ALIGNING MASS SPECTRA WITH MOLECULAR STRUCTURE FOR FAST COMPUTATIONAL OLFACTION

Ziqi Zhang[1] **Eunyeong Jin**[2] **Miguel Vasco**[1] **Farzaneh Taleb**[1] **Nona Rajabi**[1]
Alexandra Gutmann[2] **Jonathan Williams**[2] **Antônio H. Ribeiro**[3,4] **Danica Kragic**[1]

[1]Dept. of Intelligent Systems, KTH Royal Institute of Technology, Stockholm, Sweden
[2]Atmospheric Chemistry Dept., Max Planck Institute for Chemistry, Mainz, Germany
[3]Dept. of Information Technology, Uppsala University, Uppsala, Sweden
[4]Science for Life Laboratory (ScilifeLab), Uppsala, Sweden

{ziqizh, miguelsv, farzantn, nonar, dani}@kth.se
{eunyeong.jin, a.gutmann, jonathan.williams}@mpic.de
antonio.horta.ribeiro@it.uu.se

ABSTRACT

Understanding human olfaction from a computational point of view is an under-explored frontier within the machine learning community. Previous works leverage the chemical structure of odorant molecules for olfactory prediction tasks. However, obtaining the structure of unknown odorants requires significant time- and labor-intensive techniques, limiting their suitability for fast computational olfaction. In this work, we explore the use of direct electron ionization mass spectrometry (direct EI-MS), a fast sensing technique (in the order of seconds), for olfactory prediction tasks. We contribute Spectrum-to-Chemical Embedding alignmeNT (SCENT), a multi-modal contrastive learning framework to align mass spectra with explicit chemical information. In particular, we augment the mass spectrum representation with a chemical structure prior during training, requiring only mass spectrum information at test time. Our approach performs on par with state-of-the-art approaches that require chemical information explicitly during inference, and significantly better than previous mass-spectrum-only baselines.

1 INTRODUCTION

Human olfaction remains one of the least explored perceptual modalities in artificial intelligence (AI) research, and its computational modeling still faces many open challenges. In recent years, increasing attention has been devoted particularly to the structure–odor relationship, demonstrating that learned chemical structure representations can capture structure-odor mappings and align closely with human olfactory perception (Keller et al., 2017; Lee et al., 2023; Taleb et al., 2024). However, explicit chemical information is often unavailable in real-world scenarios, where odor perception must be inferred from sensor-derived signals instead.

Gas chromatography coupled with electron ionization mass spectrometry (GC-EI-MS) remains the gold standard among available technologies for linking specific chemical components to olfactory percepts. The chromatographic separation ensures the acquisition of a pure spectrum for definitive structural identification (Zhu et al., 2021). However, this process is extremely time-consuming and requires expertise, limiting its use to a few laboratory settings. In contrast, direct electron ionization mass spectrometry with a single quadrupole (direct EI-MS, hereafter referred to as MS) omits chromatographic separation while retaining structure-rich fragmentation fingerprints, enabling rapid acquisition at the cost of overlapping molecular signals. Such signal superposition creates structural ambiguity, necessitating a model that can discern the chemical semantics underlying each spectral peak to accurately infer molecular properties. To the best of our knowledge, the existing end-to-end approaches that predict odor labels directly from MS fail to account for the intrinsic chemical structure of the spectrum (Ito & Nakamoto, 2020; Hasebe et al., 2022; Debnath & Nakamoto, 2022).

Inspired by contrastive pretraining paradigms in vision and language (Chen et al., 2020; Radford et al., 2021; Zhai et al., 2022), cross-modal contrastive learning offers a principled framework for

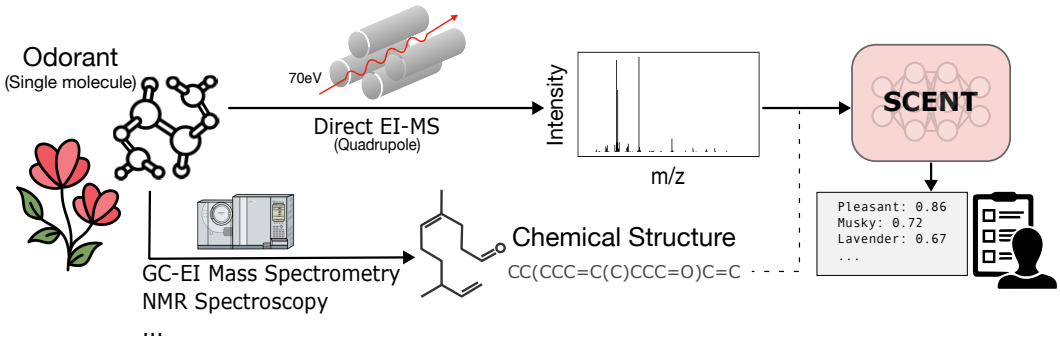


Figure 1: **Overview of Spectrum-to-Chemical Embedding Alignment (SCENT)**. We explore the prediction of olfactory perception directly from MS of single odorant molecules. We propose to use explicit chemical structure information, pertaining to the odorant molecule, to pretrain our representation space and to predict human olfactory percepts directly from MS data.

aligning heterogeneous representations of different modalities, which is an ideal access for bridging the gap between accessible signals and interpretable structures. In this work, we contribute SCENT (Spectrum-to-Chemical Embedding alignmeNT), a multi-modal contrastive learning framework for aligning MS and chemical structure representations (Figure 1). Specifically, SCENT aligns MS embeddings, which reflect the relative contributions of latent molecular substructures, with chemical structure embeddings that explicitly encode substructure identity, yielding a joint representation space. While chemical structure information is used to guide representation learning, the resulting model requires only MS data at inference time. Using odor perception label prediction as a downstream probe, we show that our aligned MS representations significantly outperform MS-only baselines and achieve performance comparable to methods that require explicit chemical structure information at test time. These results provide evidence that cross-modal alignment enables sensor-derived embeddings to serve as reliable surrogates for symbolic chemical representations and affords chemically interpretable reasoning when explicit structure information is unavailable.

2 METHOD

We introduce *Spectrum-to-Chemical Embedding alignmeNT* (SCENT), a multi-modal contrastive learning framework for aligning MS and chemical structure information pertaining to single-molecule odorants and predict human olfactory descriptors. We define two different stages of our framework, as shown in Figure 2.

Modality Alignment In the first stage, we use a learnable MS encoder and a frozen chemical structure encoder to project both modalities into a shared latent space (Figure 2(**A**)). Inspired by EIMS2Vec (Liu et al., 2024), we encode MS data by embedding peak co-occurrences and weighting the embeddings with power-scaled intensities ($p = 0.5$). Different from the original work, however, we remove the original sum pooling over embedding outputs and use a two-layer transformer encoder, with four self-attention heads each, as we found empirically to improve performance. The final MS embedding is extracted from a projection MLP applied over the CLS token.

To embed chemical structure data, we consider two different pre-trained encoders:

- **Open POM** (Barsainyan et al., 2023), the publicly available version of the POM (Lee et al., 2023) model, a message passing graph neural network trained with supervision to predict odor labels from chemical information. We extract from the model a 256-dim embedding vector for a single odorant.
- **MolFormer** (Ross et al., 2022), a pre-trained foundation model of chemical data, trained on over 1.1 billion molecules, which outputs a 768-dim embedding vector for each odorant.

Both models use SMILES (Simplified Molecular Input Line Entry System) (Weininger, 1988) data as input, which encodes molecules as a sequence of characters representing molecular structure and functional groups information. Finally, SCENT aligns the MS and chemical structure embeddings through a contrastive learning objective, inspired by CLIP (Radford et al., 2021). Formally, within

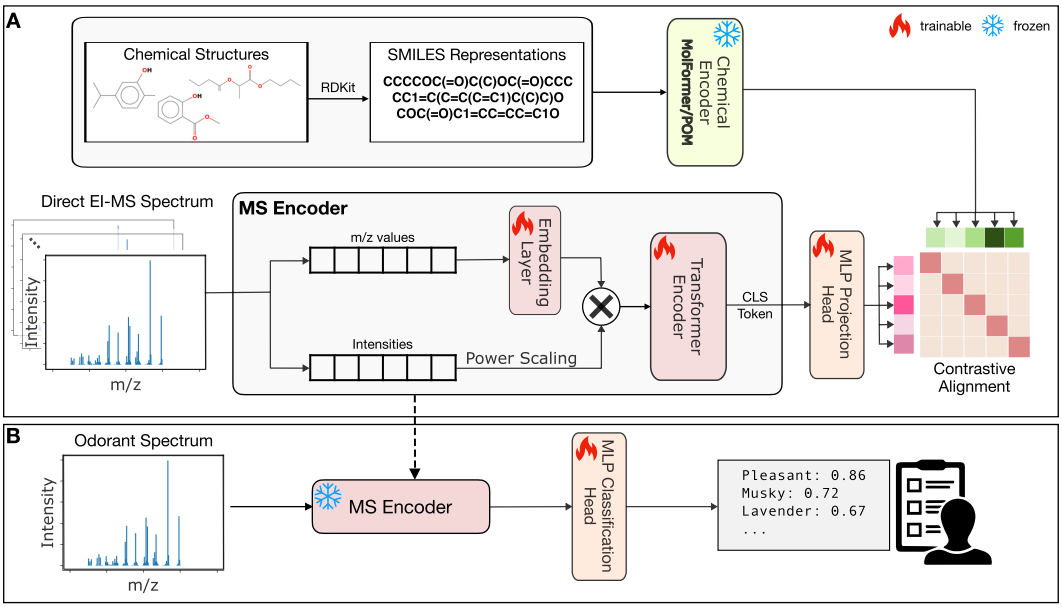


Figure 2: **The Spectrum-to-Chemical Embedding Alignment (SCENT) framework**: **(A)** In the first stage, we use a learnable MS encoder and a frozen chemical structure encoder to project both modalities into a shared latent space, forcing their alignment with a multimodal contrastive loss (Eq. 2); **(B)** On the second stage, we train a small classification head over the frozen MS encoder to directly perform multi-label odor descriptor classification directly from the spectrum.

a mini-batch of N pairs, we compute a similarity matrix using temperature-scaled cosine similarity, denoted as $s_{ij} = (u_i^T v_j)$, where τ is the temperature parameter. The training objective is to minimize the symmetric alignment loss:

$$\mathcal{L}_{\text{align}} = -\frac{1}{2N} \sum_{i=1}^{N} \left(\log \frac{e^{s_{ii}/\tau}}{\sum_{j=1}^{N} e^{s_{ij}/\tau}} + \log \frac{e^{s_{ii}/\tau}}{\sum_{j=1}^{N} e^{s_{ji}/\tau}} \right)$$

Odor Descriptor Prediction Finally, on the second stage of SCENT, we aim at using the learned representation to perform multi-label (independent) odor descriptor classification directly from the MS of odorants. To do so, we use the trained MS encoder (now frozen) and train a small MLP-based classification head, using a standard Binary Cross-Entropy (BCE) loss.

Training SCENT To train our proposed framework, we consider three datasets that provide spectral, structural, and perceptual information about odorants. To train the MS encoder and align MS embeddings with structure embeddings in the latent space we use:

- **NIST EI-MS library (2023)**: Starting with 347,100 compounds, we filtered for molecular weights between 50 and 300 Da to target volatile odorants and remove background noise. This resulted in 184,558 unique compound-spectrum pairs.
- **SMILES**: Corresponding SMILES strings are retrieved via PubChem and canonicalized using RDKit, with stereochemistry removed. This yielded a final paired dataset of 148,212 compounds for alignment.

Following Lee et al. (2023), to train the classification head we use the GS-LF dataset (GoodScent; Leffingwell & Associates, 2001), which contains SMILES representations and 138 odor descriptor labels associated with single-molecule odorants. Additionally, we filter the dataset and retain 2588 single-molecule entries with a valid MS in the filtered NIST library. The dataset is split into a fixed test set and a training pool, maintaining an approximate 8:1:1 train/validation/test ratio. To ensure statistical robustness, we perform 5-fold cross-validation over the training pool with stratified splitting to preserve label distribution across folds. We choose the micro and weighted AUC, and Precision@k ($k = 5$) as the metric to evaluate the multi-label classification performance. Detailed hardware and hyperparameter configurations are provided in Appendix A.

Table 1: Model performance on filtered GS-LF dataset. In brackets, we indicate which chemical encoder was used for training SCENT. We report the mean and standard deviation of 5 seeds, where the asterisk denotes statistical significance ($p < 0.05$), using a Welch's t-test (Welch, 1947), compared to the baseline.

Model	Input	Micro-AUC↑	Weighted-AUC↑	Adj.Precision@k↑
Open-POM	SMILES	81.06 ± 0.60	69.52 ± 1.27	37.54 ± 0.79
MolFormer	SMILES	84.71 ± 1.09	75.38 ± 1.77	41.93 ± 1.90
EIMS2Vec	MS	79.65 ± 0.48	65.40 ± 0.88	37.74 ± 0.28
SCENT(Open-POM)	MS	$83.14 \pm 0.26^*$	$72.93 \pm 0.38^*$	$40.70 \pm 0.46^*$
SCENT(MolFormer)	MS	$82.88 \pm 0.30^*$	$72.94 \pm 0.63^*$	$39.97 \pm 0.29^*$

3 RESULTS

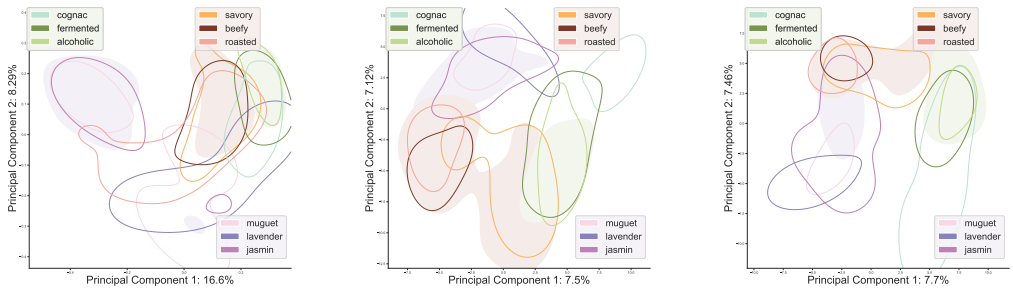
In this section, we validate SCENT by evaluating its quantitative performance on the multi-label odor classification task (Section 3.1). Furthermore, we visualize the learned representation space to qualitatively analyze the presence of odor-related clusters (Section 3.2). Finally, we explore the chemical semantics captured by SCENT by looking at the attention maps of our model (Section 3.3).

3.1 QUANTITATIVE EVALUATION ON ODOR CLASSIFICATION

Model performance is shown in Table 1. Compared with the unaligned MS embedding, the proposed chemical-aligned MS embedding space, whether aligned with Open-POM or MolFormer, significantly improves the performance in odor prediction tasks, demonstrating the effectiveness of our alignment method. Additionally, in Appendix B, we perform an ablation study that reveals the importance of both the transformer architecture and the alignment with chemical structure information to the overall performance of SCENT.

3.2 QUALITATIVE VISUALIZATION OF LEARNED REPRESENTATIONS

In Figure 3 we present the low-dimensional representations of the MS embeddings (obtained using PCA) for different odorants and models. In the aligned spaces (Figure 3B-C), odor perceptual islands are more clearly defined than in the unaligned space. In particular, since Open-POM is directly trained to predict odor labels, the corresponding aligned embedding space exhibits a stronger clustering effect for odor perceptual islands. At the same time, we also observe that there is a decrease in the explained variance ratio of the first two principal components (PCs) after alignment (compared to Figure 3A). We hypothesize this is due to the fact that our contrastive learning objective redistributes information across different dimensions to better capture fine-grained structural semantics. We present extra visualizations in Appendix C.



(A) EIMS2Vec (Liu et al., 2024) (B) SCENT (MolFormer) (C) SCENT (Open-POM)

Figure 3: **Visualization of the latent space of different MS encoder models.** The shade highlights the space associated with coarse-grained descriptors, while the contour lines mark regions with more precise descriptors. We employ PCA to reduce the dimensionality of the corresponding embeddings.

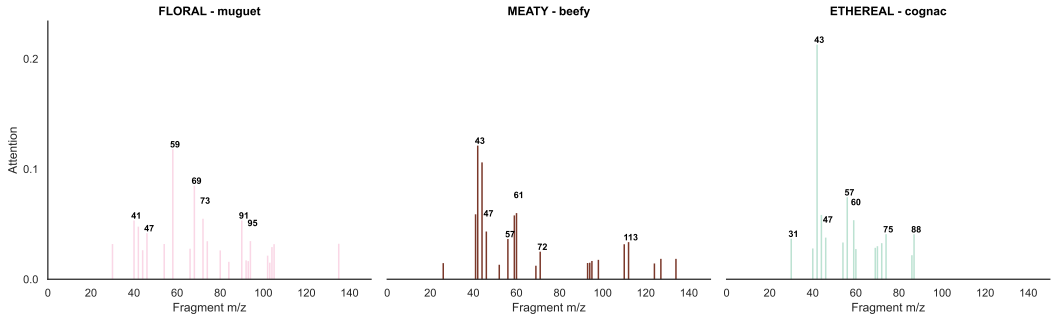


Figure 4: **Average attention map of SCENT**. We plot the attention values from the last transformer layer of SCENT (Open-POM), when generated by different molecules with the same olfactory label ('muguet', 'beefy' and 'cognac'), for different m/z values.

3.3 Attention maps of odor groups

In Figure 4, we present the average attention maps from the CLS token of the last transformer layer of SCENT (Open-POM), when generated by different molecules with the same olfactory label. We observe that the peaks attended to by the model correspond directly to odor-determining substructures. For example, the m/z 61 and 113 in "beefy" are particularly prominent and explicitly indicate sulfur-containing substructures, as described in the Appendix in Table D1. While a single m/z peak alone may not be sufficient to definitively identify the corresponding substructure, the concurrent presence of both fragment ions increases the likelihood that the signals originate from thiol groups. This phenomenon extends to the "Muguet" and "Cognac" categories, where the peaks attended to by the model explicitly point to their corresponding aromatic ether scaffolds and ester structures, respectively. Crucially, since the attention heads remained frozen during this classification task, these chemically meaningful patterns are derived directly from the alignment. That shows that our alignment strategy successfully captures the key structural basis of odor perception, rather than merely memorizing spectral fingerprints. We provide additional examples in Appendix D.

4 Discussion and Conclusions

We propose a contrastive learning framework that integrates chemical information into MS-based representations. While contrastive learning has been successfully applied for alternative mass spectrometric modalities (Chen et al., 2024; Zhang et al., 2024; Bushuiev et al., 2025), which provides explicit molecule fragmentation pathways, it remains underutilized in the context of single-stage EI-MS, that solely relies on implicit structural cues. That makes the task of recovering semantic information more challenging. In this scenario, our cross-modal alignment plays a critical role, encouraging the model to associate latent spectral patterns with chemically meaningful substructures. By bridging the gap between implicit sensor signals and molecular structures, this approach provides a generalized computational space for computational olfaction.

Despite these advantages, our approach faces certain limitations. First, we don't use stereochemical indicators in SMILES representations. While this decision aligns with the intrinsic nature of direct EI-MS, since stereoisomers can evoke different scents, this ambiguity places a theoretical bound on the model's ability to differentiate stereoisomers with distinct odor profiles. Additionally, regarding comparative analysis, we note that while prior end-to-end studies explored similar goals (Ito & Nakamoto, 2020; Hasebe et al., 2022; Debnath & Nakamoto, 2022), direct quantitative comparisons were not feasible in this study due to the unavailability of open-source implementations.

We aim to establish a direct-sampling EI-MS framework that eliminates the need for GC separation, thereby achieving rapid response speeds comparable to conventional gas sensors for computational olfaction applications, while retaining the structural information inherent to EI-MS using comparatively simple and cost-efficient instrumentation. This approach, while promising, introduces several challenges regarding the shift from using data from a curated library to using real-world samples. Future work will focus on enhancing robustness through data augmentation and developing capabilities to handle mixed signals, thereby paving the way for rapid and accurate olfactory digitization.

ACKNOWLEDGMENTS

This work was supported by the Knut and Alice Wallenberg Foundation, Swedish Research Council, ERC AdV BIRD 88480, and ERC Syn D2Smell 101118977. The computational experiments were enabled by the Berzelius resource provided by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre.

REFERENCES

Aryan Amit Barsainyan, Ritesh Kumar, Pinaki Saha, and Michael Schmuker. Openpom - open principal odor map, 2023.

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*, 2025.

Lu Chen, Bing Xia, Yu Wang, Xia Huang, Yucheng Gu, Wenlin Wu, and Yan Zhou. Cmssp: A contrastive mass spectra-structure pretraining model for metabolite identification. *Analytical Chemistry*, 96(42):16871–16881, 2024.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PmLR, 2020.

Tanoy Debnath and Takamichi Nakamoto. Extraction of sensing data for desired scent impressions using mass spectra of odorant molecules. *Scientific Reports*, 12(1):16297, 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-20388-0. URL https://doi.org/10.1038/s41598-022-20388-0.

Dewei Feng, Carol Li, Wei Dai, and Paul Pu Liang. Smellnet: A large-scale dataset for real-world smell recognition. *arXiv preprint arXiv:2506.00239*, 2025.

GoodScent. The good scents company. URL http://www.thegoodscentscompany.com/.

Daisuke Hasebe, Manuel Alexandre, and Takamichi Nakamoto. Exploration of sensing data to realize intended odor impression using mass spectrum of odor mixture. *Plos one*, 17(8):e0273011, 2022.

Keisuke Ito and Takamichi Nakamoto. Improvement of odor impression predictive model using machine learning. In *2020 IEEE SENSORS*, pp. 1–4. IEEE, 2020.

Andreas Keller, Richard C. Gerkin, Yuanfang Guan, Amit Dhurandhar, Gabor Turu, Bence Szalai, Joel D. Mainland, Yusuke Ihara, Chung Wen Yu, Russ Wolfinger, Celine Vens, leander schietgat, Kurt De Grave, Raquel Norel, DREAM Olfaction Prediction Consortium, Gustavo Stolovitzky, Guillermo A. Cecchi, Leslie B. Vosshall, and pablo meyer. Predicting human olfactory perception from chemical features of odor molecules. *Science*, 355(6327):820–826, 2017.

Brian K. Lee, Emily J. Mayhew, Benjamin Sanchez-Lengeling, Jennifer N. Wei, Wesley W. Qian, Kelsie A. Little, Matthew Andres, Britney B. Nguyen, Theresa Moloy, Jacob Yasonik, Jane K. Parker, Richard C. Gerkin, Joel D. Mainland, and Alexander B. Wiltschko. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381(6661):999–1006, 2023.

Leffingwell and Associates. Database of perfumery materials and performance, 2001. URL http://www.leffingwell.com/.

Shibo Liu, Xuan Zhang, Anlei Jiao, Shiwei Sun, Longyang Dian, and Xuefeng Cui. Deep representation learning for electron ionization mass spectra retrieval. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 552–557, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. Large-scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence*, 4(12):1256–1264, 2022.

Farzaneh Taleb, Miguel Vasco, Antonio Ribeiro, Mårten Björkman, and Danica Kragic. Can transformers smell like humans? *Advances in Neural Information Processing Systems*, 37:72032–72060, 2024.

David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988.

Bernard L Welch. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34(1-2):28–35, 1947.

Xiaohua Zhai, Xiao Wang, Basil Mustafa, Andreas Steiner, Daniel Keysers, Alexander Kolesnikov, and Lucas Beyer. Lit: Zero-shot transfer with locked-image text tuning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 18123–18133, 2022.

Hailiang Zhang, Qiong Yang, Ting Xie, Yue Wang, Zhimin Zhang, and Hongmei Lu. Msbert: Embedding tandem mass spectra into chemically rational space by mask learning and contrastive learning. *Analytical Chemistry*, 96(42):16599–16608, 2024.

JianCai Zhu, Yunwei Niu, and ZuoBing Xiao. Characterization of the key aroma compounds in laoshan green teas by application of odour activity value (oav), gas chromatography-mass spectrometry-olfactometry (gc-ms-o) and comprehensive two-dimensional gas chromatography mass spectrometry (gc$\times$ gc-qms). *Food Chemistry*, 339:128136, 2021.

A Hardware and Hyperparameters

All computational experiments were conducted on a computing node equipped with an AMD EPYC 7502 32-Core Processor and an NVIDIA Quadro RTX 6000 GPU (24GB VRAM). The models were implemented using the PyTorch framework and optimized with the Adam optimizer. The hyperparameter of network setup is provided in Table A1.

Table A1: Hyperparameter settings for training and downstream task.

Module	Hyperparameter	Symbol	Value
MS Encoder	Transformer Layers	L	2
	Attention Heads	H	4
	Embedding Dimension	d	500
	Dropout Rate	p	0.1
	Activation	-	ReLU
Alignment	Batch Size	B	512
	Encoder Learning Rate	η	1e-4
	Optimizer	-	Adam
	Temperature	τ	0.03
	Max Epochs	-	50
Downstream	Batch Size	B	64
	Learning Rate	η	1e-4
	MLP Hidden Dim	d_{mlp}	250
	Max Epochs	-	50

B Ablation Study

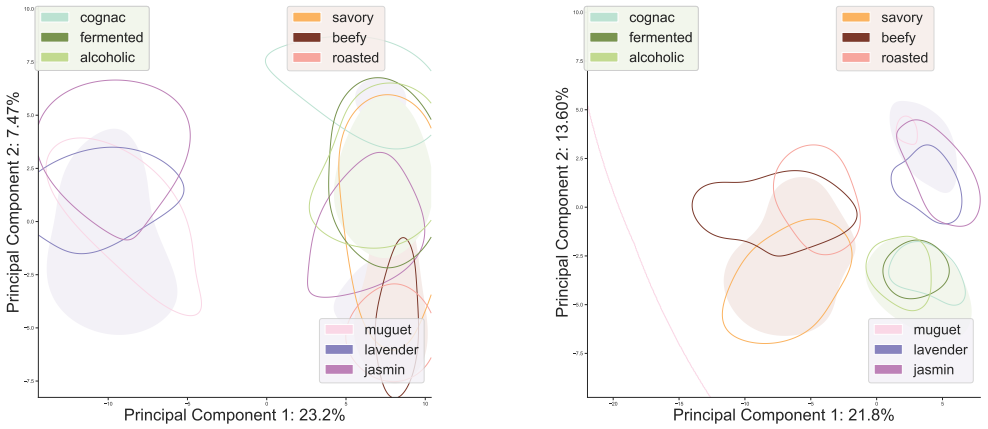
To understand how each component contributes to the overall performance of SCENT, we conduct an ablation study: (i) by removing the proposed transformer head, and (ii) the multimodal alignment objective. All variants are trained following the same experimental settings, using MS data as the sole input. In Table B1, we present the micro-averaged ROC-AUC score of the different ablated versions compared to the baseline. The results highlight how the transformer head and multimodal alignment objective are complementary in order to achieve the level of performance of SCENT.

Table B1: Ablation study of SCENT. The mean and stand deviation of 5-seed results are reported.

Model Variant	Transformer	CLIP	Micro-AUC↑	Weighted-AUC↑	Adj.Precision@k↑
SCENT(Open-POM)	✓	✓	83.14 ± 0.26	72.93 ± 0.63	40.70 ± 0.46
SCENT(MolFormer)	✓	✓	82.88 ± 0.30	72.94 ± 0.63	39.97 ± 1.29
w/o CLIP	✓	×	77.43 ± 0.07	58.68 ± 0.26	29.98 ± 0.16
w/o Transformer (Open-POM)	×	✓	81.23 ± 0.27	66.96 ± 0.81	36.04 ± 0.46
w/o Transformer (MolFormer)	×	✓	82.13 ± 0.53	68.49 ± 1.08	37.52 ± 0.60
EIMS2Vec	×	×	79.65 ± 0.48	65.40 ± 0.88	37.74 ± 0.28

C Additional PCA Visualizations

For completeness, we also present the two-dimensional PCA representation space of the chemical embedding models used in SCENT on the GS-LF dataset. Our results replicate the findings of a previous study presented in Taleb et al. (2024).



(A) MolFormer (B) Open-POM

Figure C1: **Visualization of the latent space of different chemical encoder models.** The shade highlights the space associated with coarse-grained descriptors, while the contour lines mark regions with more precise descriptors. We employ PCA to reduce the dimensionality of the corresponding embeddings.

D ATTENTION MAP OF ODOR CLASS

In Figure D1 we present the attention map of other odor groups. We observe that the model highlights common peaks across odors in the same coarse-grained class, mirroring the fact that structural similarity is a primary driver of olfactory similarity. From the figures, some peaks in the low m/z region (e.g., 41 and 43) are jointly attended by multiple odor categories. These peaks correspond to common fragment ions, such as alkyl or alkenyl chains, produced during electron ionization. The high attention weights on these shared fragments do not imply low discriminative power. Instead, this may suggests that the model has learned the spectral correspondence of organic backbones through structural guidance.

In Table D1 we provide a detailed chemical interpretation of the m/z values with high attention weights of three odor class examples. The analysis reveals a clear correspondence between the focus of the model and chemically meaningful information, confirming that the learned representations are grounded in chemical semantics rather than spurious statistical correlations.

For completeness, in Figure D2 we present the complete set of attention maps for SCENT (MolFormer). The results indicate a similar overall pattern, but with some differences in regards to the focus of the model. We hypothesis, once again, that this is due to the generalized scope of the features of MolFormer, which was trained without supervision on a wide range of odorant (and non-odorant) molecules..

E CHOICE OF SIGNAL MODALITY

In this work, we focus on direct EI-MS. Our choice is motivated by its balance between interpretability with real-time accessibility for computational olfaction. In this section, we compare direct EI-MS with other techniques and highlight the key rationale behind our selection.

- **Hard vs soft ionization**: Compared to online mass spectrometric techniques such as proton-transfer-reaction mass spectrometry (PTR-MS), which rely on softer ionization and often allow only compound-class level assignments, EI-based single-quadrupole MS enables direct comparison to established spectral libraries like NIST library. This allows us to have a certain degree toward molecular-level interpretability, with substantially reduced instrumental complexity and cost.

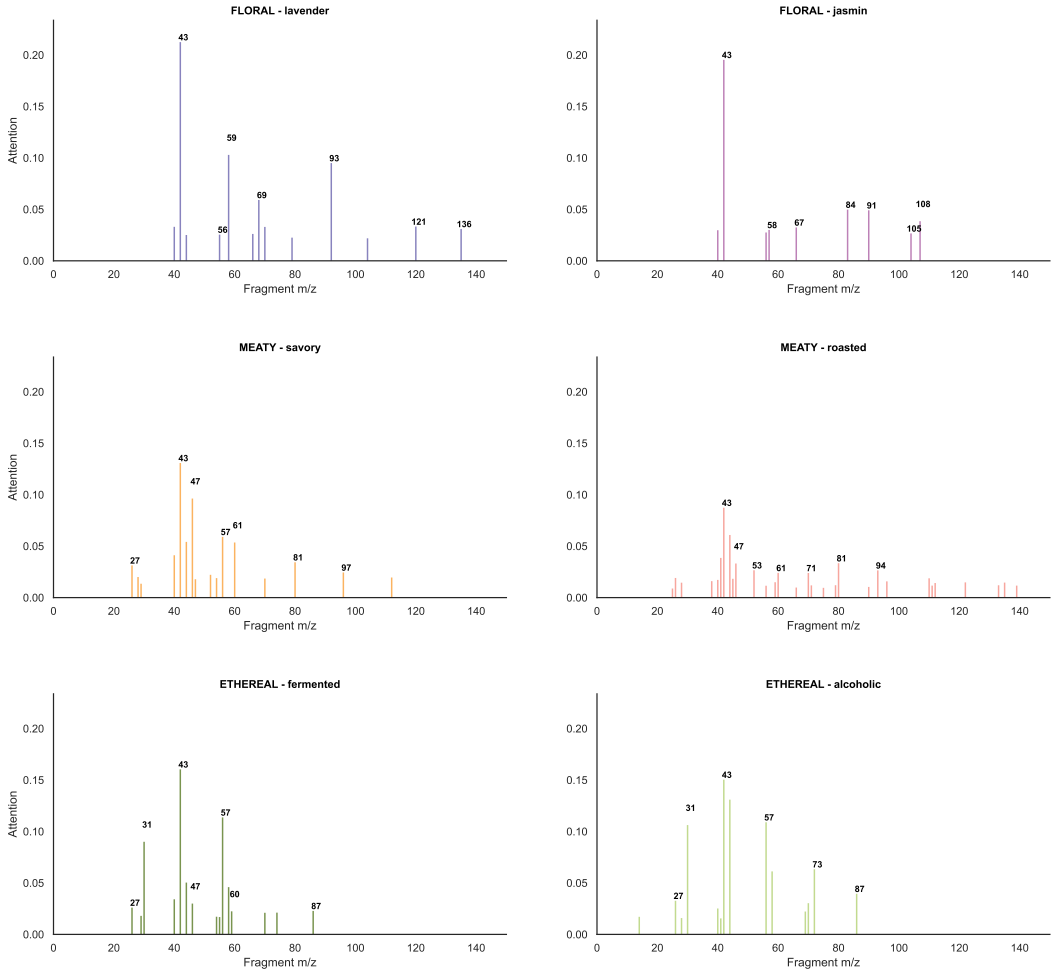


Figure D1: **Average attention map of SCENT (Open-POM)**. We plot the average attention values from the last transformer layer of SCENT (Open-POM), when generated by different molecules with the same olfactory label, for different m/z values.

- **Time vs separation**: Unlike GC-EI-MS, which requires time-consuming separation with experts' experience (in the order of hours), direct EI-MS enables rapid acquisition (in the order of seconds).
- **Electronic sensor vs analytical equipment** In the realm of rapid olfactory sensing, gas sensor arrays are widely adopted. However, they offer only non-specific digital signal

Table D1: Substructure assignments for highlighted m/z values in the attention map

Odor Class	m/z	Substructure	Substructure Assignment
Muguet	73	$[C_4H_9O]^+$	ether
	91	$[C_7H_7]^+$	benzyl
Beefy	61	$[C_2H_5S]^+$	ethyl thiol (α-Cleavage)
	113	$[C_5H_5OS]^+$	2-methylfuran-3-thiol
Cognac	70	$[C_5H_{10}]^+$	pentyl ester (α-cleavage + H migration)
	88	$[C_4H_8O_2]^+$	ethyl ester (β-cleavage)

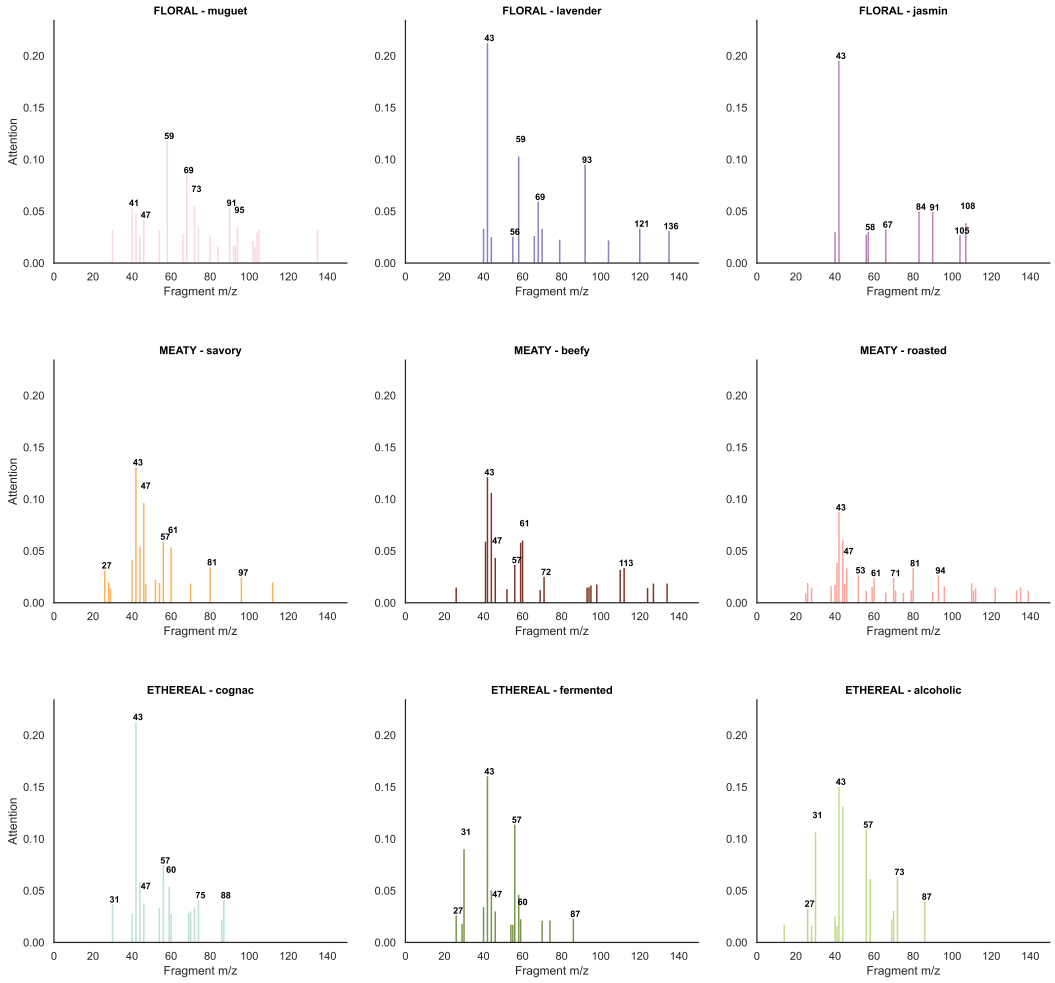


Figure D2: **Average attention map of SCENT (MolFormer)**. We plot the average attention values from the last transformer layer of SCENT (MolFormer), when generated by different molecules with the same olfactory label, for different m/z values.

fingerprints and lack explicit chemical meaning (Feng et al., 2025). In contrast, analytical instrumentation offers molecular-level information, establishing a rigorous chemical basis for olfactory computation.