcal{M})$. Let $\mathbf{y}^*$ denote its ground-truth binary fingerprint, corresponding to $\mathbf{y}_n$ in the main text. If $\boldsymbol{\mu} \to \mathbf{y}^*$, $\mathbf{u} \to \mathbf{0}$, and the limiting scores are distinct, then*

$$T_u(x) = \mathrm{Tan}\big(\boldsymbol{\mu}, \mathbf{g}(x); \mathbf{1} - \mathbf{u}\big) \longrightarrow T^*(x) = \mathrm{Tan}\big(\mathbf{y}^*, \mathbf{g}(x); \mathbf{1}\big). \tag{15}$$

Sufficiently close to this limit, T_u and T^* induce the same candidate ranking.

Proof. Write the predicted and limiting scores as

$$\begin{aligned} T_u(x) &= \mathrm{Tan}(\boldsymbol{\mu}, \mathbf{g}(x); \mathbf{1} - \mathbf{u}), \\ T^*(x) &= \mathrm{Tan}(\mathbf{y}^*, \mathbf{g}(x); \mathbf{1}). \end{aligned} \tag{16}$$

The result is immediate for a singleton candidate set. For at least two candidates, $\boldsymbol{\mu} \to \mathbf{y}^*$ and $\mathbf{u} \to \mathbf{0}$ imply $\mathbf{1} - \mathbf{u} \to \mathbf{1}$. Because $\varepsilon_T > 0$, weighted Tanimoto is continuous, giving $T_u(x) \to T^*(x)$ for every candidate. Finiteness and distinct limiting scores give a positive minimum gap,

$$\Delta_{\mathrm{gap}} = \min_{x \neq x'} |T^*(x) - T^*(x')| > 0. \tag{17}$$

Pointwise convergence on a finite set implies uniform convergence. Thus, sufficiently close to the limit,

$$\max_{x \in \mathcal{C}(S, \mathcal{M})} |T_u(x) - T^*(x)| < \frac{\Delta_{\mathrm{gap}}}{2}. \tag{18}$$

Every pairwise score difference retains its sign, preserving the candidate ordering. □

With the denominator floor set to 10^{-8}, T^* equals ordinary binary Tanimoto for a nonempty fingerprint union; an empty union receives zero. This establishes the ideal ranking limit of T_u: accurate fingerprint probabilities and vanishing uncertainty mass recover the corresponding reference ordering.

B.4 Per-bit Beta sampling and a similarity-variance bound

After the deterministic ranking limit, we ask how posterior variability enters candidate scoring. When the predicted fingerprint is sampled from the fitted Beta distributions, the standard-deviation penalty measures the resulting variability in candidate scores. With the query and trained model fixed, sampling uses a factorized model in which fingerprint bits are conditionally independent:

$$p(\boldsymbol{\rho} \mid S, \mathcal{M}) = \prod_{k=1}^{K} \mathrm{Beta}(\rho_k; \alpha_k, \beta_k), \tag{19}$$

$$\nu_k = \mathrm{Var}(\rho_k \mid S, \mathcal{M}) = \frac{\mu_k(1 - \mu_k)}{r_k + s_k + W + 1}. \tag{20}$$

For each query, we draw $B \geq 2$ independent bit-presence probability vectors. All candidates for that query share these draws. Their entries and candidate scores are

$$\rho_k^{(\ell)} \sim \mathrm{Beta}(\alpha_k, \beta_k), \qquad k = 1, \ldots, K, \quad \ell = 1, \ldots, B, \tag{21}$$

$$Z^{(\ell)}(x) = \mathrm{Tan}\big(\boldsymbol{\rho}^{(\ell)}, \mathbf{g}(x); \mathbf{1} - \mathbf{u}\big). \tag{22}$$

The sample mean centers the sampled scores when calculating their standard deviation:

$$\overline{Z}_B(x) = \frac{1}{B} \sum_{\ell=1}^{B} Z^{(\ell)}(x), \tag{23}$$

$$\widehat{\sigma}_B(x) = \sqrt{\frac{1}{B-1} \sum_{\ell=1}^{B} \big[Z^{(\ell)}(x) - \overline{Z}_B(x)\big]^2}. \tag{24}$$

The quantity $\widehat{\sigma}_B(x)$ supplies the standard-deviation penalty in the candidate score defined in Section 4.3.

Proposition 3 (Variance bound for weighted Beta-sampled similarity)**.** *Fix a candidate molecule x, its fingerprint $\mathbf{g} = \mathbf{g}(x) \in \{0,1\}^K$, and query-dependent weights $w_k = 1 - u_k \in [0,1]$. Under the factorized model, the independent variables $\rho_k \sim \mathrm{Beta}(\alpha_k, \beta_k)$ have variances ν_k. Define*

$$G_w = \sum_{k=1}^{K} w_k g_k, \qquad M = \max\{G_w, \varepsilon_T\}, \qquad Z(x) = \mathrm{Tan}(\boldsymbol{\rho}, \mathbf{g}; \mathbf{w}). \tag{25}$$

Then the score variance satisfies

$$\mathrm{Var}[Z(x)] \le \frac{1}{M^2} \sum_{k=1}^{K} w_k^2 \nu_k. \tag{26}$$

Proof. Write $\psi(\boldsymbol{\rho}) = \mathrm{Tan}(\boldsymbol{\rho}, \mathbf{g}; \mathbf{w})$. Before applying the denominator floor, weighted Tanimoto has numerator and denominator

$$N_{\mathrm{T}}(\boldsymbol{\rho}) = \sum_{k=1}^{K} w_k \rho_k g_k, \qquad D_{\mathrm{T}}(\boldsymbol{\rho}) = G_w + \sum_{k:g_k=0} w_k \rho_k. \tag{27}$$

Because $0 \le \rho_k \le 1$, we have $0 \le N_{\mathrm{T}} \le G_w$ and $D_{\mathrm{T}} \ge G_w$. When $g_k = 1$, the denominator is independent of ρ_k, giving a derivative magnitude at most w_k/M. When $g_k = 0$ and $D_{\mathrm{T}} > \varepsilon_T$, its magnitude is $N_{\mathrm{T}} w_k / D_{\mathrm{T}}^2 \le w_k/M$. When $g_k = 0$ and $D_{\mathrm{T}} < \varepsilon_T$, the derivative is zero. Therefore, wherever the derivative exists,

$$\left| \frac{\partial \psi}{\partial \rho_k} \right| \le \frac{w_k}{M}. \tag{28}$$

Continuity at $D_{\mathrm{T}} = \varepsilon_T$ gives the coordinatewise Lipschitz bound w_k/M. Let ρ_k' be an independent, identically distributed copy of ρ_k, and let $\boldsymbol{\rho}^{[k]}$ replace coordinate k by ρ_k'. Square brackets distinguish coordinate replacement from the sampling index. The Efron–Stein inequality gives

$$\begin{aligned} \mathrm{Var}[Z(x)] &\le \frac{1}{2} \sum_k \mathbb{E}\Big[\big(\psi(\boldsymbol{\rho}) - \psi(\boldsymbol{\rho}^{[k]})\big)^2\Big] \\ &\le \frac{1}{2M^2} \sum_k w_k^2 \mathbb{E}[(\rho_k - \rho_k')^2] \\ &= \frac{1}{M^2} \sum_k w_k^2 \nu_k. \end{aligned} \tag{29}$$

The final equality uses $\mathbb{E}[(\rho_k - \rho'_k)^2] = 2\nu_k$. The sample variance $\widehat{\sigma}^2_B(x)$ is an unbiased estimator of $\mathrm{Var}[Z(x)]$ and converges almost surely to it as $B \to \infty$. This bound links per-bit probability variability to candidate-score variability, providing an interpretation of the standard-deviation penalty. □

C Implementation and evaluation details

To make the comparisons in Section 5 interpretable, this appendix follows the experimental data flow from inputs to evaluation. We first fix the datasets and spectral views, then specify the evidence parameterization and training objectives, and finally define the metrics, candidate-ranking, and selective-risk protocols used to interpret the results.

C.1 Datasets and splits

To ensure that differences across configurations reflect model behavior rather than different examples, all encoder and fingerprint-size configurations within a dataset use the same splits, so their comparisons refer to the same training and evaluation spectra. CANOPUS uses the positive-ion version curated by MIST (Goldman et al., 2023), with spectra from Global Natural Products Social Molecular Networking (GNPS) (Wang et al., 2016) and assignments from `canopus_hplus_100_0.tsv`. MassSpecGym uses `split.tsv` from its data package (Bushuiev et al., 2024). SpectraVerse-positive contains positive-ion spectra (Gupta et al., 2026), with assignments from `connectivity_strict.tsv`. Labels and split records are matched by spectrum identifier. Table 3 reports the dataset statistics.

Table 3: Dataset statistics.

Dataset	Spectra	Formulae	Structure IDs	Train	Validation	Test
CANOPUS	8,030	4,804	7,131	6,810	401	819
MassSpecGym	231,104	17,634	28,929	194,119	19,429	17,556
SpectraVerse-positive	331,834	20,680	36,789	264,864	33,686	33,284

C.2 Spectrum preprocessing

Once the evaluation samples are fixed, the next question is what information each spectral view contributes. The three views expose fragment compositions and their relationships to the shared encoder while retaining the same precursor context. Among duplicate peaks, we retain the highest-intensity peak, then apply base-peak normalization and a square-root intensity transformation to obtain $\widetilde{I}_i$. After filtering peaks outside the precursor mass range and low-intensity peaks, we retain at most $L_{\max}$ peaks by intensity. Both encoders receive peaks annotated with subformulae (Goldman et al., 2023; Bohde et al., 2025). MassSpecGym configurations trained in our experiments use $L_{\max} = 10$; CANOPUS and SpectraVerse-positive retain all valid peaks after preprocessing, with $L_{\max} = L$.

Let $L_S \leq \min\{L, L_{\max}\}$ peaks remain. After reindexing, peak i has transformed intensity $\widetilde{I}_i$ and subformula vector $\mathbf{c}_i \in \mathbb{N}_0^D$, using the precursor formula's element order. Neutral-loss compositions are $\mathbf{n}_i = \mathbf{c}_{\mathrm{prec}} - \mathbf{c}_i$, with valid indices $\mathcal{I}_N = \{i \mid \mathbf{n}_i \in \mathbb{N}_0^D\}$.

Let $m(\mathbf{c})$ denote monoisotopic formula mass, while m_i remains the observed mass-to-charge ratio. Valid directed peak pairs satisfy $m(\mathbf{c}_i) > m(\mathbf{c}_j)$ and $\mathbf{d}_{ij} = \mathbf{c}_i - \mathbf{c}_j \in \mathbb{N}_0^D$. We rank pairs by $J_{ij} = \sqrt{\widetilde{I}_i \widetilde{I}_j}$, favoring pairs supported by stronger transformed intensities. At most $L_\Delta = 128$ pairs are retained to control sequence length; their index set is $\mathcal{P}_{L_\Delta}$.

The precursor token $\mathbf{t}_{\mathrm{prec}}$ encodes the precursor formula and adduct type. With $\|$ denoting token appending, the three views are

Table 4: Construction of the three spectral views.

View	Relationship	Representation
Fragment peaks (F)	Observed fragments	Subformula vectors with transformed intensities
Neutral losses (N)	Precursor–fragment	Composition differences with corresponding intensities
Peak differences (Δ)	Fragment pairs	Composition differences with pair weights

$$\begin{aligned}\mathcal{X}_{\mathrm{F}} &= \big((\mathbf{c}_i, \widetilde{I}_i)\big)_{i=1}^{L_S} \| \mathbf{t}_{\mathrm{prec}}, \\ \mathcal{X}_{\mathrm{N}} &= \big((\mathbf{n}_i, \widetilde{I}_i)\big)_{i \in \mathcal{I}_N} \| \mathbf{t}_{\mathrm{prec}}, \\ \mathcal{X}_{\Delta} &= \big((\mathbf{d}_{ij}, J_{ij})\big)_{(i,j) \in \mathcal{P}_{L_\Delta}} \| \mathbf{t}_{\mathrm{prec}}.\end{aligned} \tag{30}$$

When fewer than L_Δ valid pairs are available, all are retained. The full model uses $\mathcal{V} = \{\mathrm{F}, \mathrm{N}, \Delta\}$. Single-view ablations set $\chi_k = 0$, giving $\lambda_{v,k} = 1$ and $\delta_k = 1$, so fusion returns that view's evidence.

C.3 Evidence parameterization

The parameterization separates bit-presence probabilities from evidence strength while keeping evidence positive and bounded. Each view uses the shared encoder parameters θ_{enc}:

$$(\mathbf{h}_v, \boldsymbol{\pi}_v) = \Phi_{\theta_{\mathrm{enc}}}(\mathcal{X}_v). \tag{31}$$

The evidence head projects $\mathbf{h}_v$ to 256 dimensions, followed by layer normalization and a sigmoid linear unit (SiLU) activation. Two linear maps produce evidence preactivations and reliability scores,

$$(\boldsymbol{\ell}_v, \mathbf{q}_v) = g_{\mathrm{e}}(\mathbf{h}_v). \tag{32}$$

Here, $\boldsymbol{\pi}_v \in (0,1)^K$ and $\boldsymbol{\ell}_v, \mathbf{q}_v \in \mathbb{R}^K$. A capped softplus maps each real-valued output to total evidence $\kappa_{v,k} \in (\kappa_{\min}, \kappa_{\max}]$, with $0 < \kappa_{\min} < \kappa_{\max}$:

$$\kappa_{v,k} = \kappa_{\min} + \min\left\{\mathrm{softplus}(\ell_{v,k}), \kappa_{\max} - \kappa_{\min}\right\}. \tag{33}$$

We use $\kappa_{\min} = 10^{-4}$, $\kappa_{\max} = 300$, and shared prior strength $W = 2$. The prior mean is determined by training-set bit frequencies,

$$a_k = \mathrm{clip}\left(\frac{1}{N_{\mathrm{train}}} \sum_{n=1}^{N_{\mathrm{train}}} y_{n,k}, \varepsilon_a, 1 - \varepsilon_a\right), \qquad \varepsilon_a = 10^{-4}. \tag{34}$$

Clipping with $0 < \varepsilon_a < 1/2$ keeps both Beta prior shape parameters positive. Definition 1 gives the view-specific predictive means and uncertainty masses. Reliability scores $q_{v,k}$ and available-evidence fractions $1 - u_{v,k}$ jointly determine bitwise fusion weights in Equation 2, with $\gamma = 1$ and $\varepsilon_u = 10^{-8}$. The parameter $\gamma \geq 0$ controls the influence of the evidence fraction, and $\varepsilon_u > 0$ lower-bounds the logarithm's argument. The stop-gradient operator preserves the evidence fraction in the forward pass and blocks gradients through its contribution to the fusion weights.

The fusion coefficient and conflict strength are learned through unconstrained scalars:

$$\eta = \eta_{\min} + (\eta_{\max} - \eta_{\min})\sigma(\xi), \qquad \tau = \mathrm{softplus}(\zeta_\tau) > 0. \tag{35}$$

Here, σ is the logistic sigmoid, and $0 \leq \eta_{\min} < \eta_{\max} \leq 1$. We set $\eta_{\min} = 0.2$, $\eta_{\max} = 0.9$, and initialize $\xi = 0$ and $\zeta_\tau = 1$.

C.4 MODEL TRAINING SETTINGS

We evaluate both trained and released encoder initializations to examine UncertFP across the weight sources used in the main comparisons. MIST and DiffMS training procedures follow Goldman et al. (2023) and Bohde et al. (2025). Both MassSpecGym baselines trained in our experiments use the MIST encoder training settings released with FRIGID (Bohde et al., 2026), fixed to commit `5e53752`[1]. The same preset is used for 1024-, 2048-, and 4096-bit fingerprints.

Released-weight comparisons cover 4096-bit encoders on CANOPUS and MassSpecGym. CANOPUS MIST uses the public MIST release; MassSpecGym MIST uses `mist_msg.pt` from the MIST-FRIGID release. Both datasets use encoders from the public DiffMS release. We also train corresponding baselines for these four configurations, together with the 1024- and 2048-bit configurations and all three SpectraVerse-positive dimensionalities. On all 17,556 MassSpecGym test spectra, released MIST achieves Soft Tanimoto of 0.229119 and Binary Tanimoto of 0.282436 at threshold 0.5, reported as 0.23 and 0.28 in the main table.

UncertFP initializes from the corresponding trained baseline's best encoder checkpoint or the released 4096-bit weights and continues training. Its epoch counts denote additional epochs after initialization. Separate initial learning rates of 10^{-5} and 10^{-3} adapt the encoder and newly added evidence head at different update scales. Training uses Adam with weight decay 10^{-7}, random seed 1, and 32-bit floating-point precision (FP32) across four graphics processing units (GPUs). Dataset-specific settings are listed below.

Table 5: Training settings for each dataset.

Setting	CANOPUS	MassSpecGym	SpectraVerse-positive
Global batch size	128	256	128
Baseline initial learning rate	7.7×10^{-4}	7.7×10^{-4}	3×10^{-4}
Maximum baseline epochs	600	150	600
Maximum UncertFP epochs	300	300	300

The MassSpecGym preset uses learning-rate warmup over 10% of training steps, cosine decay, and an exponential moving average of parameters with decay 0.995. Trained baseline checkpoints are selected by validation fingerprint cosine similarity, while UncertFP checkpoints are selected by validation Soft Tanimoto.

C.5 TRAINING LOSSES

The losses supervise complementary aspects of prediction: individual bit states, complete-fingerprint overlap, and the ordering of spectrum-level uncertainty. Positive-class weighting increases the contribution of rare active bits; square-root scaling and capping control the magnitude of these weights:

$$\bar{\omega}_k = \sqrt{\frac{1 - a_k}{a_k + \varepsilon_\omega}}, \qquad \omega_k = \min(\bar{\omega}_k, \omega_{\max}), \tag{36}$$
$$\varepsilon_\omega = 10^{-8}, \qquad \omega_{\max} = 20.$$

For a minibatch $\mathcal{B}$, class-weighted binary cross-entropy (BCE) is

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{|\mathcal{B}|K} \sum_{n \in \mathcal{B}} \sum_{k=1}^{K} \left[\omega_k y_{n,k} \log \mu_{n,k} + (1 - y_{n,k}) \log(1 - \mu_{n,k}) \right]. \tag{37}$$

Logarithms without a specified base are natural logarithms. Sample error e_n and mean uncertainty mass U_n follow Section 4.3. The Soft Tanimoto loss supervises complete-fingerprint overlap through the mean minibatch error,

[1] Public MIST-FRIGID code, commit `5e53752`.

$$\mathcal{L}_{\text{sim}} = \frac{1}{|\mathcal{B}|} \sum_{n \in \mathcal{B}} e_n. \tag{38}$$

Ranking pairs are constructed from whole-fingerprint error differences as specified in Section 4.3, using $\Delta_e = 0.05$ and uncertainty-mass margin $m_{\text{rank}} = 0.02$. This objective trains the relative ordering needed for spectrum selection. The ranking loss is zero when the pair set is empty. The full model weights BCE, Soft Tanimoto, and uncertainty-mass ranking losses by 1, 1, and 0.1, respectively.

C.6 Evaluation metrics and candidate scoring

Fingerprint metrics measure agreement with the target fingerprint, while candidate ranking tests whether the prediction and scoring pipeline distinguishes structures sharing the query formula. Soft Tanimoto for spectrum n is $1 - e_n$, with denominator floor $\varepsilon_T = 10^{-8}$. Binary Tanimoto thresholds probabilities at 0.5 and compares the resulting binary fingerprint with ground truth. Both metrics are averaged over test spectra.

CANOPUS follows the MIST protocol, selecting PubChem structures with the same molecular formula as the query (Goldman et al., 2023). SpectraVerse-positive uses the same filtering procedure. MassSpecGym uses its official candidate sets for known-formula retrieval (Bushuiev et al., 2024). For each query, all methods use identical candidates with their respective predicted fingerprints and scoring procedures.

UncertFP weights fingerprint bits by $1 - u_k$, reducing the contribution of predictions with insufficient evidence. The standard deviation of Beta-sampled scores accounts for candidate-specific variability in similarity. Section 4.3 defines the complete score, and Appendix B.4 details sampling.

Top-k accuracy, $\text{Acc@}k$, is the fraction of queries whose true structure appears among the first k candidates. Mean reciprocal rank (MRR) averages the reciprocal rank of the true structure across queries. Appendix C.7 specifies coverage and selective risk.

C.7 Selective-risk protocol

This protocol asks whether uncertainty mass ranks spectra with lower fingerprint error than predictive entropy when the predictions themselves are held fixed. We therefore keep predictions fixed within each configuration and compare the two rankings at matched coverage. Figure 4 covers eight CANOPUS configurations: MIST and DiffMS with trained 1024-, 2048-, and 4096-bit encoders, plus both released 4096-bit encoders. Within each configuration, uncertainty mass and predictive entropy rank the same test spectra in ascending order and retain equal numbers at matched coverage. Mean uncertainty mass follows Section 4.3. Mean per-bit entropy is normalized by $\log 2$:

$$H_n = \frac{1}{K} \sum_{k=1}^{K} \frac{-\mu_{n,k}^{\text{clip}} \log \mu_{n,k}^{\text{clip}} - (1 - \mu_{n,k}^{\text{clip}}) \log(1 - \mu_{n,k}^{\text{clip}})}{\log 2}, \tag{39}$$

$$\mu_{n,k}^{\text{clip}} = \text{clip}(\mu_{n,k}, \varepsilon_H, 1 - \varepsilon_H), \qquad \varepsilon_H = 10^{-8}. \tag{40}$$

Let N_{test} be the test-set size and set $Q_n = U_n$ or $Q_n = H_n$. At coverage $c \in (0, 1]$, retain

$$N_c = \text{clip}\left(\lceil c N_{\text{test}} \rceil, 1, N_{\text{test}}\right). \tag{41}$$

Here, $\lceil \cdot \rceil$ is the ceiling function. If $\mathcal{I}_Q(c)$ indexes the first N_c spectra under ranking Q, selective risk is

$$\mathcal{R}_Q(c) = \frac{1}{N_c} \sum_{n \in \mathcal{I}_Q(c)} e_n. \tag{42}$$

Coverage ranges from 5% to 100% in increments of 1%. The CANOPUS test set contains 819 spectra, so 5%, 50%, and 100% coverage retain 41, 410, and 819 spectra. Within each configuration, all rankings yield the same mean error at full coverage.

For interval estimation, spectra are grouped by the connectivity block of their molecular InChIKey. We sample groups with replacement, retaining all spectra in each sampled group and preserving the multiplicity of repeated groups. Each resample uses its own total spectrum count to recompute retained counts, rerank spectra, and calculate risk at every coverage. Across 2000 resamples, the 2.5th and 97.5th percentiles of uncertainty-mass-ranked risk form pointwise 95% confidence intervals, shown in teal. These intervals describe test-sample variability with predictions held fixed.

The random baseline uses 2000 permutations of the correspondence between uncertainty mass and spectra, holding errors, labels, and predictions fixed. Black dotted curves show mean risk; gray shading gives the 2.5th–97.5th percentile range of random-ranking outcomes. The error oracle orders spectra by true prediction error, giving the lowest mean error at matched coverage, shown as gray dashed curves.

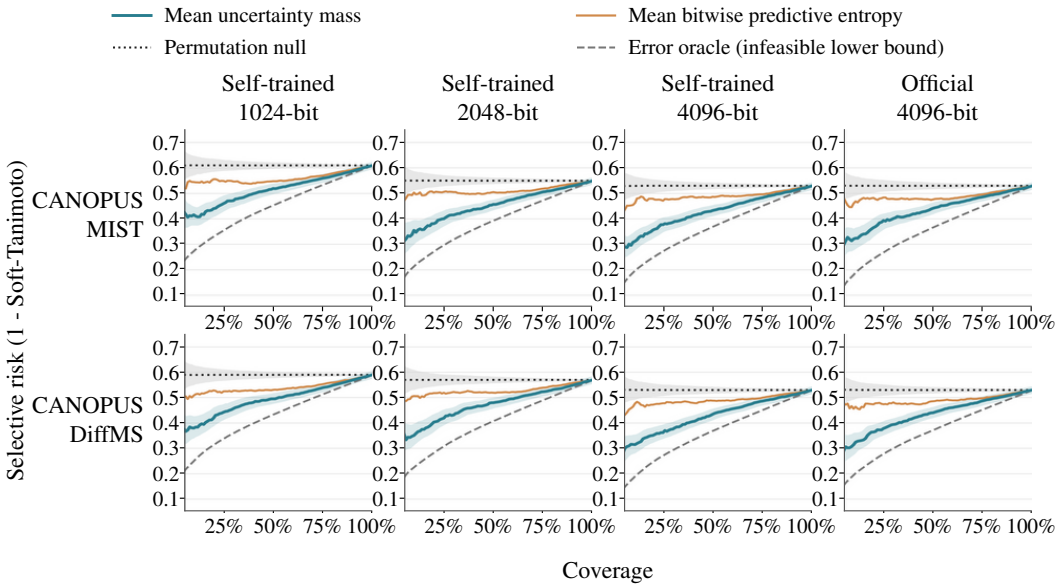

Figure 4: CANOPUS mean error $1 - \text{Soft}$ versus retained fraction. Curves use uncertainty mass (teal), predictive entropy (orange), random ranking (black dotted), or the error-oracle lower bound (gray dashed). Shading gives pointwise 95% confidence intervals from grouped resampling (teal) and 95% random-ranking ranges (gray).

D Complete candidate-structure ranking results

To test whether fingerprint gains translate into structure identification, we report the complete formula-matched retrieval results. Tables 6, 7, and 8 cover all three datasets, both encoders, and trained 1024-, 2048-, and 4096-bit configurations, together with released 4096-bit configurations on CANOPUS and MassSpecGym. Each query uses identical candidates across methods, with prediction and scoring as specified in Appendix C.6.

UncertFP improves every $\text{Acc@}k$ and MRR entry in all 22 paired comparisons. Across trained configurations, Acc@1 gains are 1.95-3.42% on CANOPUS, 5.38-10.37% on MassSpecGym, and 0.68-1.72% on SpectraVerse-positive. Released 4096-bit encoder weights follow the same ordering. The ranking benefit is therefore consistent across datasets, encoders, dimensionalities, and weight sources, but its magnitude is dataset-dependent: largest on MassSpecGym, moderate on CANOPUS, and smallest on the harder SpectraVerse-positive pool, where baseline Acc@1 is only 12-13%.

The gains are not confined to the top rank. On MassSpecGym, trained Acc@5 rises by 9.96-14.35%, exceeding the corresponding Acc@1 lifts, and MRR increases from 0.35-0.36 to 0.43-0.45 for MIST and from 0.53-0.54 to 0.60-0.61 for DiffMS. On CANOPUS the improvement is smaller and more uniform along the list; on SpectraVerse-positive, Acc@1 gains are modest but Acc@10 and Acc@100 still increase. Because improvements persist from Acc@1 through Acc@100 at every fingerprint length, UncertFP changes the rank of the true structure throughout the candidate list, rather than only increasing aggregate fingerprint overlap.

Table 6: Molecular retrieval performance on CANOPUS. Acc@k is the percentage of test queries whose true structure appears among the top k candidates; MRR denotes mean reciprocal rank. Bold marks the better value in each baseline–UncertFP pair with the same encoder, fingerprint dimensionality, and weight source.

	Retrieval accuracy (%)↑						**MRR ↑**
Method	Acc@1	Acc@5	Acc@10	Acc@20	Acc@50	Acc@100	Value
MIST-1024	26.62	56.65	66.30	73.63	82.42	86.08	0.39
UncertFP-MIST-1024	**29.67**	**58.61**	**68.13**	**75.82**	**84.00**	**87.79**	**0.43**
MIST-2048	26.13	56.65	66.42	73.75	82.78	88.40	0.40
UncertFP-MIST-2048	**29.30**	**59.22**	**69.84**	**76.80**	**85.23**	**89.50**	**0.43**
MIST-4096	26.25	57.39	67.16	73.75	82.78	85.71	0.40
UncertFP-MIST-4096	**28.82**	**61.90**	**70.82**	**77.17**	**84.37**	**89.13**	**0.44**
MIST-4096 (released)	26.13	54.58	65.69	72.65	81.93	86.81	0.39
UncertFP-MIST-4096 (released)	**28.33**	**58.36**	**67.89**	**77.53**	**85.71**	**89.74**	**0.42**
DiffMS-1024	24.42	55.07	65.69	73.75	81.32	86.32	0.38
UncertFP-DiffMS-1024	**26.37**	**57.26**	**68.25**	**76.68**	**84.13**	**87.91**	**0.41**
DiffMS-2048	26.62	55.80	66.42	73.99	81.20	85.35	0.40
UncertFP-DiffMS-2048	**30.04**	**58.73**	**68.62**	**76.19**	**84.37**	**88.03**	**0.43**
DiffMS-4096	25.89	55.31	65.20	75.09	82.66	86.69	0.39
UncertFP-DiffMS-4096	**28.21**	**58.24**	**68.74**	**76.80**	**84.13**	**87.67**	**0.42**
DiffMS-4096 (released)	26.25	55.92	65.45	73.63	81.20	86.69	0.39
UncertFP-DiffMS-4096 (released)	**29.06**	**59.95**	**70.09**	**79.24**	**86.32**	**90.60**	**0.43**

Table 7: Molecular retrieval performance on MassSpecGym. Acc@k is the percentage of test queries whose true structure appears among the top k candidates; MRR denotes mean reciprocal rank. Bold marks the better value in each baseline–UncertFP pair with the same encoder, fingerprint dimensionality, and weight source.

	Retrieval accuracy (%)↑						**MRR ↑**
Method	Acc@1	Acc@5	Acc@10	Acc@20	Acc@50	Acc@100	Value
MIST-1024	27.51	42.49	51.40	61.67	76.98	86.61	0.35
UncertFP-MIST-1024	**34.08**	**56.80**	**61.75**	**68.36**	**81.60**	**90.30**	**0.43**
MIST-2048	28.73	42.97	51.84	61.95	76.76	86.68	0.36
UncertFP-MIST-2048	**34.52**	**57.32**	**62.17**	**69.19**	**81.79**	**90.48**	**0.43**
MIST-4096	28.00	43.00	51.47	60.41	75.66	87.11	0.35
UncertFP-MIST-4096	**38.37**	**57.12**	**61.29**	**68.21**	**81.72**	**91.10**	**0.45**
MIST-4096 (released)	26.48	43.01	51.38	60.00	75.52	86.26	0.35
UncertFP-MIST-4096 (released)	**34.88**	**53.87**	**60.93**	**69.40**	**82.41**	**90.27**	**0.44**
DiffMS-1024	47.46	56.51	62.37	69.63	82.00	88.59	0.53
UncertFP-DiffMS-1024	**53.33**	**66.58**	**71.43**	**78.22**	**86.50**	**92.46**	**0.60**
DiffMS-2048	48.75	56.94	62.86	69.97	81.73	88.70	0.54
UncertFP-DiffMS-2048	**54.61**	**67.24**	**72.03**	**79.16**	**86.68**	**92.50**	**0.61**
DiffMS-4096	47.96	57.01	62.42	68.39	80.62	89.08	0.53
UncertFP-DiffMS-4096	**53.34**	**66.97**	**71.27**	**78.20**	**86.74**	**93.10**	**0.60**
DiffMS-4096 (released)	47.28	56.91	61.76	68.20	79.94	88.78	0.52
UncertFP-DiffMS-4096 (released)	**52.59**	**66.21**	**70.83**	**78.27**	**85.94**	**92.73**	**0.59**

Table 8: Molecular retrieval performance on SpectraVerse-positive. Acc@k is the percentage of test queries whose true structure appears among the top k candidates; MRR denotes mean reciprocal rank. Bold marks the better value in each baseline–UncertFP pair with the same encoder, fingerprint dimensionality, and weight source.

	Retrieval accuracy (%)↑						**MRR ↑**
Method	Acc@1	Acc@5	Acc@10	Acc@20	Acc@50	Acc@100	Value
MIST-1024	12.70	30.97	40.20	50.00	63.28	70.01	0.21
UncertFP-MIST-1024	**14.42**	**31.62**	**42.15**	**52.64**	**66.05**	**72.41**	**0.24**
MIST-2048	13.01	30.96	40.15	49.51	63.84	70.59	0.22
UncertFP-MIST-2048	**13.69**	**31.63**	**41.70**	**52.31**	**66.58**	**73.28**	**0.23**
MIST-4096	13.43	31.46	40.75	49.72	64.02	71.08	0.22
UncertFP-MIST-4096	**14.51**	**32.38**	**43.11**	**52.25**	**66.06**	**73.40**	**0.24**
DiffMS-1024	11.81	29.43	39.16	49.63	63.58	70.20	0.20
UncertFP-DiffMS-1024	**12.80**	**30.74**	**40.31**	**50.00**	**64.79**	**71.45**	**0.22**
DiffMS-2048	12.45	30.14	39.98	48.98	63.27	70.06	0.21
UncertFP-DiffMS-2048	**13.77**	**31.74**	**41.19**	**51.48**	**65.32**	**71.81**	**0.22**
DiffMS-4096	12.42	31.37	40.88	50.44	65.05	71.24	0.21
UncertFP-DiffMS-4096	**13.34**	**32.86**	**42.15**	**52.03**	**65.85**	**72.07**	**0.23**

E ADDITIONAL SPECTRAL-VIEW ABLATION RESULTS

The key question is whether a spectral view can add information that is not visible in its standalone score. We therefore extend the two CANOPUS-4096 comparisons in Table 2 to the remaining 16 configurations. Together, these comparisons cover three datasets, two encoders, and three fingerprint dimensionalities. We vary the input combination of fragment-peak (F), neutral-loss (N), and peak-difference (Δ) views while fixing splits, initialization, training budgets, and checkpoint selection within each group. Throughout Appendices E–H, Soft and Binary denote Soft Tanimoto and Binary Tanimoto. Tables pair MIST and DiffMS within each dataset and fingerprint dimensionality.

Single-view UncertFP remains close to the F-only encoder baseline, so most of the gain appears only after views are fused. Two-view models provide the first substantial lift, and FN is typically the strongest pair, consistent with F being the primary compositional view. Adding the third view then further improves both metrics. Complementarity is not predicted by standalone scores. For CANOPUS MIST-1024 in Table 9, Δ alone obtains Soft of 0.25, below 0.29 for F, yet adding Δ to FN increases Soft from 0.35 to 0.39 and Binary from 0.45 to 0.48. The converse also holds: Δ can be the strongest standalone view, as on MassSpecGym MIST-2048 and SpectraVerse-positive MIST-4096, while FNΔ still exceeds Δ and every two-view combination. Across all 18 configurations, FNΔ attains the unique maximum on both Soft and Binary, and adding Δ to FN never decreases either metric. In summary, F, N, and Δ contribute complementary evidence rather than redundant copies of the same signal. This further validates the effectiveness of our UncertFP in fusing multi-view evidence for molecular fingerprint prediction.

E.1 CANOPUS

Table 9: Spectral-view ablation on CANOPUS (1024-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-1024		DiffMS-1024	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-1024 / DiffMS-1024	✓			0.28	0.35	0.28	0.39
UncertFP-F	✓			0.29	0.36	0.29	0.42
UncertFP-N		✓		0.29	0.35	0.30	0.41
UncertFP-Δ			✓	0.25	0.40	0.27	0.39
UncertFP-FN	✓	✓		0.35	0.45	0.34	0.45
UncertFP-FΔ	✓		✓	0.34	0.42	0.33	0.42
UncertFP-NΔ		✓	✓	0.31	0.40	0.33	0.42
UncertFP-FNΔ	✓	✓	✓	**0.39**	**0.48**	**0.41**	**0.46**

Table 10: Spectral-view ablation on CANOPUS (2048-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-2048		DiffMS-2048	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-2048 / DiffMS-2048	✓			0.31	0.37	0.29	0.38
UncertFP-F	✓			0.34	0.39	0.31	0.39
UncertFP-N		✓		0.32	0.40	0.29	0.38
UncertFP-Δ			✓	0.31	0.39	0.31	0.35
UncertFP-FN	✓	✓		0.42	0.44	0.40	0.42
UncertFP-FΔ	✓		✓	0.42	0.41	0.39	0.42
UncertFP-NΔ		✓	✓	0.40	0.40	0.39	0.39
UncertFP-FNΔ	✓	✓	✓	**0.45**	**0.50**	**0.43**	**0.49**

Table 11: Spectral-view ablation on CANOPUS (4096-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-4096		DiffMS-4096	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-4096 / DiffMS-4096	✓			0.34	0.43	0.31	0.40
UncertFP-F	✓			0.36	0.45	0.32	0.41
UncertFP-N		✓		0.35	0.43	0.31	0.40
UncertFP-Δ			✓	0.39	0.40	0.29	0.40
UncertFP-FN	✓	✓		0.43	0.50	0.42	0.50
UncertFP-FΔ	✓		✓	0.41	0.50	0.41	0.50
UncertFP-NΔ		✓	✓	0.41	0.49	0.41	0.49
UncertFP-FNΔ	✓	✓	✓	**0.48**	**0.52**	**0.47**	**0.51**

E.2 MASSSPECGYM

Table 12: Spectral-view ablation on MassSpecGym (1024-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-1024		DiffMS-1024	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-1024 / DiffMS-1024	✓			0.21	0.26	0.29	0.37
UncertFP-F	✓			0.22	0.26	0.31	0.40
UncertFP-N		✓		0.22	0.27	0.30	0.39
UncertFP-Δ			✓	0.20	0.31	0.31	0.41
UncertFP-FN	✓	✓		0.31	0.38	0.36	0.43
UncertFP-FΔ	✓		✓	0.28	0.38	0.32	0.41
UncertFP-NΔ		✓	✓	0.28	0.37	0.31	0.41
UncertFP-FNΔ	✓	✓	✓	**0.33**	**0.45**	**0.38**	**0.48**

Table 13: Spectral-view ablation on MassSpecGym (2048-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-2048		DiffMS-2048	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-2048 / DiffMS-2048	✓			0.22	0.31	0.31	0.42
UncertFP-F	✓			0.23	0.33	0.33	0.43
UncertFP-N		✓		0.24	0.33	0.31	0.43
UncertFP-Δ			✓	0.26	0.30	0.31	0.45
UncertFP-FN	✓	✓		0.32	0.37	0.37	0.46
UncertFP-FΔ	✓		✓	0.32	0.34	0.35	0.45
UncertFP-NΔ		✓	✓	0.31	0.34	0.33	0.45
UncertFP-FNΔ	✓	✓	✓	**0.36**	**0.43**	**0.39**	**0.50**

Table 14: Spectral-view ablation on MassSpecGym (4096-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-4096		DiffMS-4096	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-4096 / DiffMS-4096	✓			0.24	0.35	0.35	0.46
UncertFP-F	✓			0.26	0.37	0.37	0.48
UncertFP-N		✓		0.27	0.35	0.37	0.47
UncertFP-Δ			✓	0.21	0.37	0.38	0.43
UncertFP-FN	✓	✓		0.37	0.47	0.41	0.50
UncertFP-FΔ	✓		✓	0.35	0.43	0.40	0.49
UncertFP-NΔ		✓	✓	0.35	0.43	0.39	0.47
UncertFP-FNΔ	✓	✓	✓	**0.38**	**0.49**	**0.42**	**0.52**

E.3 SPECTRAVERSE-POSITIVE

Table 15: Spectral-view ablation on SpectraVerse-positive (1024-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-1024		DiffMS-1024	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-1024 / DiffMS-1024	✓			0.24	0.19	0.26	0.30
UncertFP-F	✓			0.25	0.20	0.29	0.30
UncertFP-N		✓		0.25	0.20	0.26	0.30
UncertFP-Δ			✓	0.26	0.21	0.26	0.32
UncertFP-FN	✓	✓		0.32	0.28	0.39	0.37
UncertFP-FΔ	✓		✓	0.31	0.27	0.38	0.35
UncertFP-NΔ		✓	✓	0.31	0.26	0.36	0.35
UncertFP-FNΔ	✓	✓	✓	**0.38**	**0.31**	**0.41**	**0.43**

Table 16: Spectral-view ablation on SpectraVerse-positive (2048-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-2048		DiffMS-2048	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-2048 / DiffMS-2048	✓			0.23	0.20	0.25	0.27
UncertFP-F	✓			0.23	0.21	0.28	0.28
UncertFP-N		✓		0.23	0.22	0.28	0.27
UncertFP-Δ			✓	0.28	0.20	0.29	0.31
UncertFP-FN	✓	✓		0.33	0.31	0.37	0.42
UncertFP-FΔ	✓		✓	0.30	0.30	0.37	0.39
UncertFP-NΔ		✓	✓	0.29	0.30	0.35	0.38
UncertFP-FNΔ	✓	✓	✓	**0.35**	**0.37**	**0.40**	**0.45**

Table 17: Spectral-view ablation on SpectraVerse-positive (4096-bit fingerprints). Checkmarks denote included views and blanks omitted views. Bold marks the highest score in each column.

	Spectral views			MIST-4096		DiffMS-4096	
Method	F	N	Δ	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-4096 / DiffMS-4096	✓			0.27	0.32	0.29	0.37
UncertFP-F	✓			0.29	0.33	0.32	0.38
UncertFP-N		✓		0.28	0.32	0.30	0.39
UncertFP-Δ			✓	0.31	0.29	0.28	0.38
UncertFP-FN	✓	✓		0.36	0.41	0.40	0.45
UncertFP-FΔ	✓		✓	0.34	0.41	0.39	0.42
UncertFP-NΔ		✓	✓	0.32	0.40	0.37	0.39
UncertFP-FNΔ	✓	✓	✓	**0.39**	**0.43**	**0.44**	**0.48**

F COMPLETE LOSS ABLATION RESULTS

These ablations ask whether bitwise prediction, whole-fingerprint overlap, and uncertainty ordering provide complementary supervision. Therefore, we first denote class-weighted BCE, Soft Tanimoto loss, and uncertainty-mass ranking loss as B, S, and R, respectively. Then, we compare B, S, BS, BR, SR, and BSR across three datasets, two encoders, and three fingerprint dimensionalities. Within each configuration, view construction, fusion, splits, initialization, and training budgets remain fixed. Removing a loss sets its weight to zero while retaining the others.

We observe that individual objectives emphasize different aspects. Across all configurations, S outperforms B on the Soft Tanimoto metric, while B generally performs better on the Binary Tanimoto metric. Combining the two captures the advantages of both. In all 18 configurations, BS

matches or surpasses either individual objective on both metrics, and the two-loss model already accounts for most of the performance gain relative to B or S. Introducing R on top of this yields a strictly positive improvement: BSR improves over BS by 0.01–0.03 on both metrics and achieves the unique maximum. BR and SR still underperform BSR, indicating that ranking supervision cannot replace either the bit-wise loss or the full fingerprint loss.

Overall, class-weighted BCE (B), Soft Tanimoto loss (S), and uncertainty-mass ranking loss (R) provide supervision over different but complementary aspects of the same prediction: bit-wise accuracy, fingerprint overlap, and the ranking of spectrum-level uncertainty, and should therefore be optimized jointly.

F.1 CANOPUS

Table 18: Training-loss ablation on CANOPUS (1024-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			**MIST-1024**		**DiffMS-1024**	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.33	0.43	0.36	0.40
UncertFP-S		✓		0.34	0.40	0.37	0.37
UncertFP-BS	✓	✓		0.38	0.47	0.40	0.44
UncertFP-BR	✓		✓	0.38	0.46	0.39	0.43
UncertFP-SR		✓	✓	0.38	0.44	0.39	0.41
UncertFP-BSR	✓	✓	✓	**0.39**	**0.48**	**0.41**	**0.46**

Table 19: Training-loss ablation on CANOPUS (2048-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			**MIST-2048**		**DiffMS-2048**	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.40	0.44	0.38	0.42
UncertFP-S		✓		0.44	0.42	0.39	0.44
UncertFP-BS	✓	✓		0.44	0.48	0.42	0.48
UncertFP-BR	✓		✓	0.44	0.47	0.42	0.47
UncertFP-SR		✓	✓	0.44	0.45	0.41	0.45
UncertFP-BSR	✓	✓	✓	**0.45**	**0.50**	**0.43**	**0.49**

Table 20: Training-loss ablation on CANOPUS (4096-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			**MIST-4096**		**DiffMS-4096**	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.42	0.47	0.42	0.46
UncertFP-S		✓		0.44	0.45	0.45	0.42
UncertFP-BS	✓	✓		0.47	0.51	0.46	0.49
UncertFP-BR	✓		✓	0.47	0.48	0.46	0.49
UncertFP-SR		✓	✓	0.46	0.48	0.46	0.47
UncertFP-BSR	✓	✓	✓	**0.48**	**0.52**	**0.47**	**0.51**

F.2 MASSSPECGYM

Table 21: Training-loss ablation on MassSpecGym (1024-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			MIST-1024		DiffMS-1024	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.28	0.40	0.33	0.43
UncertFP-S		✓		0.30	0.39	0.34	0.40
UncertFP-BS	✓	✓		0.31	0.44	0.37	0.47
UncertFP-BR	✓		✓	0.31	0.41	0.37	0.44
UncertFP-SR		✓	✓	0.31	0.42	0.36	0.44
UncertFP-BSR	✓	✓	✓	**0.33**	**0.45**	**0.38**	**0.48**

Table 22: Training-loss ablation on MassSpecGym (2048-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			MIST-2048		DiffMS-2048	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.31	0.38	0.34	0.46
UncertFP-S		✓		0.32	0.39	0.35	0.44
UncertFP-BS	✓	✓		0.35	0.42	0.37	0.49
UncertFP-BR	✓		✓	0.35	0.42	0.38	0.47
UncertFP-SR		✓	✓	0.35	0.41	0.37	0.47
UncertFP-BSR	✓	✓	✓	**0.36**	**0.43**	**0.39**	**0.50**

Table 23: Training-loss ablation on MassSpecGym (4096-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			MIST-4096		DiffMS-4096	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.33	0.45	0.37	0.48
UncertFP-S		✓		0.35	0.43	0.38	0.48
UncertFP-BS	✓	✓		0.37	0.48	0.41	0.51
UncertFP-BR	✓		✓	0.37	0.48	0.41	0.50
UncertFP-SR		✓	✓	0.37	0.46	0.41	0.49
UncertFP-BSR	✓	✓	✓	**0.38**	**0.49**	**0.42**	**0.52**

F.3 SPECTRAVERSE-POSITIVE

Table 24: Training-loss ablation on SpectraVerse-positive (1024-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			MIST-1024		DiffMS-1024	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.29	0.29	0.36	0.36
UncertFP-S		✓		0.32	0.26	0.38	0.36
UncertFP-BS	✓	✓		0.35	0.30	0.40	0.41
UncertFP-BR	✓		✓	0.34	0.30	0.40	0.40
UncertFP-SR		✓	✓	0.35	0.30	0.40	0.39
UncertFP-BSR	✓	✓	✓	**0.38**	**0.31**	**0.41**	**0.43**

Table 25: Training-loss ablation on SpectraVerse-positive (2048-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			MIST-2048		DiffMS-2048	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.29	0.32	0.35	0.41
UncertFP-S		✓		0.33	0.32	0.37	0.41
UncertFP-BS	✓	✓		0.34	0.36	0.39	0.44
UncertFP-BR	✓		✓	0.34	0.34	0.38	0.43
UncertFP-SR		✓	✓	0.34	0.33	0.39	0.42
UncertFP-BSR	✓	✓	✓	**0.35**	**0.37**	**0.40**	**0.45**

Table 26: Training-loss ablation on SpectraVerse-positive (4096-bit fingerprints). Checkmarks denote included losses and blanks omitted losses. Bold marks the highest score in each column.

	Training losses			MIST-4096		DiffMS-4096	
Method	B	S	R	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP-B	✓			0.34	0.38	0.39	0.44
UncertFP-S		✓		0.36	0.36	0.42	0.43
UncertFP-BS	✓	✓		0.38	0.42	0.43	0.47
UncertFP-BR	✓		✓	0.37	0.41	0.43	0.46
UncertFP-SR		✓	✓	0.37	0.39	0.43	0.45
UncertFP-BSR	✓	✓	✓	**0.39**	**0.43**	**0.44**	**0.48**

G Complete view-fusion ablation results

Since all three views are available in every comparison, the question is whether the gain comes from adding multiple views alone or from modeling their bitwise reliability and inter-view disagreement. We compare fusion configurations across three datasets, two encoders, and three fingerprint dimensionalities, fixing splits, initialization, training budgets, and checkpoint selection within each group. EqualAvg averages view-specific bit-presence probabilities equally, while DetFusion weights these probabilities bitwise. Full UncertFP uses the evidence fusion and training configuration: positive and negative evidence are aggregated, discounted according to inter-view disagreement, and combined with the shared prior.

We observe that uniform averaging of the three views is not sufficient and is often harmful. UncertFP-EqualAvg falls below the F-only encoder baseline in 15 of 18 configurations, so additional views do not by themselves improve fingerprint prediction. Bitwise probability weighting recovers this loss: DetFusion exceeds both EqualAvg and the encoder baseline on both metrics in every configuration. Full UncertFP then adds a further strictly positive margin, and EqualAvg $<$ DetFusion $<$ UncertFP holds for Soft and Binary Tanimoto across all settings.

Therefore, uniform probability averaging cannot allocate per-bit reliability or reduce evidence under conflict, whereas our uncertainty-aware multi-view evidence fusion strategy does both. This ordering further separates the benefit of using multiple views from the benefit of allocating and discounting their evidence. Appendix H separates the contributions of disagreement discounting and the learnable fusion coefficient within that configuration.

G.1 CANOPUS

Table 27: View-fusion ablation on CANOPUS (1024-bit fingerprints). Bold marks the highest score in each column.

	MIST-1024		DiffMS-1024	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-1024 / DiffMS-1024	0.28	0.35	0.28	0.39
UncertFP-EqualAvg	0.29	0.36	0.23	0.37
UncertFP-DetFusion	0.37	0.43	0.33	0.42
UncertFP	**0.39**	**0.48**	**0.41**	**0.46**

Table 28: View-fusion ablation on CANOPUS (2048-bit fingerprints). Bold marks the highest score in each column.

	MIST-2048		DiffMS-2048	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-2048 / DiffMS-2048	0.31	0.37	0.29	0.38
UncertFP-EqualAvg	0.26	0.33	0.22	0.32
UncertFP-DetFusion	0.40	0.47	0.32	0.41
UncertFP	**0.45**	**0.50**	**0.43**	**0.49**

Table 29: View-fusion ablation on CANOPUS (4096-bit fingerprints). Bold marks the highest score in each column.

	MIST-4096		DiffMS-4096	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-4096 / DiffMS-4096	0.34	0.43	0.31	0.40
UncertFP-EqualAvg	0.35	0.44	0.25	0.33
UncertFP-DetFusion	0.38	0.47	0.39	0.47
UncertFP	**0.48**	**0.52**	**0.47**	**0.51**

G.2 MassSpecGym

Table 30: View-fusion ablation on MassSpecGym (1024-bit fingerprints). Bold marks the highest score in each column.

	MIST-1024		DiffMS-1024	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-1024 / DiffMS-1024	0.21	0.26	0.29	0.37
UncertFP-EqualAvg	0.15	0.23	0.20	0.31
UncertFP-DetFusion	0.32	0.40	0.37	0.42
UncertFP	**0.33**	**0.45**	**0.38**	**0.48**

Table 31: View-fusion ablation on MassSpecGym (2048-bit fingerprints). Bold marks the highest score in each column.

	MIST-2048		DiffMS-2048	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-2048 / DiffMS-2048	0.22	0.31	0.31	0.42
UncertFP-EqualAvg	0.17	0.27	0.29	0.41
UncertFP-DetFusion	0.28	0.36	0.32	0.43
UncertFP	**0.36**	**0.43**	**0.39**	**0.50**

Table 32: View-fusion ablation on MassSpecGym (4096-bit fingerprints). Bold marks the highest score in each column.

	MIST-4096		DiffMS-4096	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-4096 / DiffMS-4096	0.24	0.35	0.35	0.46
UncertFP-EqualAvg	0.21	0.34	0.31	0.44
UncertFP-DetFusion	0.29	0.46	0.40	0.50
UncertFP	**0.38**	**0.49**	**0.42**	**0.52**

G.3 SPECTRAVERSE-POSITIVE

Table 33: View-fusion ablation on SpectraVerse-positive (1024-bit fingerprints). Bold marks the highest score in each column.

	MIST-1024		DiffMS-1024	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-1024 / DiffMS-1024	0.24	0.19	0.26	0.30
UncertFP-EqualAvg	0.26	0.20	0.18	0.25
UncertFP-DetFusion	0.36	0.27	0.39	0.40
UncertFP	**0.38**	**0.31**	**0.41**	**0.43**

Table 34: View-fusion ablation on SpectraVerse-positive (2048-bit fingerprints). Bold marks the highest score in each column.

	MIST-2048		DiffMS-2048	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-2048 / DiffMS-2048	0.23	0.20	0.25	0.27
UncertFP-EqualAvg	0.17	0.17	0.22	0.23
UncertFP-DetFusion	0.34	0.32	0.32	0.35
UncertFP	**0.35**	**0.37**	**0.40**	**0.45**

Table 35: View-fusion ablation on SpectraVerse-positive (4096-bit fingerprints). Bold marks the highest score in each column.

	MIST-4096		DiffMS-4096	
Method	Soft ↑	Binary ↑	Soft ↑	Binary ↑
MIST-4096 / DiffMS-4096	0.27	0.32	0.29	0.37
UncertFP-EqualAvg	0.22	0.26	0.24	0.34
UncertFP-DetFusion	0.28	0.33	0.41	0.45
UncertFP	**0.39**	**0.43**	**0.44**	**0.48**

H COMPLETE CONFLICT-DISCOUNTING AND FUSION-COEFFICIENT ABLATIONS

After the preceding comparison establishes the benefit of evidence-aware fusion, we ask which two operations account for that gain. Evidence accumulation and inter-view conflict affect different aspects of fusion: the coefficient η controls how evidence is accumulated, whereas disagreement discounting controls how much evidence is retained under conflict. We cross these components across all 18 configurations.

Discounting is disabled with $\delta_k \equiv 1$ or enabled with $\delta_k = \exp(-\tau\chi_k)$. The coefficient η is fixed at $\eta = 0.5$ or learned using Appendix C.3. The three-view evidence model and all loss terms and weights remain identical. Within each group, models are trained separately from the same initial encoder and evidence-head weights, with identical splits, random seeds, optimizers, training budgets, and checkpoint selection. Full UncertFP rows use the corresponding trained-model results from Table 1.

We can observe that each factor helps on its own. Enabling discounting at $\eta = 0.5$ improves both metrics in every configuration, as does learning η without discounting. Neither operation subsumes the other: discounting can outperform a learned η on Binary Tanimoto, while a learned η can outperform discounting on Soft Tanimoto, as on CANOPUS MIST-4096 (0.50 vs. 0.49 Binary; 0.46 vs. 0.45 Soft). Their combination is strictly best. Full UncertFP attains the unique maximum in every column of Tables 36–44 and improves Soft Tanimoto by 0.04–0.07 over the no-discount, fixed-η

evidence model. For CANOPUS DiffMS-4096, matching the main-text example, discounting raises Soft from 0.41 to 0.45 at $\eta = 0.5$, and learning η with discounting raises it further to 0.47.

Overall, η and disagreement discounting are complementary controls inside the same fusion rule, rather than interchangeable implementations of it. Learning η sets how evidence is accumulated across views; discounting then reduces retained evidence when those views conflict. Both are required for the full UncertFP framework.

H.1 CANOPUS

Table 36: Discounting and fusion-coefficient ablation on CANOPUS (1024-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		**MIST-1024**		**DiffMS-1024**	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.32	0.42	0.34	0.41
UncertFP (fixed fusion)	✓		0.35	0.45	0.38	0.44
UncertFP (no discount)		✓	0.37	0.47	0.39	0.45
UncertFP	✓	✓	**0.39**	**0.48**	**0.41**	**0.46**

Table 37: Discounting and fusion-coefficient ablation on CANOPUS (2048-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		**MIST-2048**		**DiffMS-2048**	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.38	0.46	0.36	0.43
UncertFP (fixed fusion)	✓		0.43	0.49	0.40	0.47
UncertFP (no discount)		✓	0.44	0.48	0.41	0.48
UncertFP	✓	✓	**0.45**	**0.50**	**0.43**	**0.49**

Table 38: Discounting and fusion-coefficient ablation on CANOPUS (4096-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		**MIST-4096**		**DiffMS-4096**	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.42	0.46	0.41	0.45
UncertFP (fixed fusion)	✓		0.45	0.50	0.45	0.47
UncertFP (no discount)		✓	0.46	0.49	0.46	0.47
UncertFP	✓	✓	**0.48**	**0.52**	**0.47**	**0.51**

H.2 MASSSPECGYM

Table 39: Discounting and fusion-coefficient ablation on MassSpecGym (1024-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		**MIST-1024**		**DiffMS-1024**	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.27	0.39	0.34	0.43
UncertFP (fixed fusion)	✓		0.31	0.43	0.36	0.45
UncertFP (no discount)		✓	0.32	0.44	0.37	0.47
UncertFP	✓	✓	**0.33**	**0.45**	**0.38**	**0.48**

Table 40: Discounting and fusion-coefficient ablation on MassSpecGym (2048-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		**MIST-2048**		**DiffMS-2048**	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.31	0.39	0.33	0.46
UncertFP (fixed fusion)	✓		0.34	0.41	0.36	0.48
UncertFP (no discount)		✓	0.35	0.42	0.38	0.47
UncertFP	✓	✓	**0.36**	**0.43**	**0.39**	**0.50**

Table 41: Discounting and fusion-coefficient ablation on MassSpecGym (4096-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		**MIST-4096**		**DiffMS-4096**	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.33	0.44	0.38	0.48
UncertFP (fixed fusion)	✓		0.36	0.46	0.40	0.50
UncertFP (no discount)		✓	0.36	0.47	0.41	0.51
UncertFP	✓	✓	**0.38**	**0.49**	**0.42**	**0.52**

H.3 SPECTRAVERSE-POSITIVE

Table 42: Discounting and fusion-coefficient ablation on SpectraVerse-positive (1024-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		**MIST-1024**		**DiffMS-1024**	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.31	0.25	0.34	0.38
UncertFP (fixed fusion)	✓		0.34	0.29	0.38	0.42
UncertFP (no discount)		✓	0.37	0.29	0.40	0.42
UncertFP	✓	✓	**0.38**	**0.31**	**0.41**	**0.43**

Table 43: Discounting and fusion-coefficient ablation on SpectraVerse-positive (2048-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		**MIST-2048**		**DiffMS-2048**	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.29	0.31	0.34	0.40
UncertFP (fixed fusion)	✓		0.32	0.34	0.35	0.43
UncertFP (no discount)		✓	0.33	0.36	0.36	0.44
UncertFP	✓	✓	**0.35**	**0.37**	**0.40**	**0.45**

Table 44: Discounting and fusion-coefficient ablation on SpectraVerse-positive (4096-bit fingerprints). Checkmarks indicate enabled disagreement discounting or a learned fusion coefficient; a blank in the Learn-η column means that η is fixed at 0.5; bold marks the highest score in each column.

	Fusion components		MIST-4096		DiffMS-4096	
Method	Discount	Learn η	Soft ↑	Binary ↑	Soft ↑	Binary ↑
UncertFP (no discount, fixed fusion)			0.34	0.37	0.38	0.42
UncertFP (fixed fusion)	✓		0.36	0.40	0.41	0.45
UncertFP (no discount)		✓	0.38	0.42	0.42	0.46
UncertFP	✓	✓	**0.39**	**0.43**	**0.44**	**0.48**

I Visual analysis of molecular fingerprint prediction

Whole-fingerprint metrics show whether predictions improve; the gallery asks which local structural bits account for those gains and whether the pattern extends across datasets. We therefore inspect recovered positive bits at the molecular-environment level, using the two main-text cases as the reference point and the additional galleries as cross-dataset checks. The gallery is qualitative and illustrates local structural contexts rather than estimating their prevalence.

A positive bit is recovered when its ground-truth value is 1 and its predicted probability increases from below 0.5 under the baseline to at least 0.5 under UncertFP. Each figure presents the molecular structure, baseline prediction, UncertFP prediction, and ground-truth fingerprint. The three grids display the same fingerprint subset in matching bit order. Predictions share a probability scale from 0 to 1, ground truth is binary, and gray cells provide padding. Soft Tanimoto uses all 4096 bits.

Letters connect local atomic environments to fingerprint bits; the associated annotations report bit indices, environment radii, and predicted probabilities. In these figures, p, y, and r denote predicted probability, binary ground truth, and environment radius in bonds. A hashed fingerprint bit can represent multiple local environments, and the annotations identify the environments shown in each example.

I.1 Bit-level results for the main-text examples

The main-text cases examine two structural patterns associated with spectral ambiguity: larger rings, whose spectra can resemble those of chain structures, and repeated heterocyclic substructures that can produce similar spectral patterns across molecules. Figure 3 in the main-text uses SpectraVerse-positive MIST-4096 predictions for a larger-ring example and an example with repeated oxygen-containing heterocycles.

Table 45: Representative fingerprint bits from the two main-text examples. All listed bits have ground-truth value 1; letter labels are assigned separately within each panel.

Panel	Label	Bit	Radius	Baseline probability	UncertFP probability
(a)	B	2910	1	0.022	0.993
	C	2212	1	0.023	0.993
	L	1274	1	0.136	0.995
(b)	B	550	2	0.009	0.983
	C	222	2	0.040	0.993
	F	3652	1	0.036	0.993

Panel (a) corresponds to `SPECTRAVERSE0000143996`, with formula $C_{46}H_{60}N_2O_{13}$. The 12 annotated positive-bit probabilities rise from 0.022–0.136 to 0.993–0.995, while Soft Tanimoto increases from 0.083 to 0.699. Panel (b), `SPECTRAVERSE0000155991`, has formula $C_{39}H_{50}O_{20}$. Its seven annotated positive-bit probabilities increase from 0.003–0.051 to 0.933–0.994, and Soft rises from 0.124 to 0.838.

All annotated bits cross the 0.5 threshold, recovering local structural features missed by the baseline. These observations illustrate the structural predictions supported by UncertFP's per-bit evidence representation. Table 45 lists representative values; Figures 17 and 23 show all annotations.

I.2 MassSpecGym

Do the same bit-level recoveries appear outside the SpectraVerse-positive cases? The MassSpecGym gallery uses MIST-FRIGID-4096 to inspect this question across local environments within a molecule. In Figure 5, eight annotated positive bits describe environments in different molecular regions. Their baseline probabilities of 0.030–0.187 increase to 0.801–0.982 under UncertFP, changing all eight bit predictions from absent to present. Whole-fingerprint Soft Tanimoto rises from 0.128 to 0.369.

A further example, shown in Figure 6, recovers seven annotated positive bits and increases Soft from 0.122 to 0.718. The recovered environments have radii of one and two bonds, connecting the improved fingerprint predictions to identifiable local molecular neighborhoods.

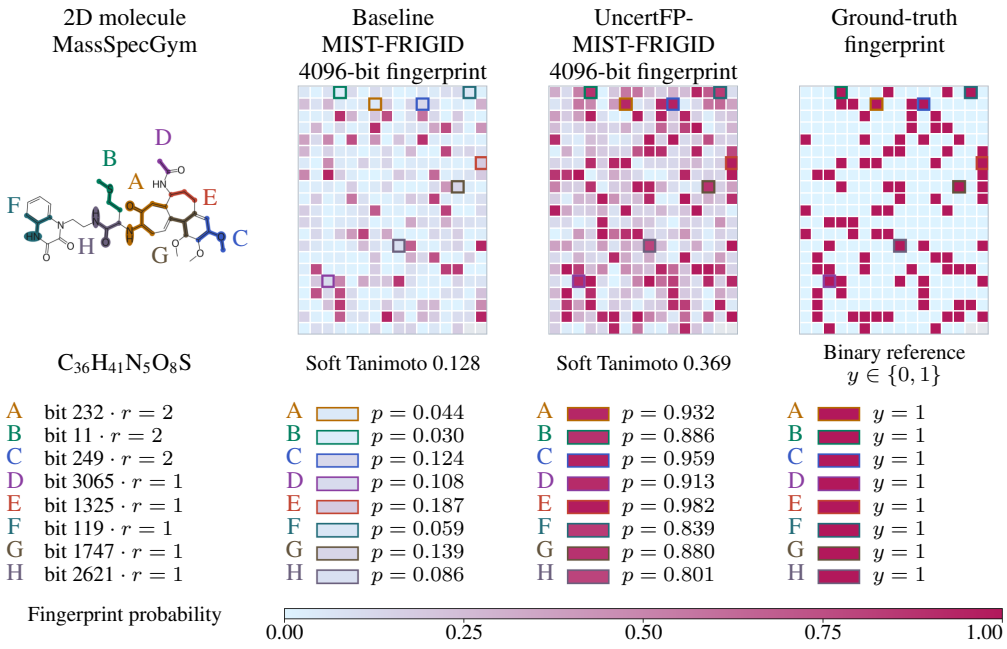

Figure 5: Recovery of 8 positive fingerprint bits.

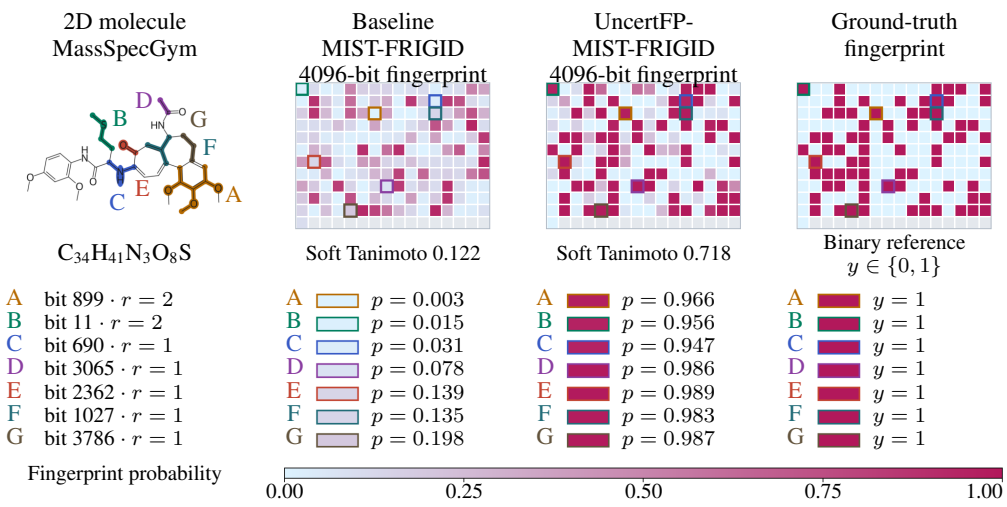

Figure 6: Recovery of 7 positive fingerprint bits.

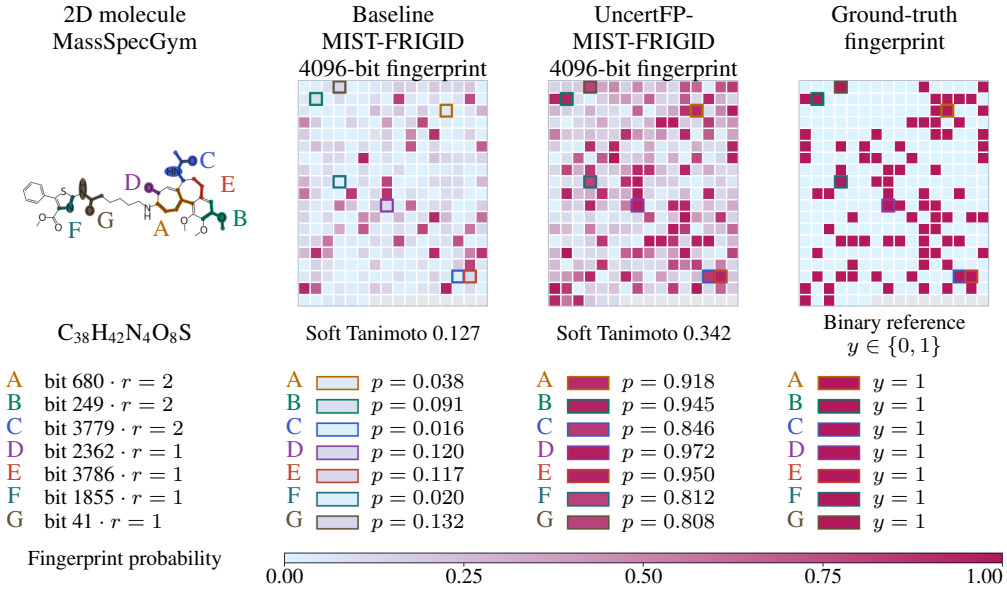

Figure 7: Recovery of 7 positive fingerprint bits.

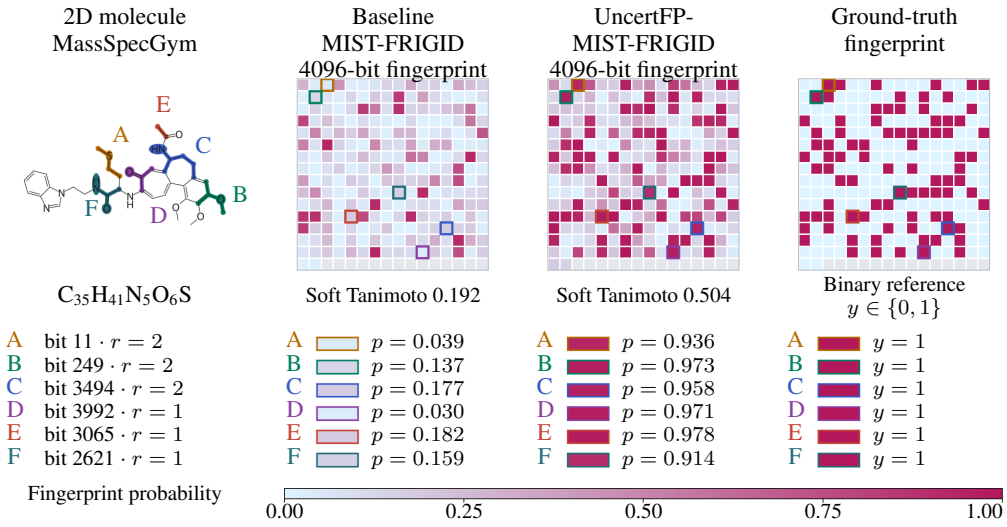

Figure 8: Recovery of 6 positive fingerprint bits.

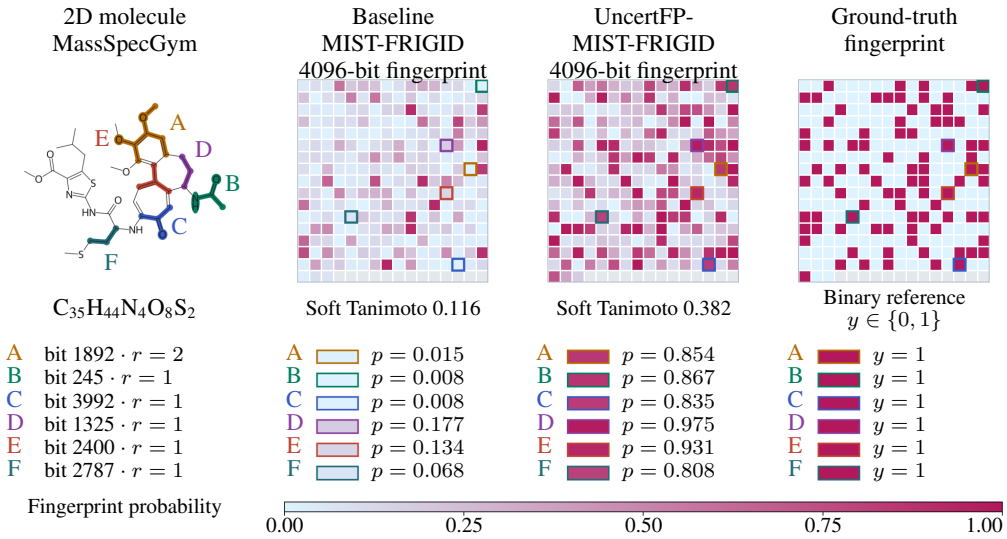

Figure 9: Recovery of 6 positive fingerprint bits.

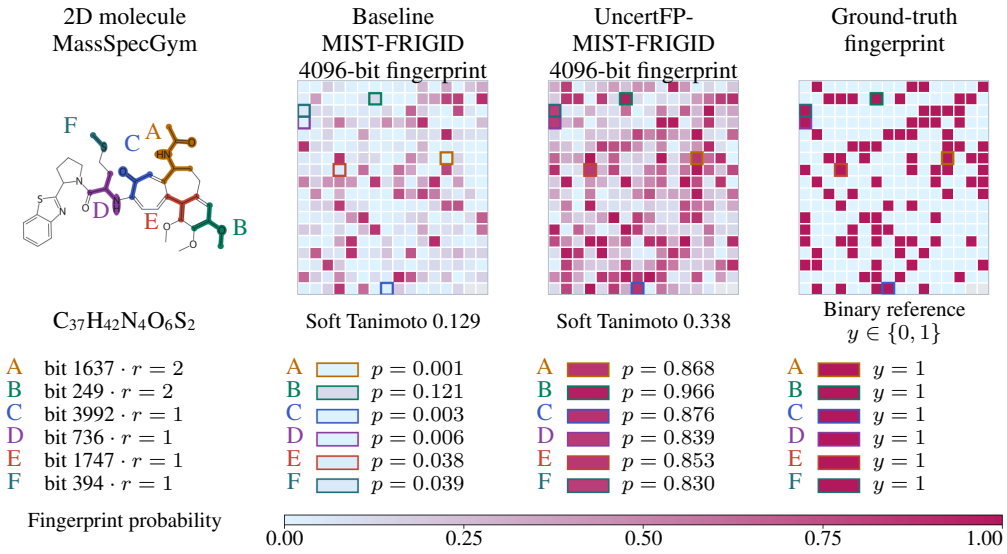

Figure 10: Recovery of 6 positive fingerprint bits.

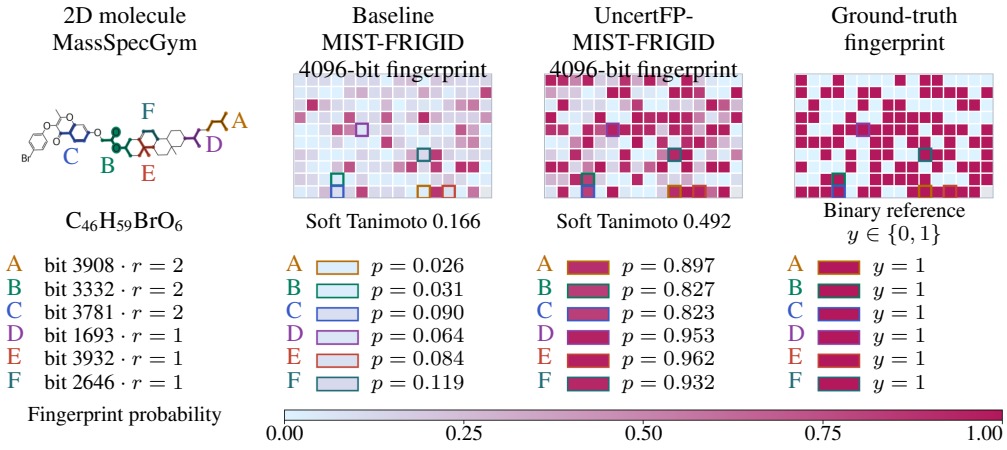

Figure 11: Recovery of 6 positive fingerprint bits.

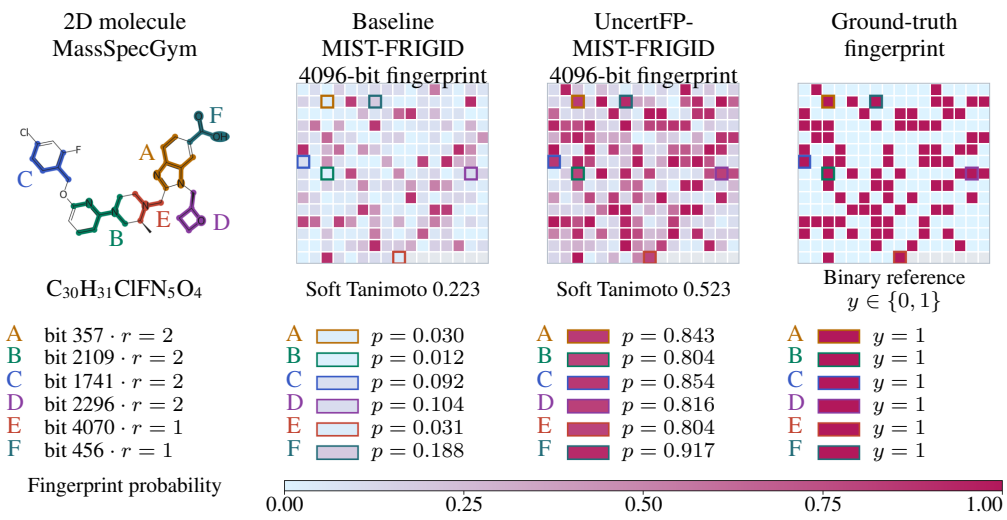

Figure 12: Recovery of 6 positive fingerprint bits.

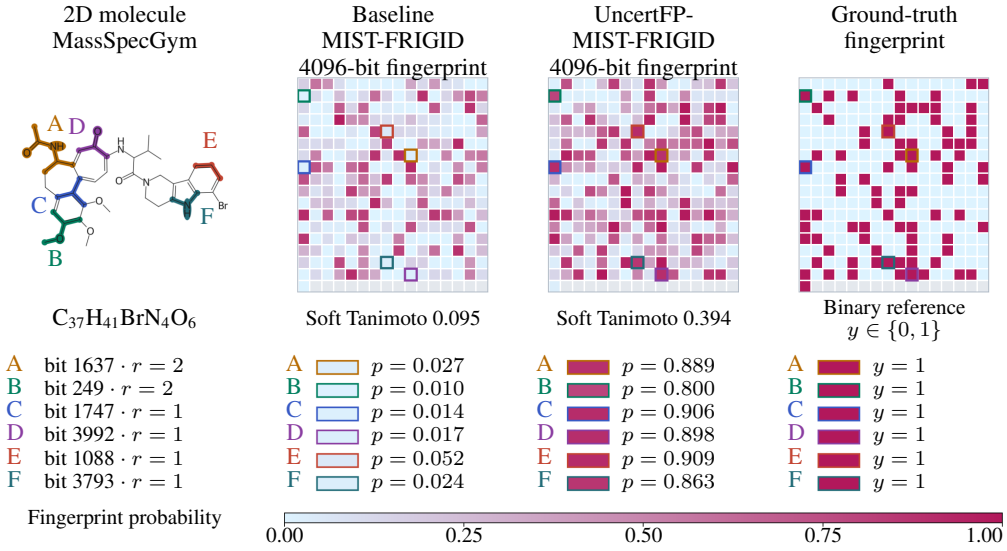

Figure 13: Recovery of 6 positive fingerprint bits.

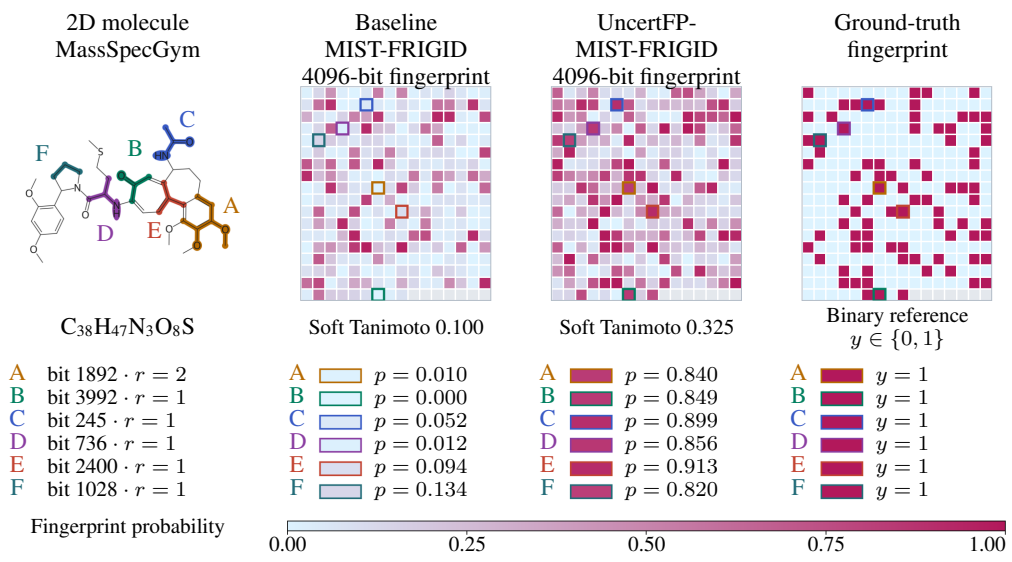

Figure 14: Recovery of 6 positive fingerprint bits.

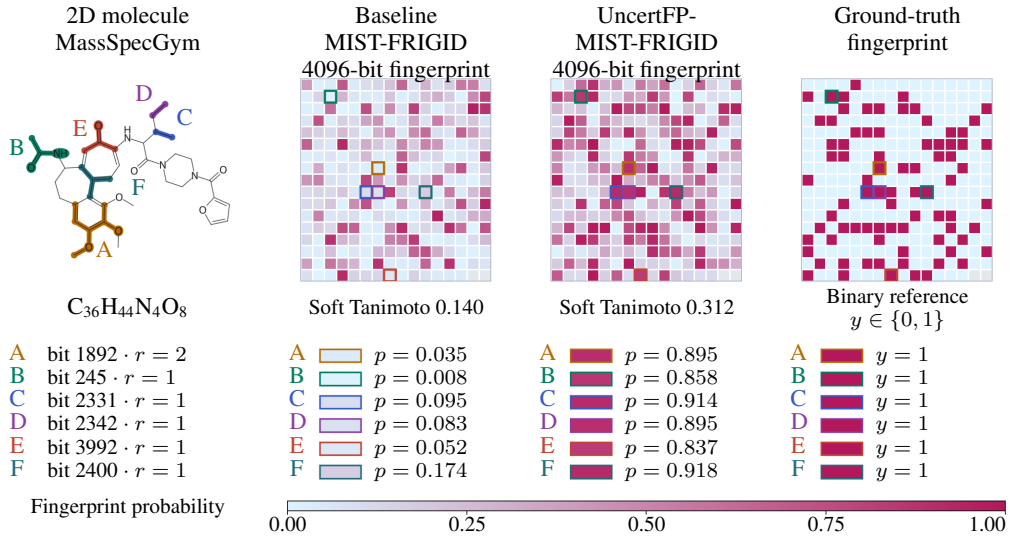

Figure 15: Recovery of 6 positive fingerprint bits.

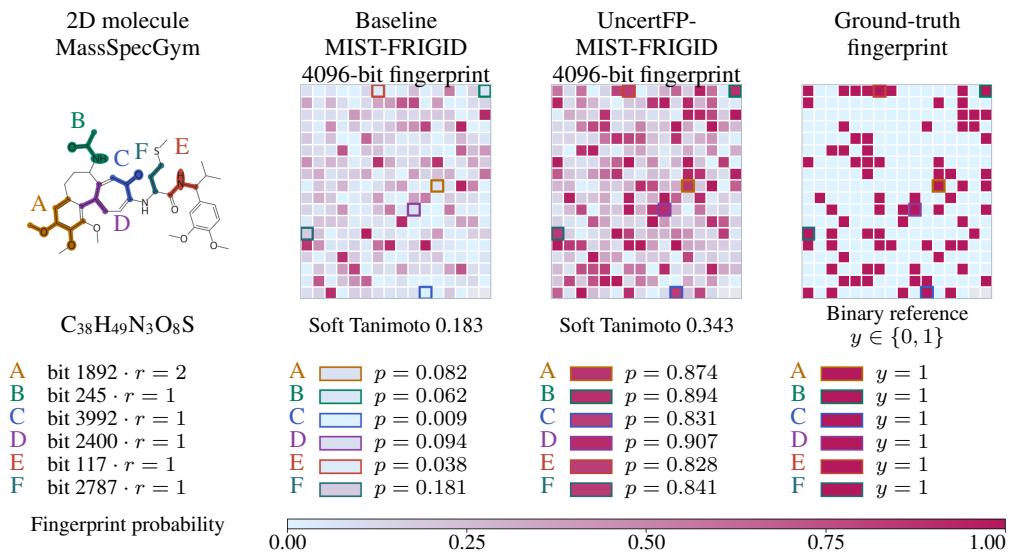

Figure 16: Recovery of 6 positive fingerprint bits.

I.3 SpectraVerse-positive

To extend the main-text analysis within the same dataset, we inspect additional molecules with different local environments. The representative example below reports the quantitative change, while the remaining panels provide qualitative structural context for the recovered bits. In Figure 63, UncertFP recovers six annotated positive bits, while whole-fingerprint Soft Tanimoto increases from 0.103 to 0.223. The accompanying structures and bit annotations connect these changes to the corresponding local environments, extending the case analysis beyond the two examples in the main text.

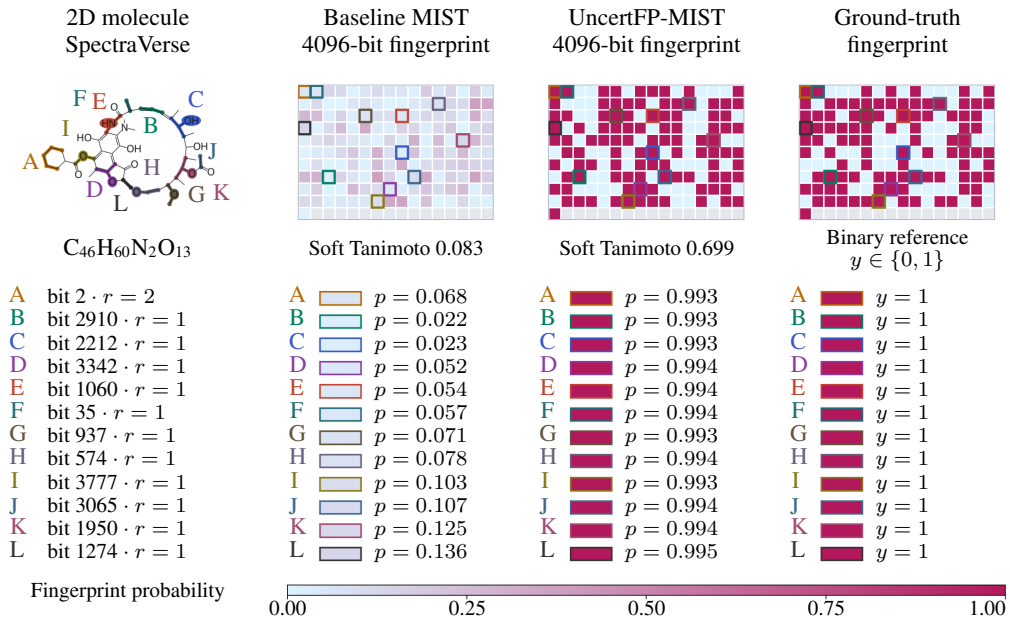

Figure 17: Recovery of 12 positive fingerprint bits.

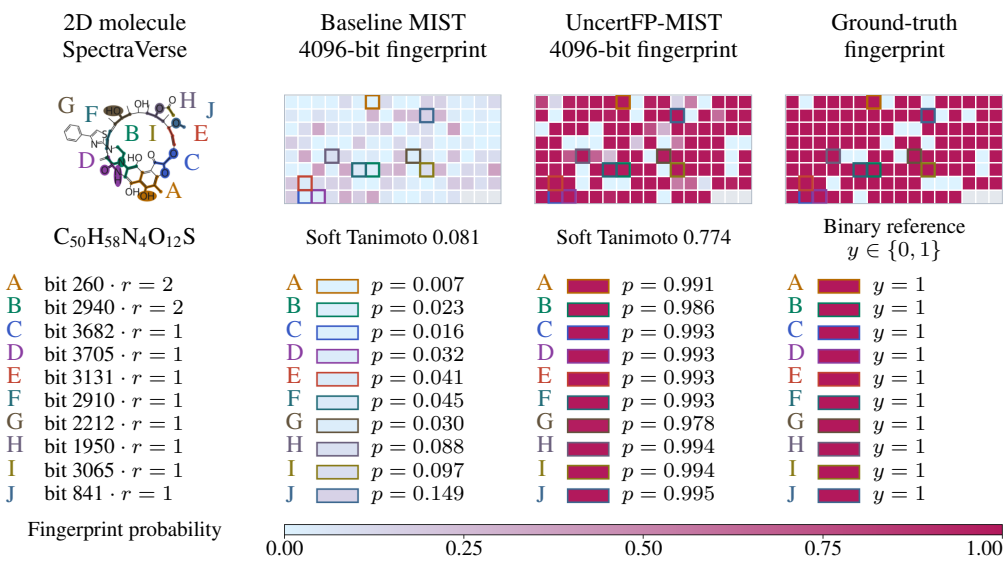

Figure 18: Recovery of 10 positive fingerprint bits.

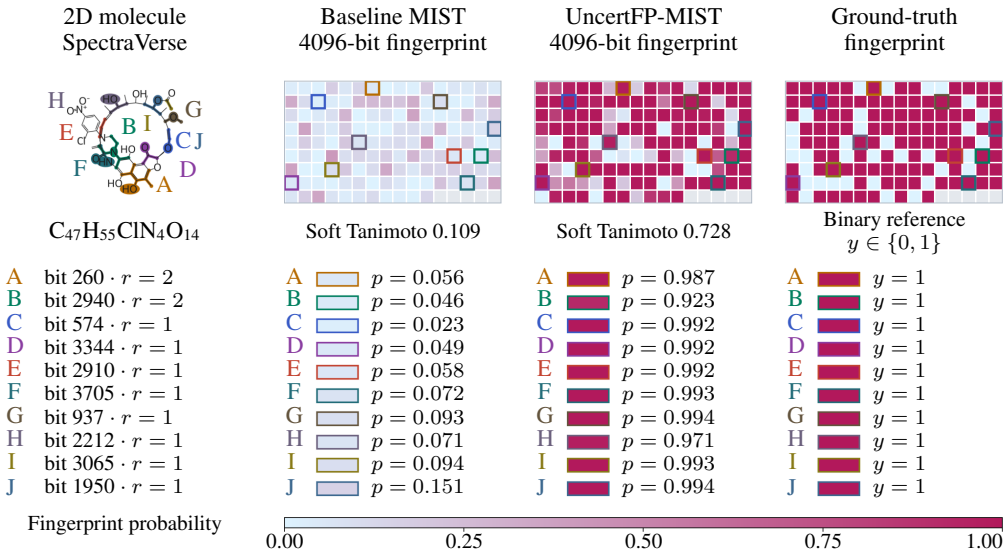

Figure 19: Recovery of 10 positive fingerprint bits.

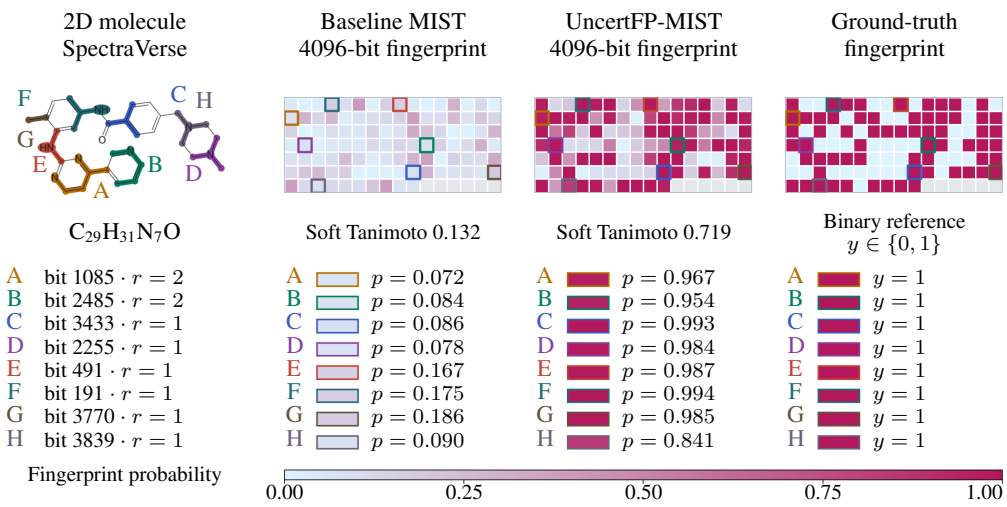

Figure 20: Recovery of 8 positive fingerprint bits.

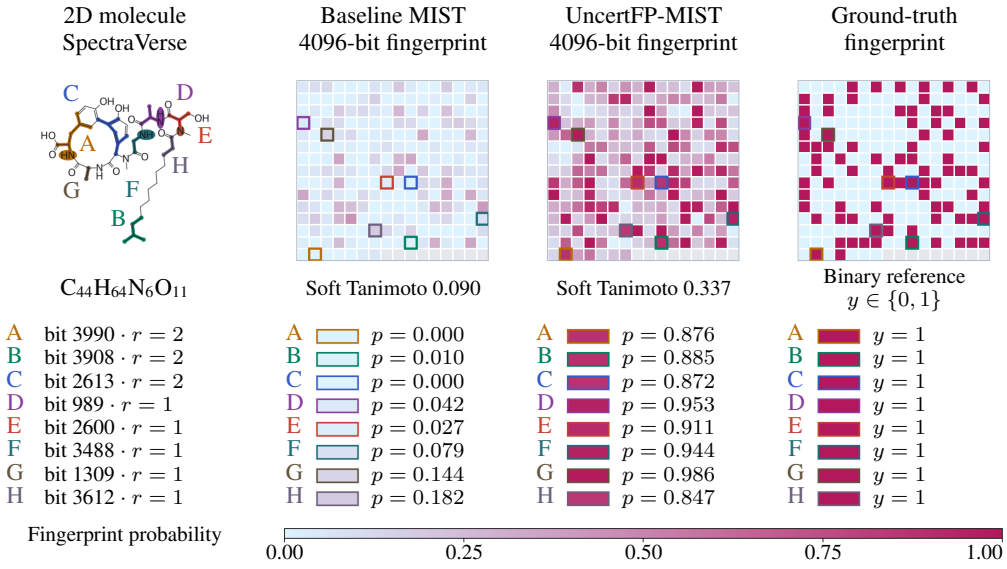

Figure 21: Recovery of 8 positive fingerprint bits.

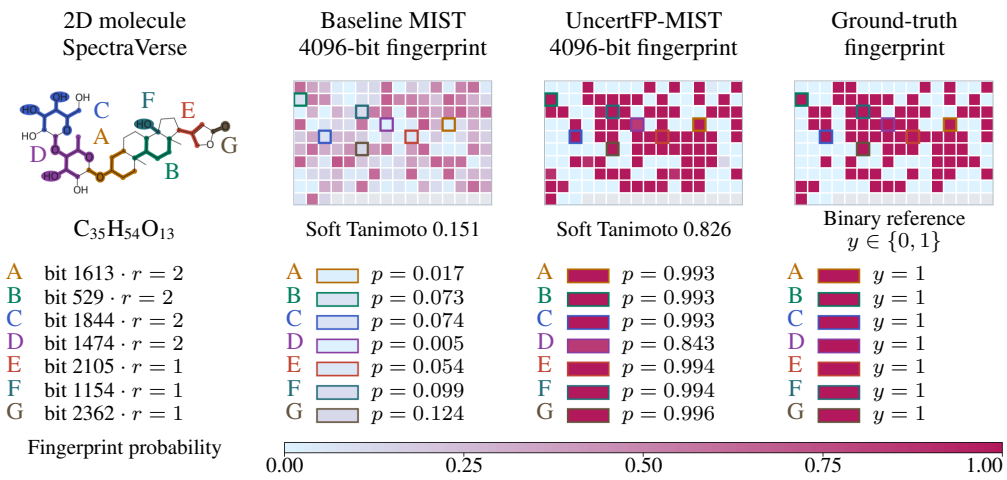

Figure 22: Recovery of 7 positive fingerprint bits.

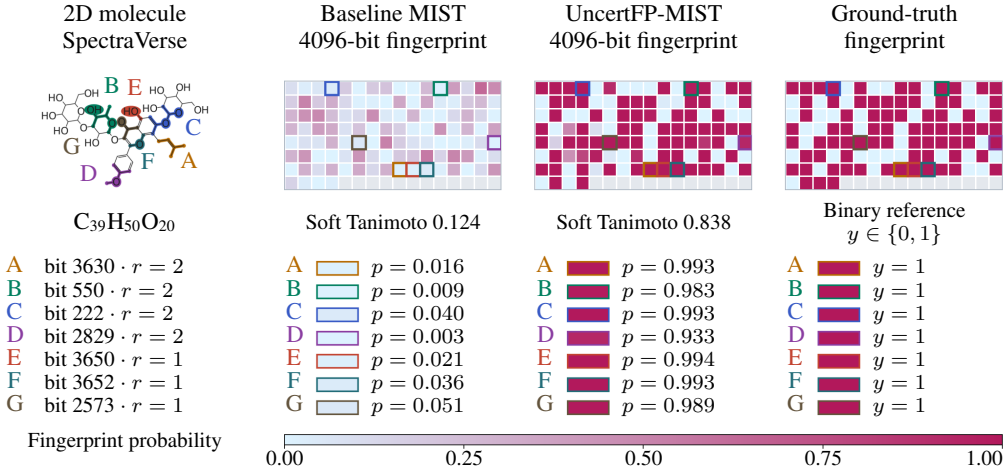

Figure 23: Recovery of 7 positive fingerprint bits.

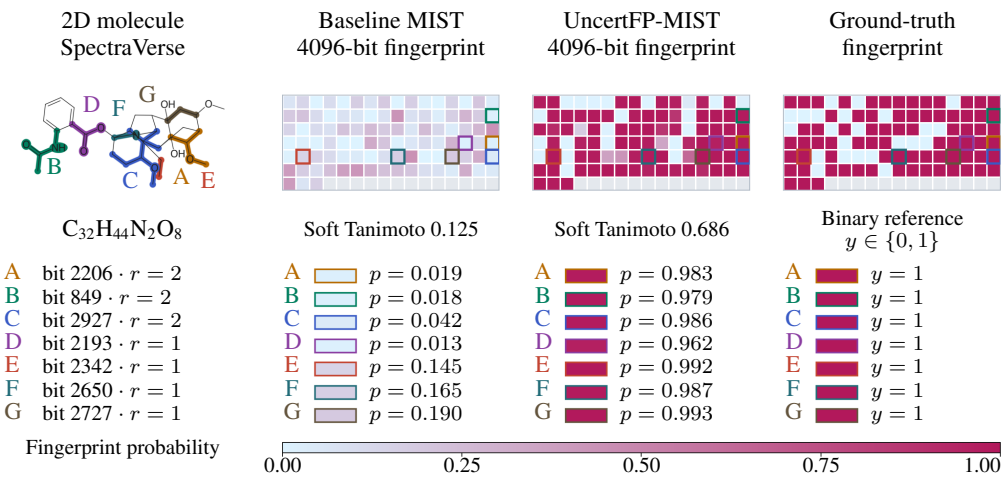

Figure 24: Recovery of 7 positive fingerprint bits.

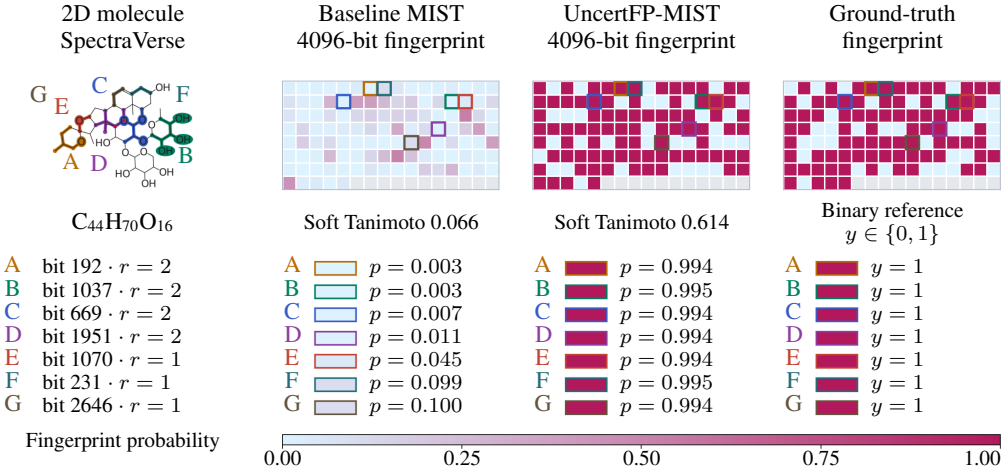

Figure 25: Recovery of 7 positive fingerprint bits.

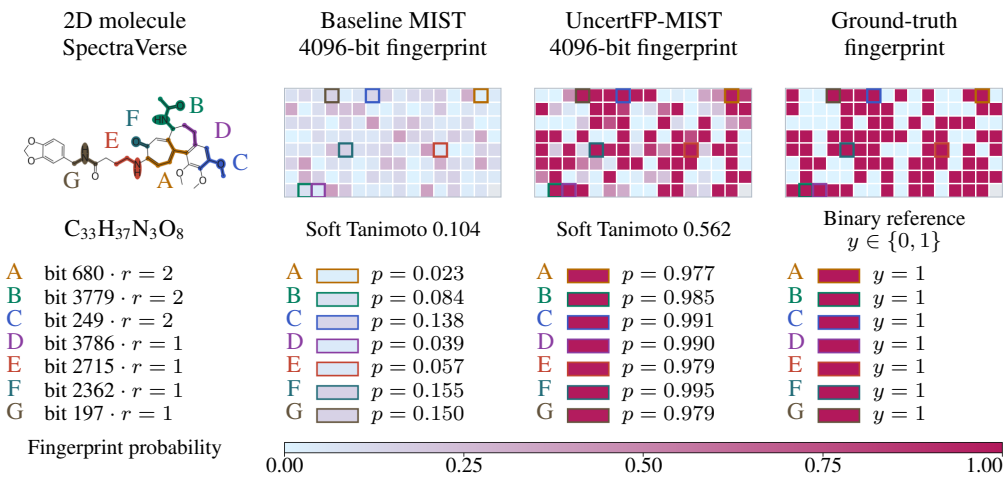

Figure 26: Recovery of 7 positive fingerprint bits.

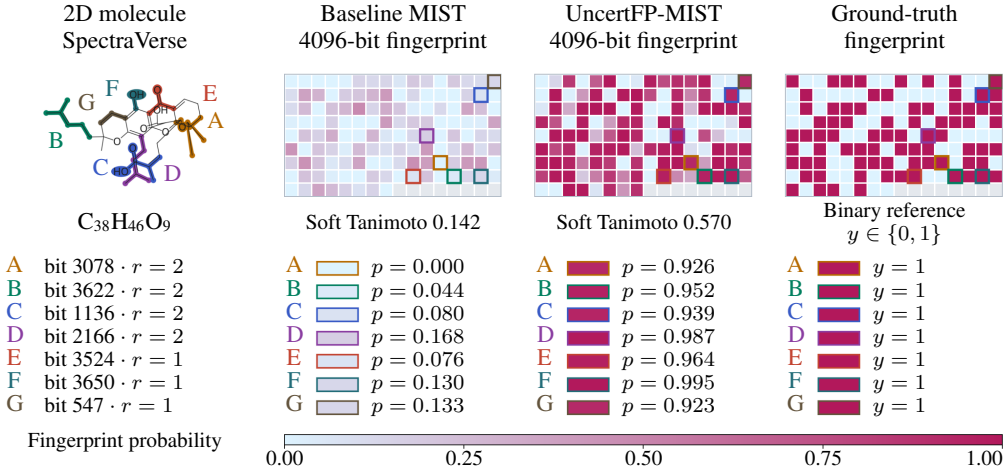

Figure 27: Recovery of 7 positive fingerprint bits.

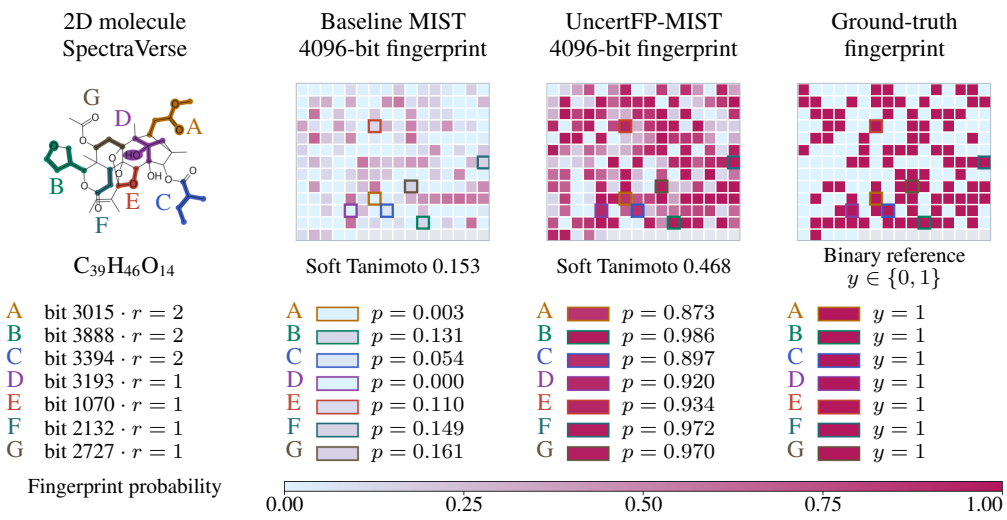

Figure 28: Recovery of 7 positive fingerprint bits.

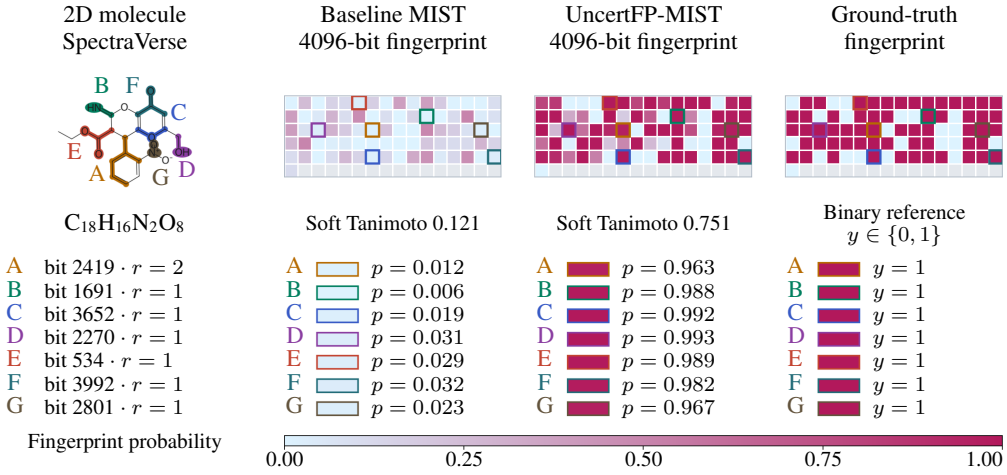

Figure 29: Recovery of 7 positive fingerprint bits.

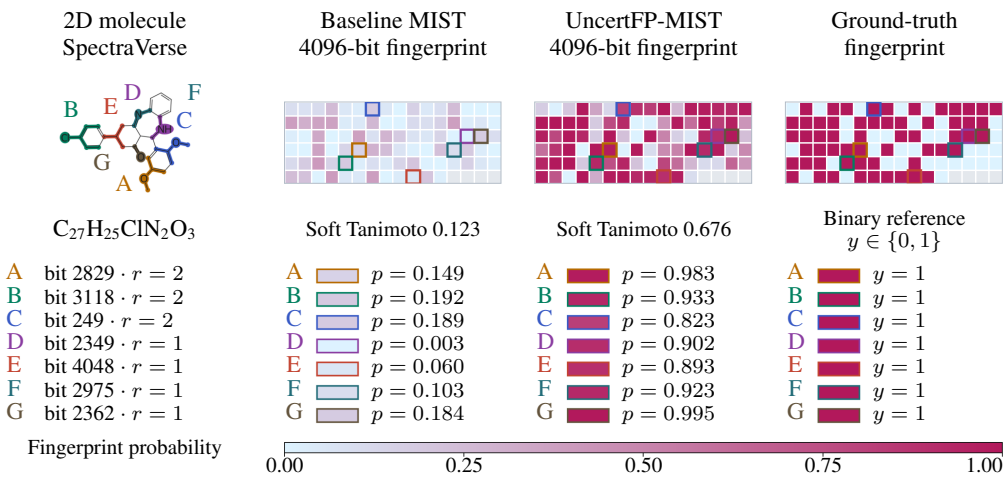

Figure 30: Recovery of 7 positive fingerprint bits.

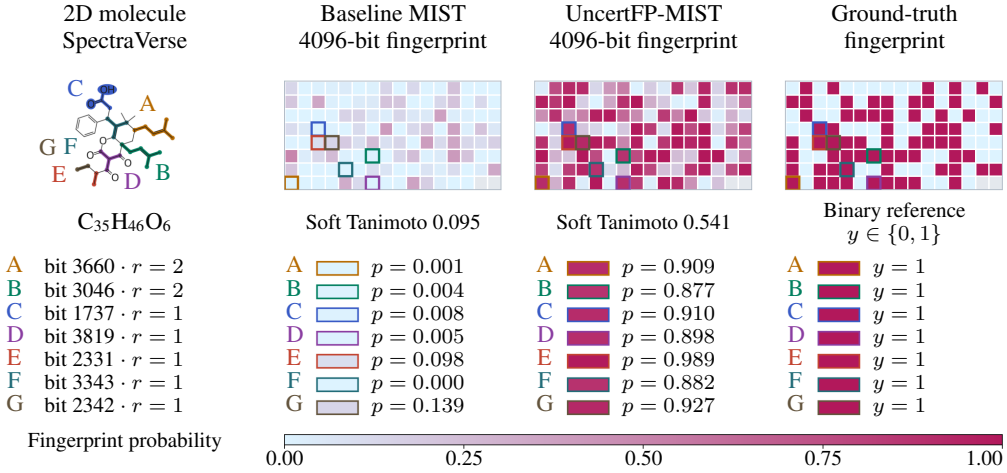

Figure 31: Recovery of 7 positive fingerprint bits.

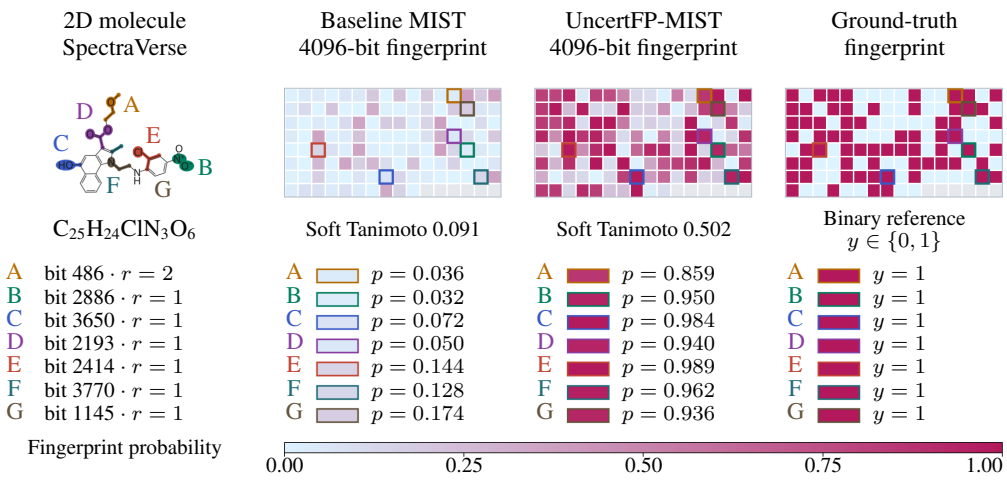

Figure 32: Recovery of 7 positive fingerprint bits.

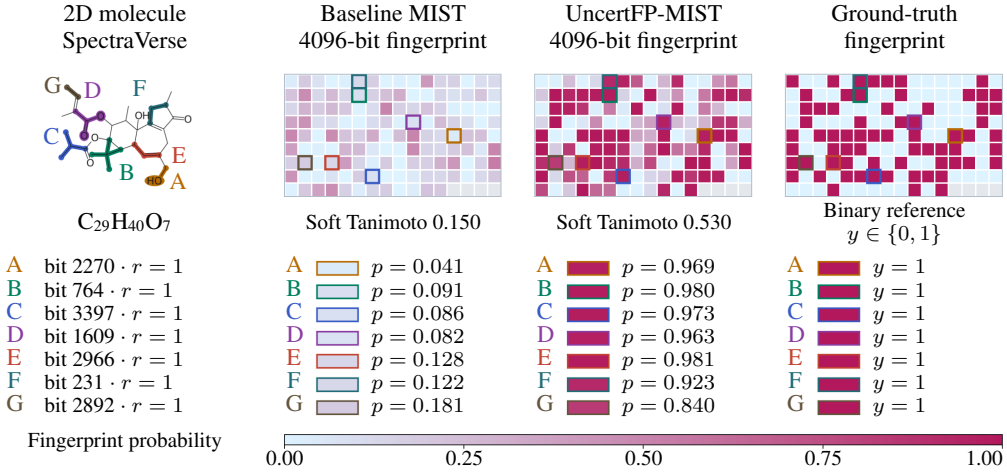

Figure 33: Recovery of 7 positive fingerprint bits.

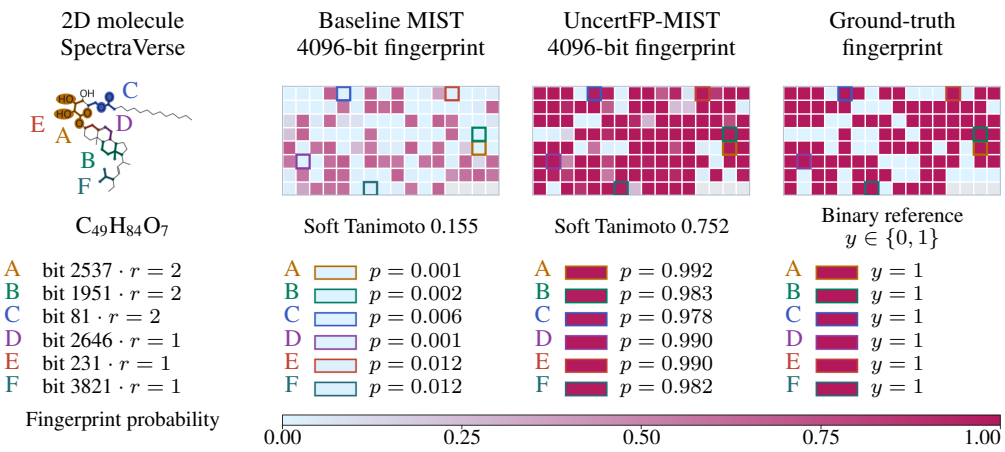

Figure 34: Recovery of 6 positive fingerprint bits.

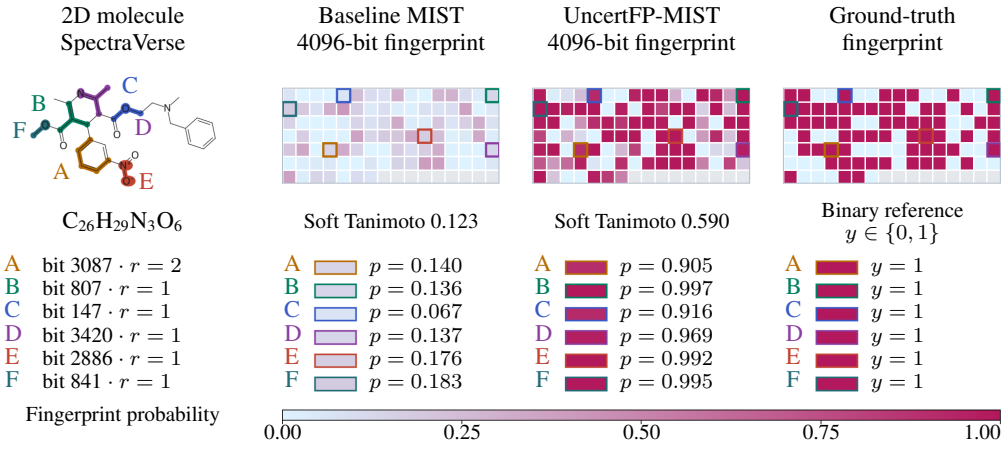

Figure 35: Recovery of 6 positive fingerprint bits.

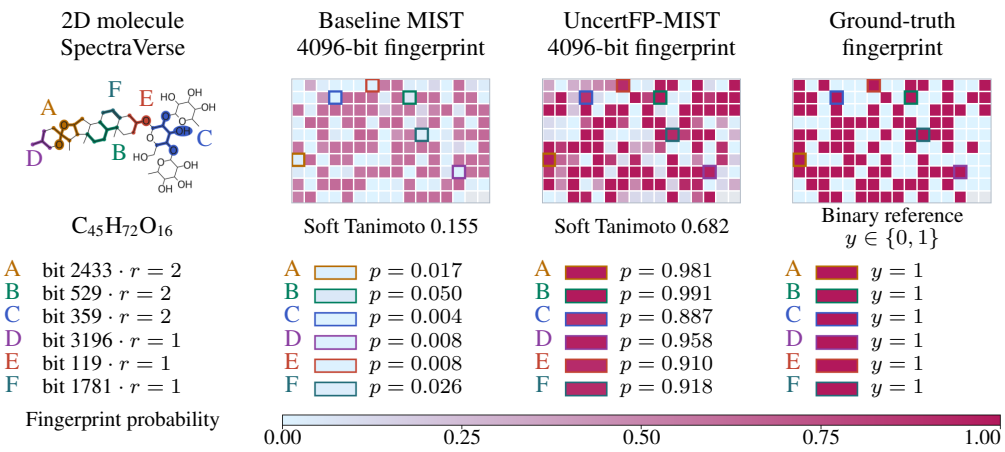

Figure 36: Recovery of 6 positive fingerprint bits.

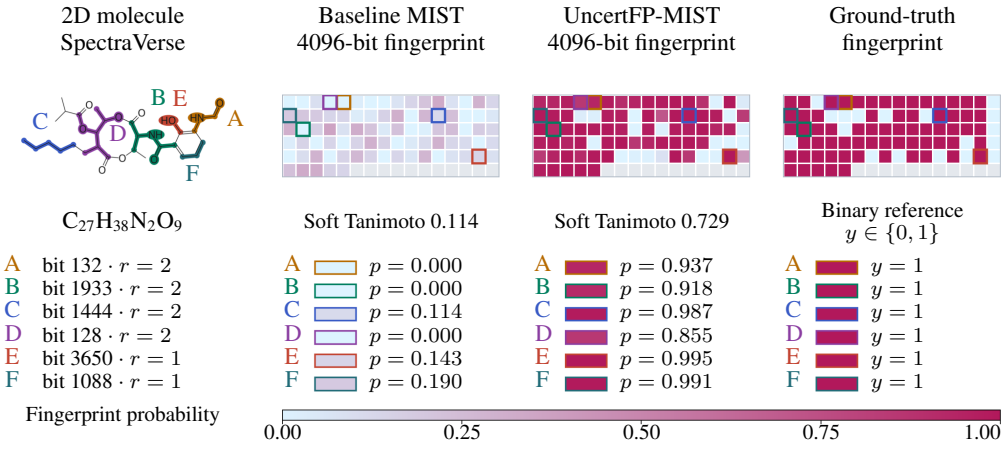

Figure 37: Recovery of 6 positive fingerprint bits.

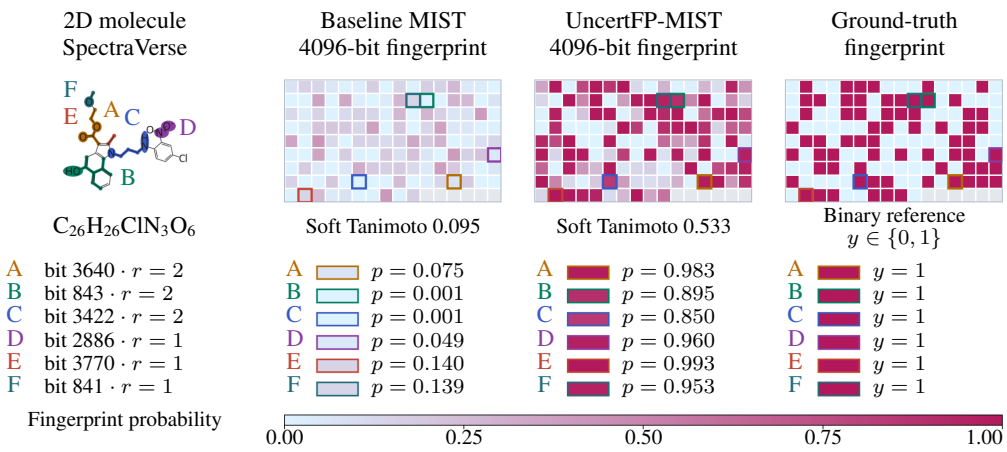

Figure 38: Recovery of 6 positive fingerprint bits.

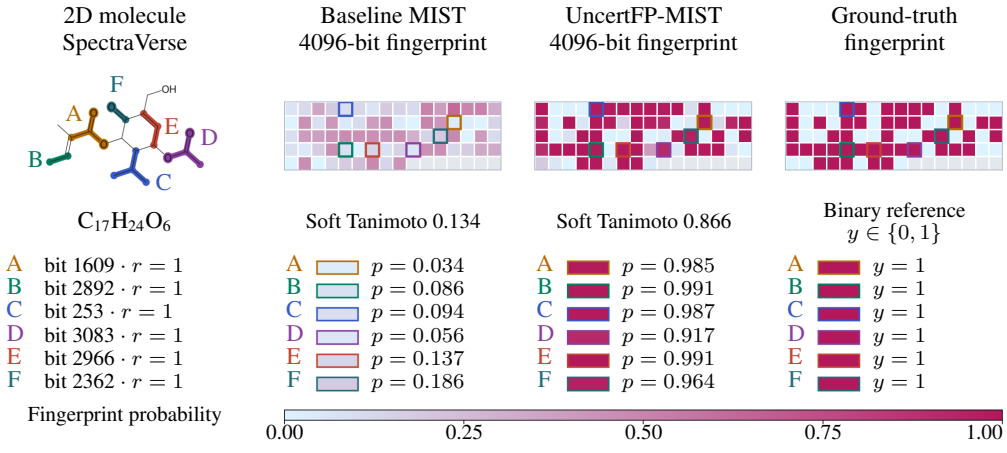

Figure 39: Recovery of 6 positive fingerprint bits.

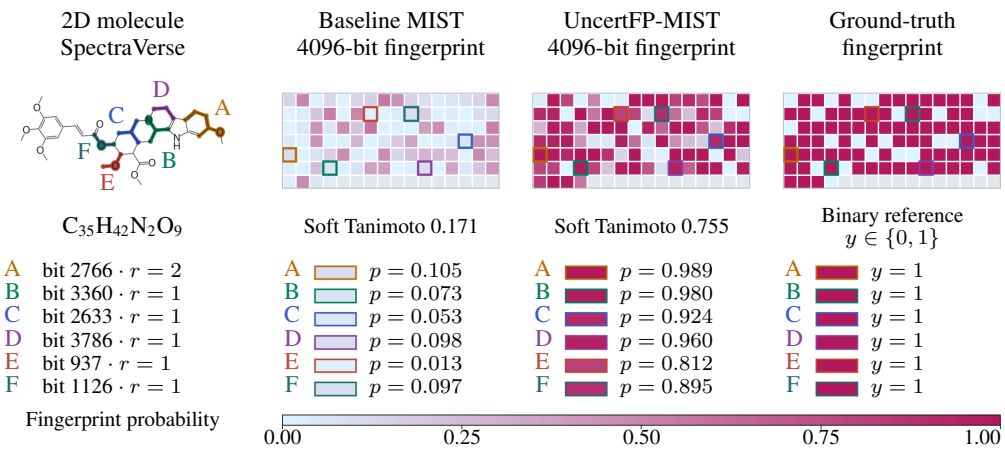

Figure 40: Recovery of 6 positive fingerprint bits.

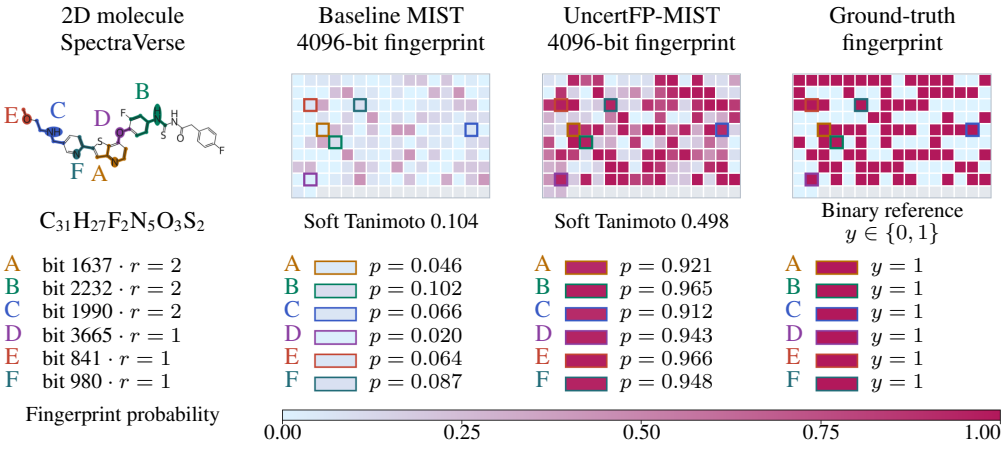

Figure 41: Recovery of 6 positive fingerprint bits.

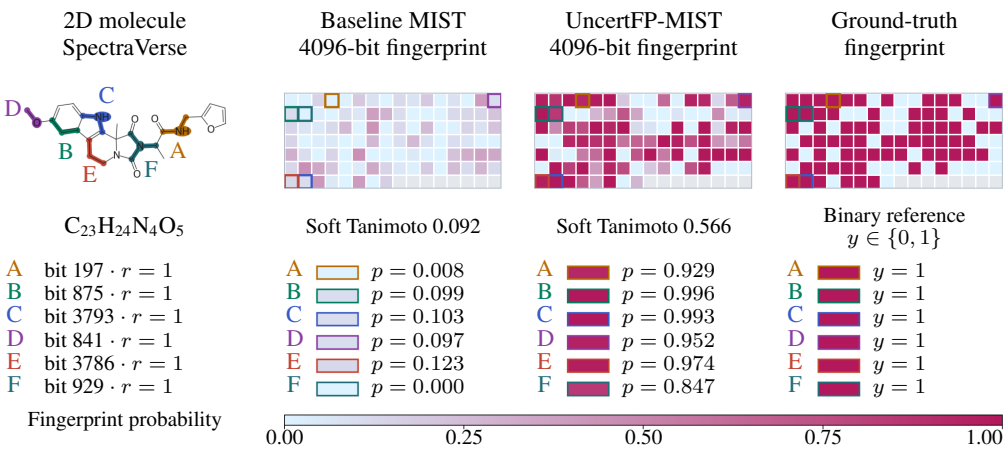

Figure 42: Recovery of 6 positive fingerprint bits.

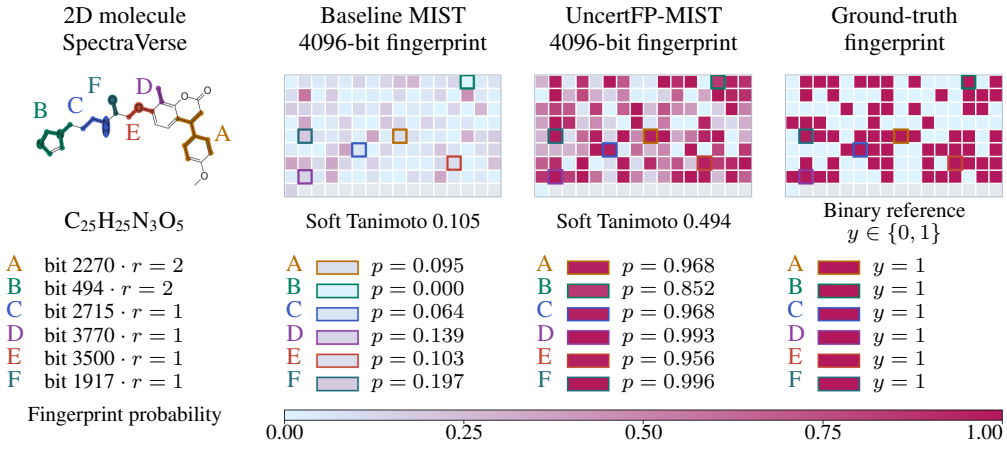

Figure 43: Recovery of 6 positive fingerprint bits.

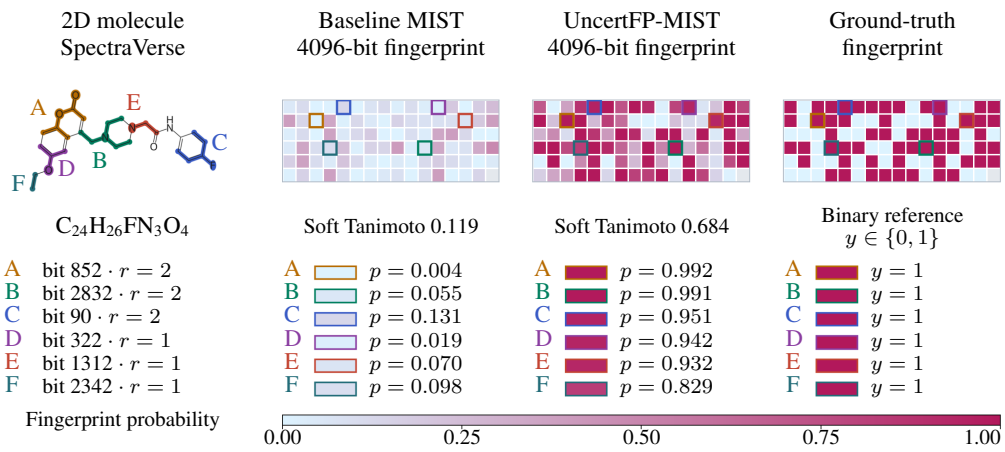

Figure 44: Recovery of 6 positive fingerprint bits.

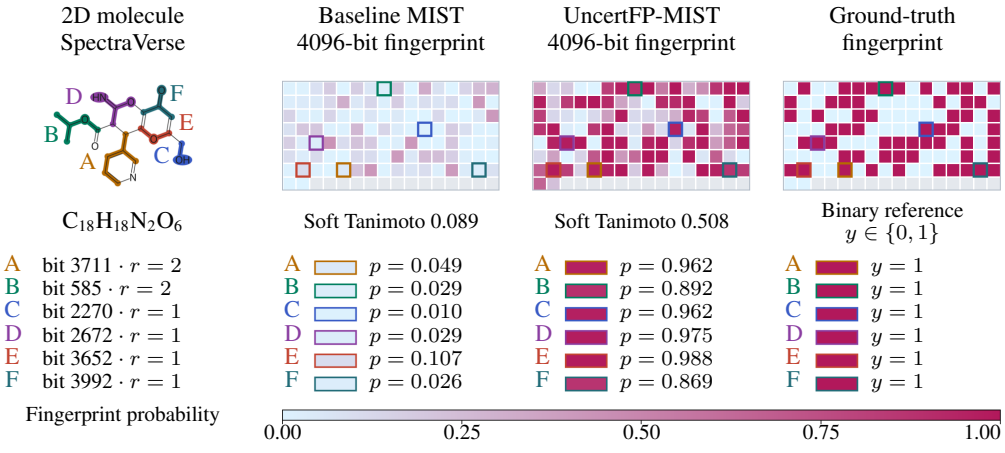

Figure 45: Recovery of 6 positive fingerprint bits.

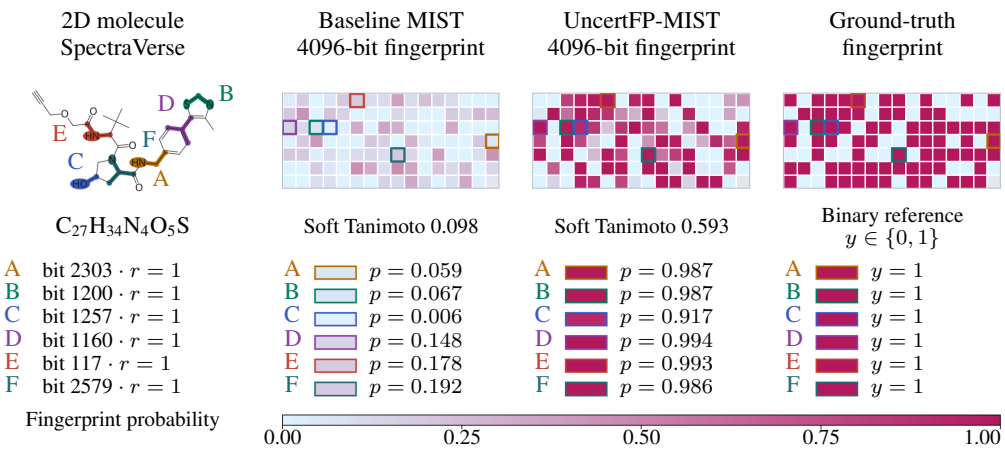

Figure 46: Recovery of 6 positive fingerprint bits.

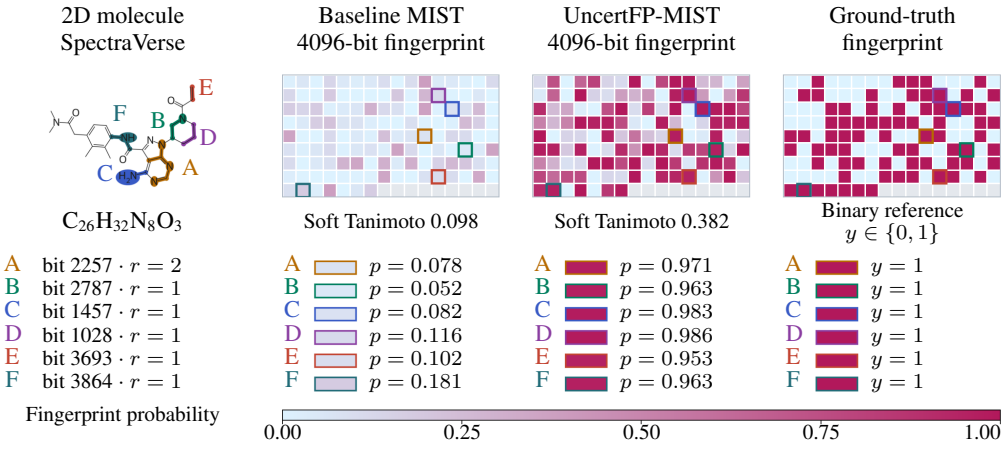

Figure 47: Recovery of 6 positive fingerprint bits.

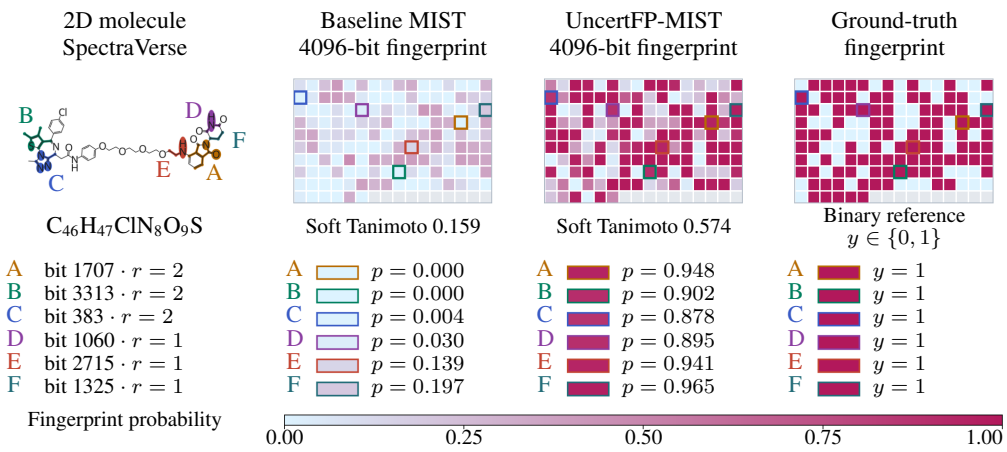

Figure 48: Recovery of 6 positive fingerprint bits.

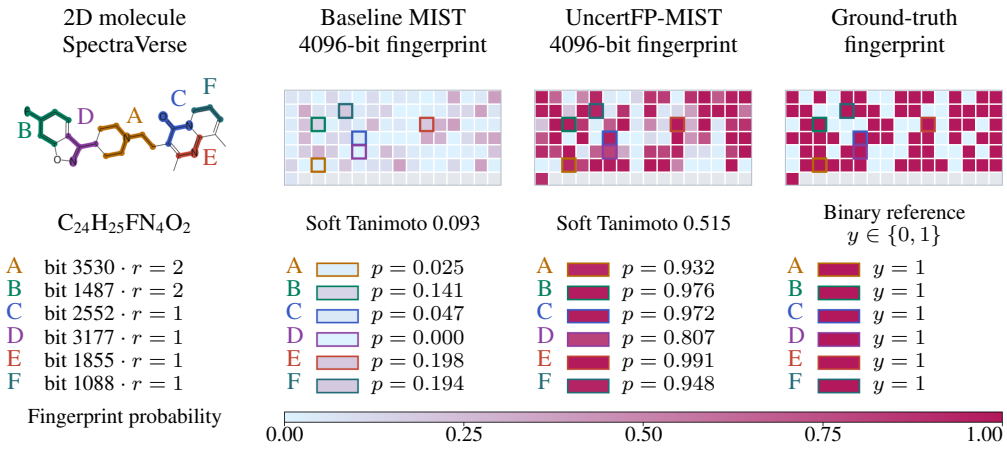

Figure 49: Recovery of 6 positive fingerprint bits.

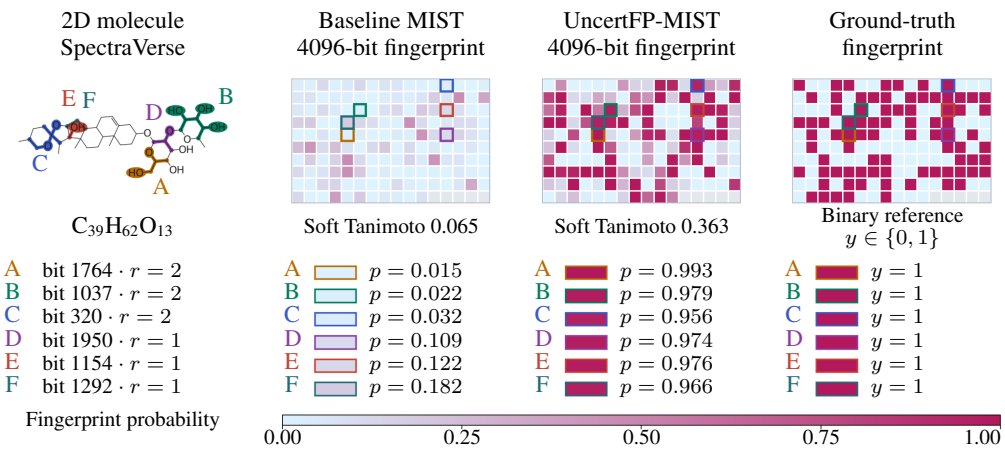

Figure 50: Recovery of 6 positive fingerprint bits.

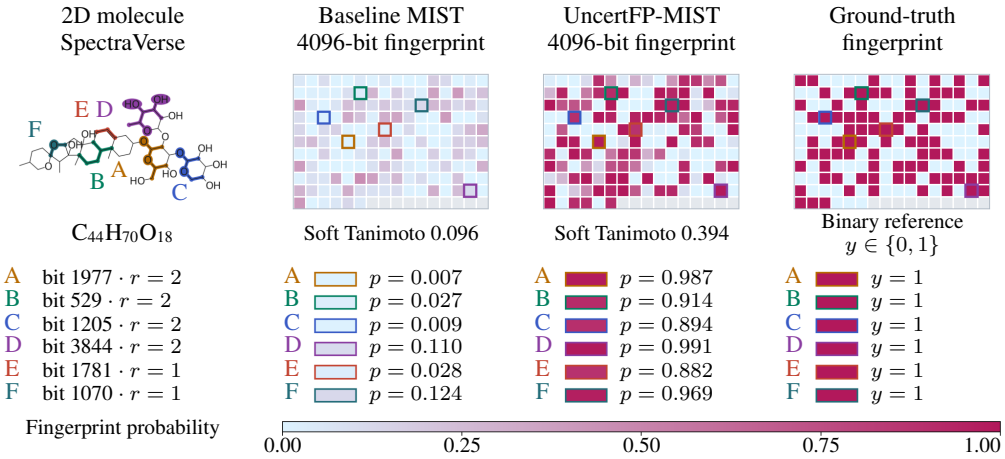

Figure 51: Recovery of 6 positive fingerprint bits.

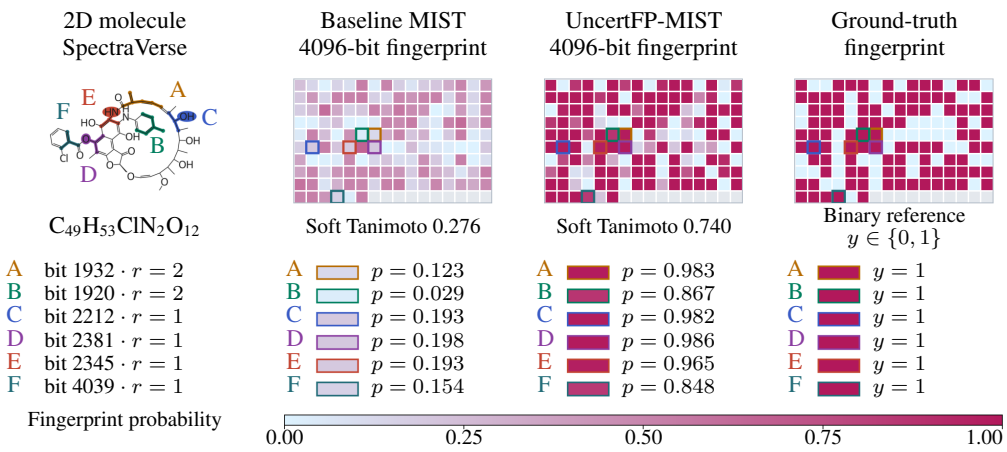

Figure 52: Recovery of 6 positive fingerprint bits.

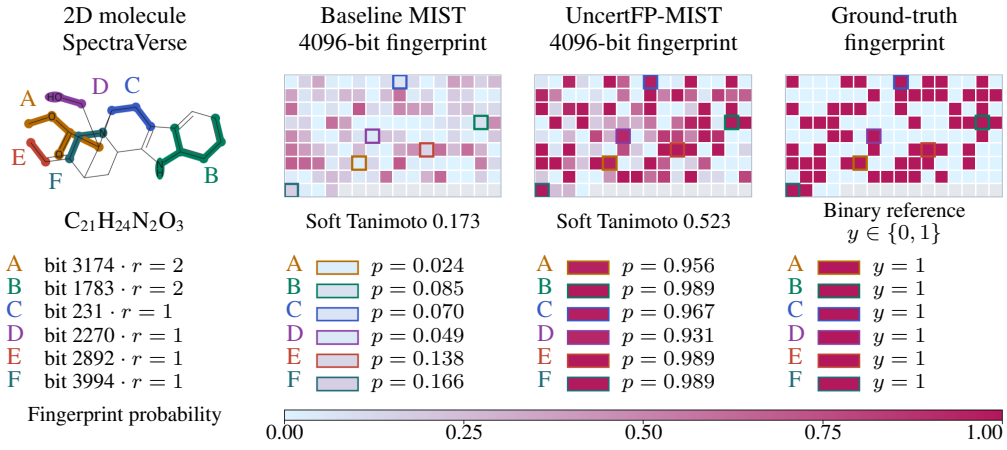

Figure 53: Recovery of 6 positive fingerprint bits.

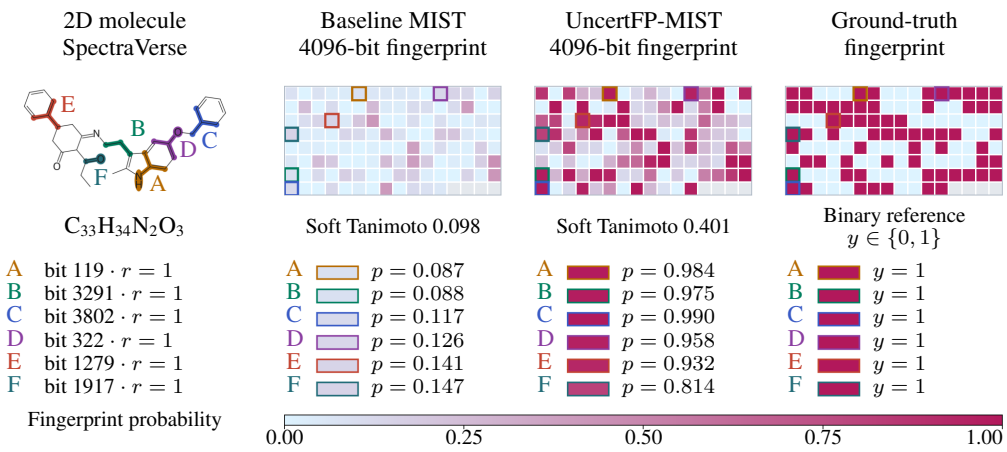

Figure 54: Recovery of 6 positive fingerprint bits.

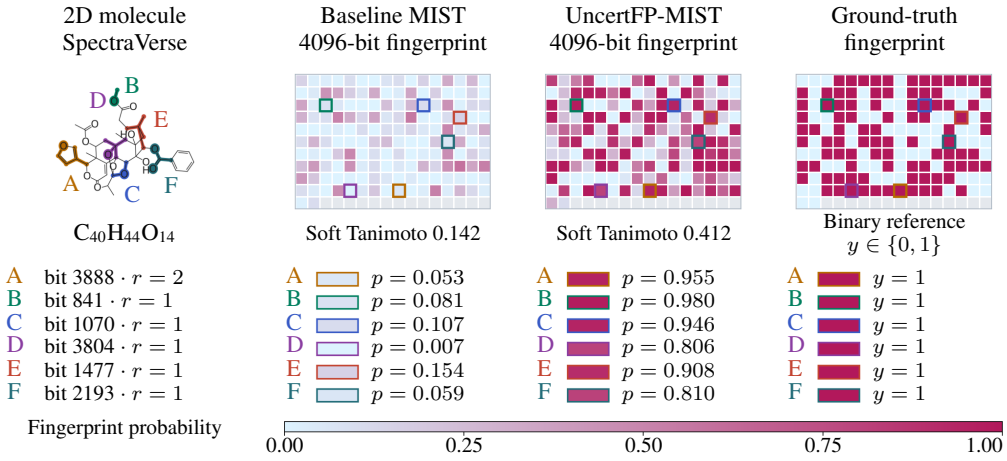

Figure 55: Recovery of 6 positive fingerprint bits.

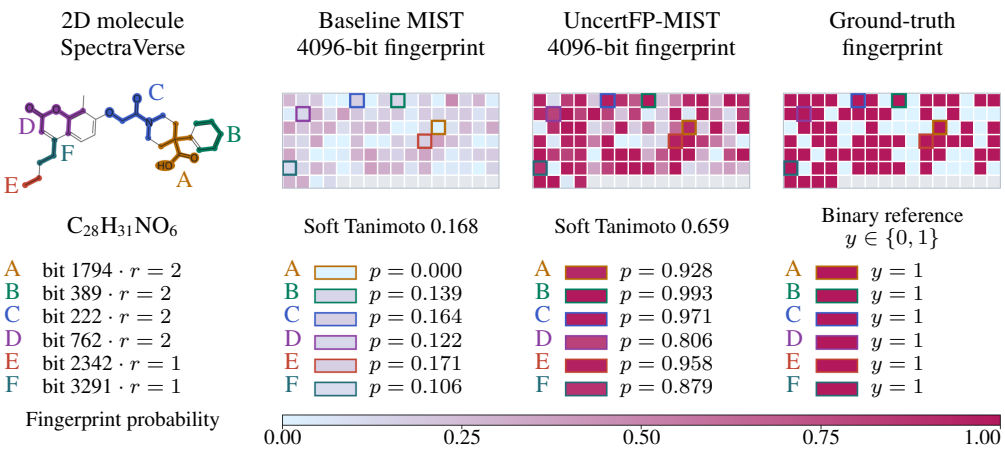

Figure 56: Recovery of 6 positive fingerprint bits.

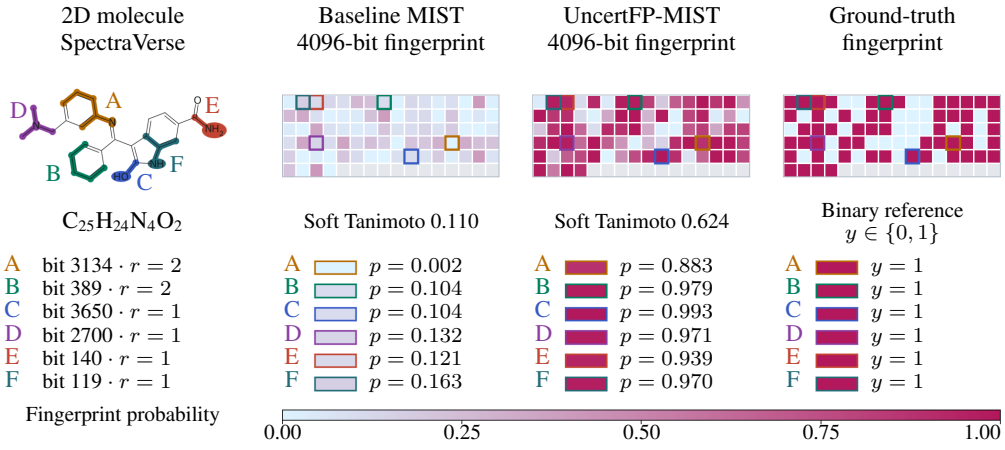

Figure 57: Recovery of 6 positive fingerprint bits.

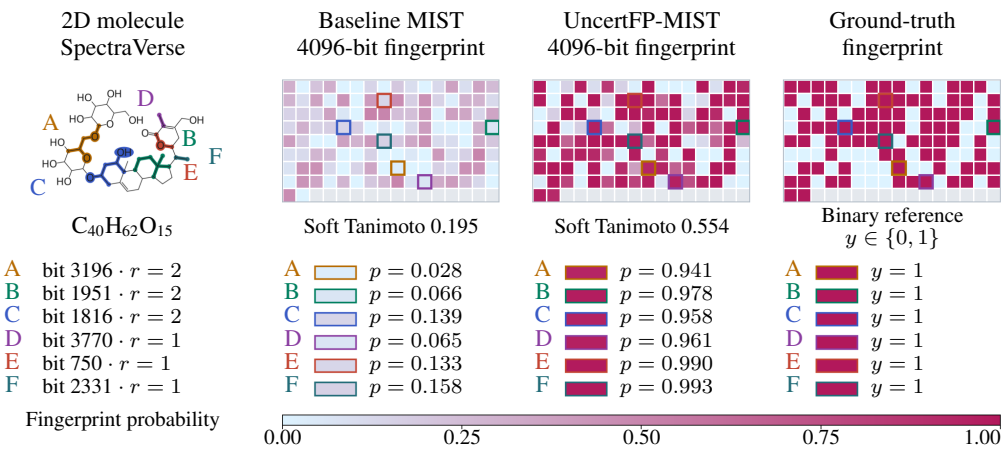

Figure 58: Recovery of 6 positive fingerprint bits.

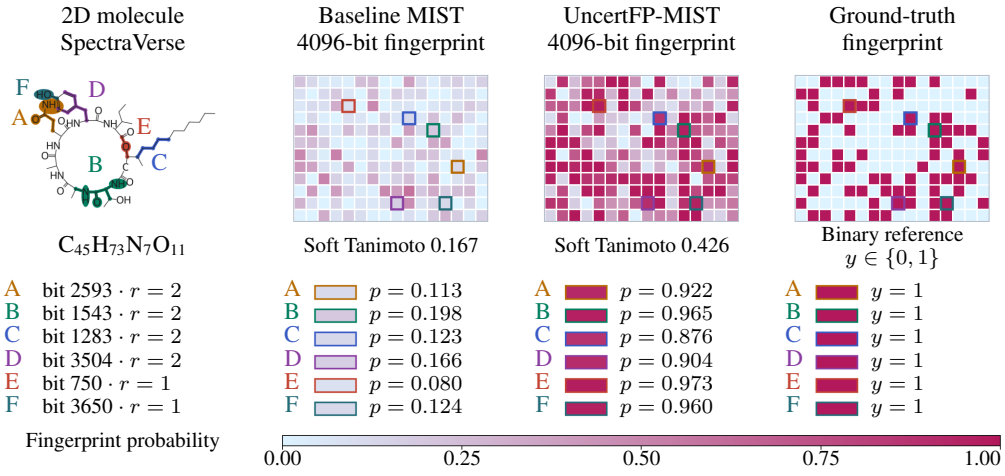

Figure 59: Recovery of 6 positive fingerprint bits.

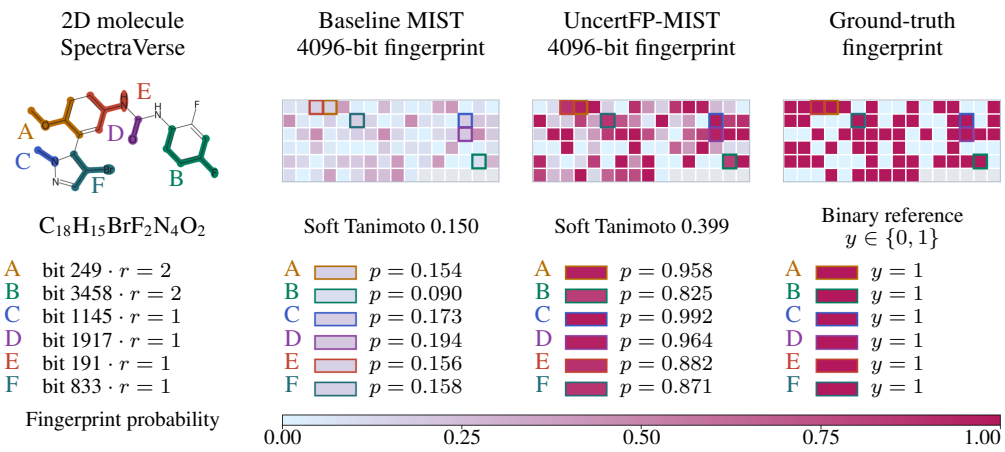

Figure 60: Recovery of 6 positive fingerprint bits.

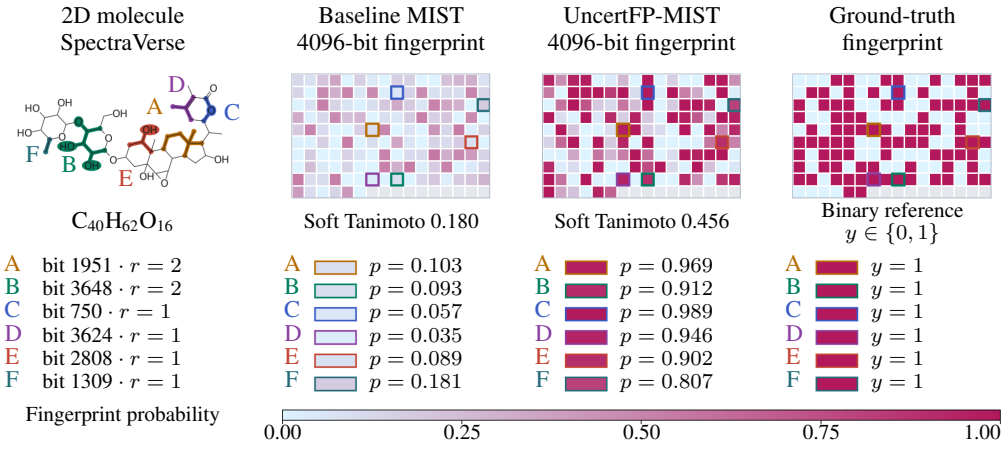

Figure 61: Recovery of 6 positive fingerprint bits.

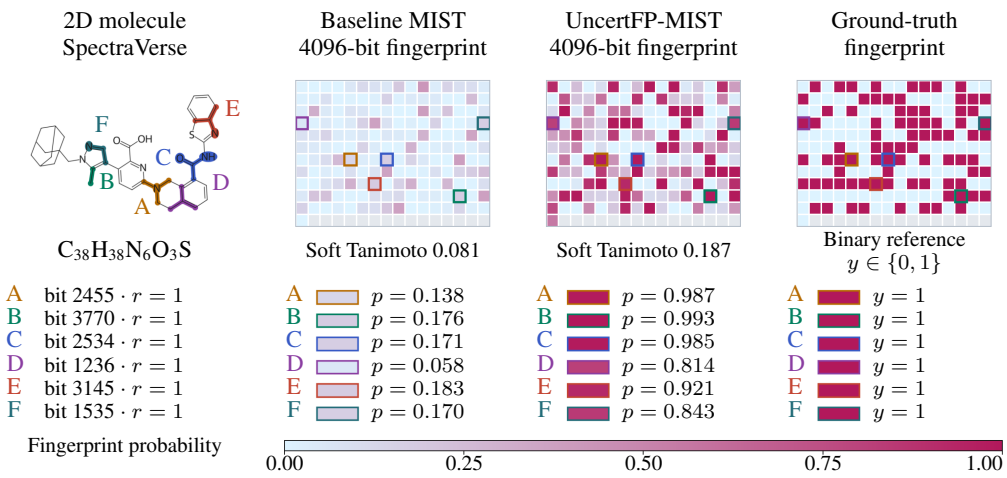

Figure 62: Recovery of 6 positive fingerprint bits.

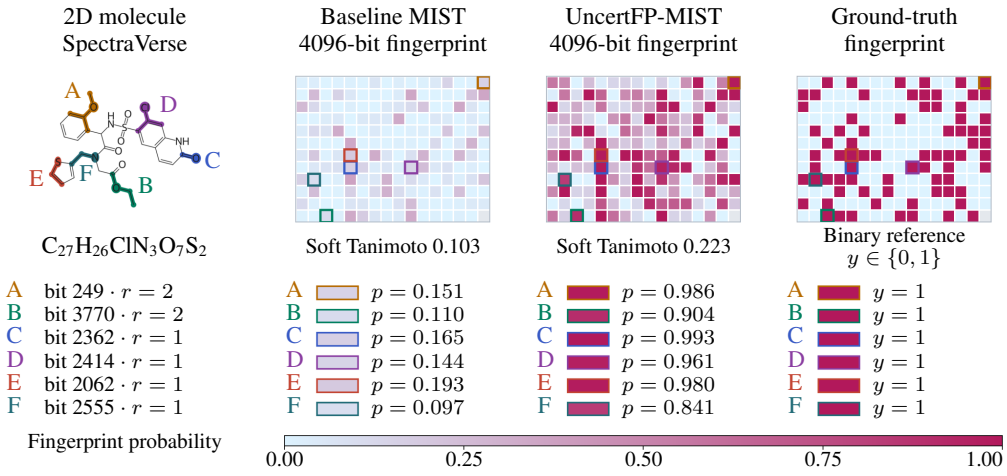

Figure 63: Recovery of 6 positive fingerprint bits.