

# **Development of Olfaction Inspired Odor Sensor Based on Structure-odor Relationships**

匂いと分子構造の関係に基づく嗅覚模倣匂いセンサの  
開発



**Liang Shang**

Graduate School of Information Science and Electrical Engineering  
Kyushu University

This dissertation is submitted for the degree of  
*Doctor of Engineering*

June 2018



## Acknowledgements

I am indebted to many for their help and generosity during my doctoral study at Kyushu University at Japan.

Words are not enough to express my deep gratitude to my supervisor, Prof. Kenshi Hayashi. Without his guidance, patience and encouragement, this dissertation would not have been finished. His diligent, wisdom and personality motivate me tremendously on research, and impact on my life.

I sincere appreciate Dr. Chuanjun Liu who gives me the most helps and advices on my research. He gives me a lot of helpful suggestions on the journal paper writing, and helps me to amend my papers before and after submission.

I would like to give my special thanks to Ms. Maiko Moribe. She supported my research and helped me much about my study and life in Japan.

I would also like to express my sincere appreciation to the vice-supervisors, Prof. Yoichi Tomiura and Prof. Junya Suehiro, for their kind work on the dissertation and defense. Their many suggestions have made this work deeper.

I am very grateful to Dr. Wanming Hao, Mr. Chengkun Jiang, Mr. Xiangwei Wang, Dr. Xiaochen Yang, Dr. Kai Wen, Dr. Ting Chen, Dr. Shiyang Feng and Dr. Binglin Guo for their kind support and friendship. In addition, I would like to thank former and present members of my research group: Mr. Jingshao Li, Mr. Shuanghong Wu, Mr. Lingpu Ge, Mr. Hao Guo, Mr. Zihan Luo and all friends in Hayashi lab for their support during my stay in Japan.

I would like to thank Dr. Midori Watanabe, Dr. Masashi Watanabe and Dr. Hirotaka Yoshioka in Kyushu University for their patient guidance and help on many of experiments. I am also thankful for China Scholarship Council (CSC) for the financial support, which enable me to complete the doctoral study.

Finally, I am particularly indebted to my loving family for their financial support and constant encouragement, but primarily for their love.



## Abstract

As the most primitive sense, olfaction plays an essential role in our daily lives, which provides us an opportunity to explore our chemical environment. Progress in the molecular biology of olfaction has revealed a close relationship between the structural features of odorants and the response patterns they elicit in the olfactory bulb. Molecular feature-related response patterns, termed odor maps (OMs), may represent information related to basic odor quality. To understand the structure-odor relationship is very helpful for clarifying the mechanism of bio-olfaction and developing olfaction-inspired odor sensors. Accordingly, this research aims for exploring the relationship between olfaction information and molecularly information of odorants, and developing a molecular recognized optical sensor platform for volatile organic compounds (VOC) detection and identification. This dissertation consists of seven chapters and the chapter outlines are described as follows:

Chapter 1 composed the background of the present study. The general introductions on the mechanism of bio-olfaction model and odor sensors were reviewed. Also, the basic characteristics and mechanism of localized surface plasmon resonance (LSPR) and molecularly imprinted sol-gels (MISGs) were presented.

Chapter 2 explored the correlation between OMs and the molecular parameters (MPs) of odorants by taking OMs from rat olfactory bulbs and extracting feature profiles of the corresponding odorant molecules. Correlation analysis between the two matrixes was first carried out by establishing coefficient maps. Results from hierarchical clustering showed that all parameters could be segregated into seven clusters, and each cluster showed a relatively similar response pattern in the olfactory bulb. Using the information from the OMs and MPs, we mapped odorants in 2D space by incorporating dimension-reducing techniques based on principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). Artificial neural network models based on the OM and MP feature values were proposed as a means to identify odorant functional groups. An OM-PCA-based model calibrated via extreme learning machine (ELM) was 94.81% and 93.02% accuracy for the calibration and validation sets, respectively. Similarly, an MP-t-SNE-based model calibrated by ELM was 86.67% and 93.35% accuracy for the calibration set and the validation set, respectively. This research supports a structure-odor relationship from a data-analysis perspective.

Chapter 3 presented a proof-of-concept model by which odor information can be obtained by machine-learning-based prediction from MPs of odorant molecules. The features of the MPs were extracted via either unsupervised or supervised approaches and then used as input to calibrate machine-learning models. Predictions were performed by various machine-learning approaches. A support vector machine model combined with feature extraction by Boruta (confirmed only) was found to afford the best results with an accuracy of 97.08%. The result indicated that odor descriptors can be predicted by machine-learning-based models from MPs.

Chapter 4 explored a possibility to use LSPR of Au nanoparticles (AuNPs) and MISGs as the sensitive layer to recognize typical organic acid odorants. The LSPR layer was prepared by vacuum sputtering of AuNPs on a glass substrate and consequently thermal annealing. The sensitive layer was fabricated by spin-coating molecularly imprinted titanate sol-gel on the AuNPs layer. For the MISG coated sensors, the LSPR sensitivity was affected by the spin coating speed. In addition, a sensor array based on MISGs with different templates was constructed to detect the organic acids in single and their binary mixtures. A 100% classification rate was achieved by leave-one-out cross-validation technique for linear discriminant analysis model. It demonstrated that the MISGs coated LSPR sensor array has a great potential in organic acid odor recognition of human body odor.

Chapter 5 developed a sensitive and selective nanocomposite-imprinted, LSPR sensor for cis-jasmone vapor. The functional monomer and the ratio of matrix materials to functional monomers in the MISGs were investigated and optimized. MISGs that contained the functional monomer trimethoxyphenylsilane at a 3:1(v:v) ratio exhibited a higher sensitivity and selectivity than other films.

In Chapter 6, AuNPs were doped in the MISG to enhance the sensitivity of the LSPR sensor through hot spot generation. The size and amount of AuNPs added to the MISG were investigated and optimized. The sensor coated with the MISG containing 20  $\mu$ L of 30-nm AuNPs exhibited higher sensitivity than that of the sensors coated with other films. Furthermore, an optical multi-channel sensor platform containing different channels that were bare and coated with four types of MISGs was developed to detect plant VOCs in single and binary mixtures. k-nearest neighbor model had good potential to identify plant VOCs quickly and efficiently (96.03%). This study demonstrated that an LSPR sensor array coated with an AuNP-embedded MISG combined with a pattern recognition approach can be used for plant VOCs detection and identification.

Chapter 7 summarized the experimental works and concludes the dissertation with recommendations for future work.

# Table of contents

<b>List of figures</b>	<b>xi</b>
<b>List of tables</b>	<b>xv</b>
<b>List of Symbols and Abbreviations</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Olfaction and olfaction system . . . . .	1
1.2 Structure-odor relationship . . . . .	3
1.3 Odor detection method . . . . .	4
1.3.1 Gas chromatography/mass spectrometer . . . . .	4
1.3.2 Gas chromatography-olfactometry . . . . .	5
1.3.3 Gas sensors . . . . .	6
1.3.4 Electronic-nose . . . . .	8
1.4 Localized surface plasmon resonance . . . . .	9
1.5 Molecular imprinted sol-gel . . . . .	11
1.6 Motivation and objectives . . . . .	13
1.7 Organization of dissertation . . . . .	14
<b>2 Odorant clustering based on molecular parameters and odor maps</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.2 Materials and method . . . . .	18
2.2.1 Glomerular activity patterns and molecular parameters . . . . .	18
2.2.2 Construction of PCC-maps for molecular features . . . . .	19
2.2.3 Selection methods for characteristic variables . . . . .	21
2.2.4 Sample division method . . . . .	22
2.2.5 Modeling methods . . . . .	23
2.3 Results and discussion . . . . .	23
2.3.1 Molecular parameters and functional group labels . . . . .	23

2.3.2	Clustering characterization for PCC-maps of molecular parameters . . . . .	25
2.3.3	Mapping the odorants in t-SNE space . . . . .	26
2.3.4	Functional group-identification models . . . . .	30
2.3.5	Discussion . . . . .	35
2.4	Conclusion . . . . .	35
<b>3</b>	<b>Prediction of odor perception from molecular parameters</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Materials and method . . . . .	38
3.2.1	Odor data collection . . . . .	38
3.2.2	Data analysis . . . . .	39
3.2.3	Model validation . . . . .	39
3.3	Results and discussion . . . . .	41
3.3.1	Odor descriptors . . . . .	41
3.3.2	Feature extraction . . . . .	42
3.3.3	Model calibration . . . . .	43
3.3.4	Model comparison . . . . .	47
3.3.5	Model validation . . . . .	48
3.3.6	Discussion . . . . .	48
3.4	Conclusion . . . . .	50
<b>4</b>	<b>LSPR sensor based on MISGs for volatile organic acid detection</b>	<b>51</b>
4.1	Introduction . . . . .	51
4.2	Materials and method . . . . .	53
4.2.1	Chemicals and reagents . . . . .	53
4.2.2	Synthesis of MISG reaction solutions . . . . .	53
4.2.3	MISG coated Au nano-island film preparation . . . . .	53
4.2.4	Vapor generating system . . . . .	53
4.2.5	Sensing system . . . . .	54
4.3	Results and discussion . . . . .	55
4.3.1	UV-vis spectra and vapor absorption characteristics . . . . .	55
4.3.2	Sensor array response for organic acid odors . . . . .	58
4.3.3	Discrimination of organic acid vapors . . . . .	60
4.3.4	Discussion . . . . .	62
4.4	Conclusion . . . . .	63

<b>5 Development of MISG based LSPR sensor for detection of volatile cis-jasmone</b>	<b>65</b>
5.1 Introduction . . . . .	65
5.2 Materials and Methods . . . . .	66
5.2.1 Materials and instrumentations . . . . .	66
5.2.2 Preparation of MISG reaction solution . . . . .	67
5.2.3 MISG-coated LSPR sensor preparation . . . . .	67
5.3 Results and discussion . . . . .	68
5.3.1 Functional monomer selection . . . . .	68
5.3.2 Optimization of the TBOT/TMP ratio . . . . .	72
5.3.3 CJ detection with a MISG-coated Au nano-island sensor . . . . .	73
5.3.4 Discussion . . . . .	76
5.4 Conclusion . . . . .	76
<b>6 LSPR sensor array coated AuNPs@MISGs for PVOCs recognition</b>	<b>79</b>
6.1 Introduction . . . . .	79
6.2 Experimental . . . . .	80
6.2.1 Materials, chemicals and instrumentations . . . . .	80
6.2.2 Preparation of AuNPs@MISGs reaction solution . . . . .	81
6.2.3 AuNPs@MISGs coated LSPR sensor fabrication . . . . .	81
6.3 Results and discussion . . . . .	82
6.3.1 Effect of Au NP size on LSPR . . . . .	82
6.3.2 Optimization of the amount of AuNPs . . . . .	83
6.3.3 Optimization of spin coating speed . . . . .	84
6.3.4 CJ detection with the optimized AuNPs@MISG-coated LSPR sensor	85
6.3.5 AuNPs@MISGs-coated LSPR sensor array . . . . .	87
6.3.6 Identification model calibration . . . . .	90
6.4 Conclusion . . . . .	92
<b>7 Conclusion and prospect</b>	<b>95</b>
7.1 Conclusion . . . . .	95
7.1.1 Structure–odor relationship . . . . .	95
7.1.2 Molecular imprinted material coated optical odor sensor . . . . .	96
7.2 Prospect . . . . .	97
7.2.1 Basic odor clusters and molecular parameters . . . . .	97
7.2.2 Possibility of machine-learning-based GC-O . . . . .	97
7.2.3 Olfaction inspired optical odor sensing system . . . . .	98

<b>References</b>	<b>99</b>
<b>Appendix A Chapter 2 support information</b>	<b>111</b>
<b>Appendix B Chapter 3 support information</b>	<b>115</b>
<b>Appendix C Chapter 6 support information</b>	<b>149</b>

# List of figures

1.1	Schematic diagram of olfaction . . . . .	2
1.2	Schematic diagram of SPME-GC/MS . . . . .	4
1.3	Schematic diagram of GC-O . . . . .	5
1.4	Schematic diagram of MOS sensors . . . . .	7
1.5	Device configuration of conducting polymer films coated chemresistor sensor	8
1.6	QCM gas sensor and sensing principle . . . . .	9
1.7	Comparison of the mammalian olfactory system and the e-nose system . . . . .	10
1.8	Schematic diagrams representing localized surface plasmon resonance . . . . .	11
1.9	Schematic diagram of MISGs . . . . .	13
1.10	Motivation and objectives . . . . .	14
1.11	Outline of the thesis . . . . .	15
2.1	Schematic diagram of data processing . . . . .	18
2.2	Schematic diagram of obtaining matrix of odor maps and molecular parameters	21
2.3	Response regions for 12 types of functional groups . . . . .	24
2.4	Correlation map for molecular parameters . . . . .	26
2.5	Contribution rates and standard deviations for PCC-maps . . . . .	27
2.6	Heat map and hierarchical dendograms . . . . .	27
2.7	The PCC-maps for molecular parameter . . . . .	28
2.8	Contribution rates and standard deviations of PCs for Mom and Mop . . . . .	29
2.9	Odorant clustering maps by t-SNE method . . . . .	32
2.10	Select the numbers of hidden layer nodes for LVQ networks . . . . .	33
2.11	Select the numbers of hidden layer nodes for ELM networks . . . . .	34
3.1	Machine-learning-based olfactometer . . . . .	38
3.2	Concept diagram to predict odor descriptors using molecular parameters . . . . .	38
3.3	Data processing diagram of prediction models calibration and validation . . . . .	41
3.4	The most 20 frequent ODs in Sigma-Aldrich database . . . . .	42

3.5	MPs selection based on BR method for the ten ODs . . . . .	44
3.6	Identification average accuracies for ten odor descriptors . . . . .	46
3.7	Comparison of average identification accuracies . . . . .	48
4.1	The schematic graph of MISG-LSPR multichannel sensor platform . . . . .	52
4.2	Schematics of vapor generation system and sensing system . . . . .	55
4.3	Transmittance spectra and optical features of MISGs . . . . .	56
4.4	Real-time response of MISGs/NISG with different coating speeds to HA vapor	57
4.5	SEM images of bare Au nano-island and MISGs coated samples . . . . .	57
4.6	Real-time responses of HA-MISG-LSPR sensor to three fatty acid vapors .	59
4.7	Transmittance spectra for MISG coated samples . . . . .	59
4.8	Typical response of HA-MISG coated LSPR sensor to fatty acid vapors .	60
4.9	PCA score plots of the multichannel responses . . . . .	62
4.10	LDA score plots of the multichannel responses . . . . .	63
5.1	Schematic of MISG-coated Au nano-islands with functional monomers .	66
5.2	Chemical structures of the functional monomers and PVOCS . . . . .	68
5.3	Absorption spectra for the bare substrate, NISGs/MISGs coatings with functional monomers . . . . .	69
5.4	SEM images of bare, NISG/MISGs coated samples . . . . .	69
5.5	AFM images of bare, NISG/MISGs coated samples . . . . .	70
5.6	FT-IR spectra of samples before and after CJ vapor absorption . . . . .	71
5.7	Real-time and quantitative responses for NISG-/MISGs-coated samples .	72
5.8	Absorption spectra for TMP-MISG-coated samples at different TBOT/TMP	73
5.9	SEM images of samples coated with MISG at different TBOT/TMP . . . .	73
5.10	AFM images of samples coated with MISG at different TBOT/TMP . . . .	74
5.11	Real-time responses of samples coated with TMP-MISG at TBOT/TMP=75/125, 100/100, 125/75, 150/50, 175/25 and their response summary . . . . .	74
5.12	Real-time responses of TMP-MISG-modified Au-islands to PVOCs . . . .	75
5.13	Linear response vs. CJ concentration in air . . . . .	76
6.1	Schematic of AuNPs@MISG-coated Au nano-islands for PVOCs detection	80
6.2	Schematic of AuNPs@MISG-coated sensor array for PVOCs detection . .	81
6.3	Absorption spectra of MISG-coated samples with different size of AuNPs .	82
6.4	SEM images of control and MISGs with AuNPs . . . . .	83
6.5	Response effected by AuNPs size in MISGs . . . . .	84
6.6	SEM images of bare (a), MISG <sub>cis-jasmone-</sub> , MISG <sub>α-pinene-</sub> , MISG <sub>limonene-</sub> , and MISG <sub>γ-terpiene-</sub> , NISG-, and AuNPs doped NISG-coated sample . . . .	84

6.7	Particle size distribution histogram of spherical AuNPs determined from bare sample . . . . .	85
6.8	Absorption spectra of MISG-coated samples with different amount of 30-nm AuNPs . . . . .	85
6.9	SEM images of control, and MISGs with 30 nm AuNPs 5- $\mu$ L, 10- $\mu$ L, 20- $\mu$ L, 30- $\mu$ L, 50- $\mu$ L, 70- $\mu$ L, and 90- $\mu$ L . . . . .	86
6.10	Response effected by Au nanoparticles amount in MISGs . . . . .	86
6.11	Response effected by spin coating speed . . . . .	87
6.12	Real-time responses of AuNPs@MISG <sub>CJ</sub> -modified Au-islands to four types of PVOCs . . . . .	87
6.13	Linear responses vs. CJ concentration in air . . . . .	88
6.14	Correlation matrix for sensor array . . . . .	89
6.15	PCA score plots of the responses of 72 samples . . . . .	90
6.16	LDA score plots of the responses of 72 samples . . . . .	91
7.1	Schematic of hyperspectral camera based muli-MISG-LSPR optical sensing platform . . . . .	98
B.1	The accumulative contribution rates for ten OPs . . . . .	115
B.2	The grid search process of penalty factor ( <i>c</i> ) and RBF parameter ( <i>g</i> ) for SVM models (all parameters) . . . . .	117
B.3	The grid search process of penalty factor ( <i>c</i> ) and RBF parameter ( <i>g</i> ) for SVM models (PCA) . . . . .	118
B.4	The grid search process of penalty factor ( <i>c</i> ) and RBF parameter ( <i>g</i> ) for SVM models (BR-CT) . . . . .	119
B.5	The grid search process of penalty factor ( <i>c</i> ) and RBF parameter ( <i>g</i> ) for SVM models (BR-C) . . . . .	120
B.6	Impact of number of features on the misclassification error for RF models under all parameters . . . . .	121
B.7	Impact of number of features on the misclassification error for RF models under features obtained by PCA . . . . .	122
B.8	Impact of number of features on the misclassification error for RF models under features obtained by BR-CT . . . . .	123
B.9	Impact of number of features on the misclassification error for RF models under features obtained by BR-C . . . . .	124
B.10	Impact of number of trees on the misclassification error for RF models under all parameters . . . . .	125

B.11 Impact of number of trees on the misclassification error for RF models under features obtained by PCA . . . . .	126
B.12 Impact of number of trees on the misclassification error for RF models under features obtained by BR-CT . . . . .	127
B.13 Impact of number of trees on the misclassification error for RF models under features obtained by BR-C . . . . .	128
B.14 The accuracies under different numbers of hidden nodes for ELM models under all parameters . . . . .	129
B.15 The accuracies under different numbers of hidden nodes for ELM models under features obtained by PCA . . . . .	130
B.16 The accuracies under different numbers of hidden nodes for ELM models under features obtained by BR-CT . . . . .	131
B.17 The accuracies under different numbers of hidden nodes for ELM models under features obtained by BR-C . . . . .	132

# List of tables

2.1	46 types of molecular parameters extracted by ChemBio 3D. . . . .	19
2.2	The list of 14 functional group labels considered in this study. . . . .	25
2.3	Summary for the cluster accuracies. . . . .	30
2.4	The optimal training parameters of LVQ, SVM and ELM models. . . . .	31
2.5	Identification accuracies of LVQ, SVM and ELM models. . . . .	31
3.1	Data sets, division of samples, principal components, and molecular parameters . . . . .	40
3.2	Modeling parameters for SVM, RF, and ELM calibration . . . . .	45
3.3	Models validation by golden delicious apple sample . . . . .	49
5.1	The roughness of Bare, NISGs and MISGs with different functional monomers coated samples . . . . .	70
5.2	The roughness of MISGs coated samples with different TBOT/TMP . . . . .	72
6.1	PVOCs identification accuracies of LDA, KNN, and NBC models based on the sensor array response matrix . . . . .	93
A.1	The list of 178 odorants and their CAS numbers, chemical clusters and t-SNE results . . . . .	111
B.1	Odor perceptions appeared in Sigma Aldrich database . . . . .	116
B.2	Comparison of identification results for train set . . . . .	133
B.3	Comparison of identification results for test set . . . . .	133
B.4	The list of molecular parameters in chapter 3 . . . . .	134
C.1	Standard performance measures for train set . . . . .	149
C.2	Standard performance measures for test set . . . . .	150



# List of Symbols and Abbreviations

$\Delta A$	The change of absorption	
$A_{air}$	Absorption in air	
$A_{gas}$	Absorption in gas	
$C$	Concentration of gas	ppm
$D_r$	Diffusion rate	$\mu\text{g}/\text{min}$
$F$	Flow rate of diluent gas	L/min
$k$	Factor for converting gas weight to gas volume	
$M$	molecular weight	g/mol
$P$	Gas pressure	mmHg
$T$	Transmittance in gas	%
$t$	Gas temperature	$^{\circ}\text{C}$
$T_0$	Transmittance in air	%
AFM	Atomic force microscope	
AuNPs	Au nanoparticles	
E-nose	Electronic-nose	
ELM	Extreme learning machine	
FT-IR	Fourier transform infrared spectrometer	
GC-O	Gas chromatography-olfactometry	

GC/MS	Gas chromatography/mass spectrometer
LDA	Linear discriminant analysis
LSPR	Localized surface plasmon resonance
LVQ	Learning vector quantization
MIP	Molecular imprinted polymer
MISG	Molecular imprinted sol-gel
MOS	Metal oxide semiconductor
MP	Molecular parameter
NISG	Non-imprinted sol gel
OB	Olfactory bubble
OM	Odor map
OR	Olfactory receptor
PC	Principal component
PCA	Principal component analysis
PCC-maps	Pearson correlation coefficient maps
PVOC	Plant volatile organic compound
QCM	Quartz crystal microbalance
RI	Refractive index
SEM	Scanning electron microscope
SPME	Solid phase micro extraction
SVM	Support vector machine
t-SNE	t-distributed stochastic neighbor-embedding
VOC	Volatile organic compound

# Chapter 1

## Introduction

*“Did you ever measure a smell? Can you tell whether one smell is just twice strong as another? Can you measure the difference between two kinds of smell and another? It is very obvious that we have very many different kinds of smells, all the way from the odour of violets and roses up to asafetida. But until you can measure their likeness and differences, you can have no science of odour. If you are ambitious to find a new science, measure a smell.”*

---

Alexander Graham Bell, 1914

### 1.1 Olfaction and olfaction system

As the most primitive sense, olfaction plays an essential role in our daily lives [1]. Although olfaction has been considered of secondary importance to senses (such as vision or hearing) for a long time, it is recognized as one of the most necessary senses recently [2, 3]. Over 400,000 chemicals are known to produce a sense of smell in people. Recent research has reported that we can discriminate more than 1 trillion olfactory stimuli [4]. The mechanism through which olfactory perception is achieved remained basically unknown until Richard and Buck discovered odorant receptors and described the organization of the olfactory system [5, 6]. They found a large gene family contained 1,000 different genes that give rise to an equivalent number of olfactory receptor (OR) types. In addition, these ORs are located on the olfactory receptor cells, which occupy a small area in the upper part of the nasal epithelium and detect the inhaled odorant molecules [7]. Since then, our knowledge regarding olfactory perception, particularly at the molecular level, has grown significantly [8–10].

Research into the response patterns of neurons in the olfactory bulb (OB) has helped clarify the mechanism of biological olfaction [11]. The first relay station of the olfactory system

is OB, which has a cortical structure with distinct layers and numerous glomerular modules [12]. As illustrated in Figure 1.1, olfactory stimuli (odorant molecules) are typically volatile chemicals that connected with ORs on olfactory sensory neurons in the olfactory epithelium [13]. These chemicals are firstly absorbed on the mucus in nasal cavity, and detected by the olfactory cilia in ORs. Then, the signal is sensed and processed by the OB. Finally, molecular features of odorants have been shown to be represented by spatiotemporal patterns of activity across olfactory sensory neurons in the OB [14, 15]. Odorants are discriminated and recognized in mammal brains through analysis of these glomerular activity patterns (odor maps) [16, 17].

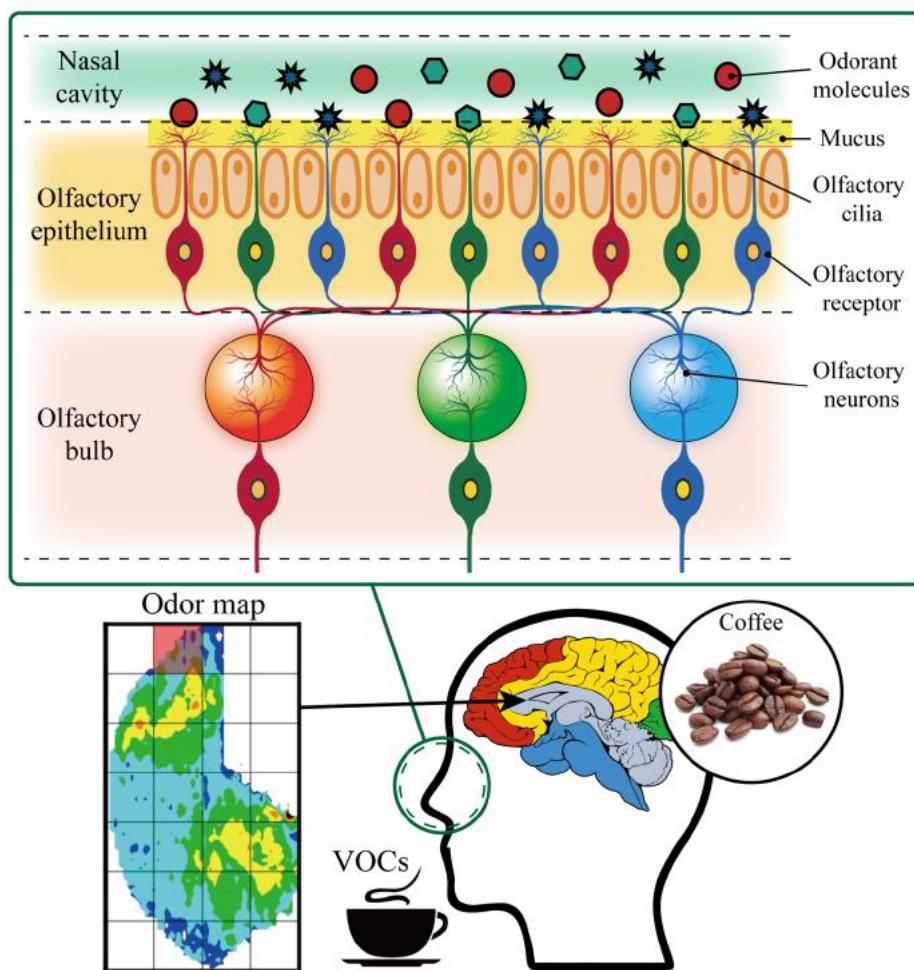


Fig. 1.1 Schematic diagram of olfaction.

By imaging the 2-deoxyglucose uptake in rat glomeruli, Johnson and his team systematically mapped spatially odorant-evoked activity into two dimensional (2D) images for more than 300 odorants [18], and these odor maps now comprise a database (OdorMapDB) that

can be freely accessed (<http://gara.bio.uci.edu/>) [19]. They concluded that clustering responses on the glomerular surface to the molecular features of odorants is likely a general strategy for odor encoding [20]. For example, responses in odor map shift progressively according to odorant carbon chain length. Besides, functional groups, include alicyclic, ester, carboxyl, ketone, aromatic, and alkane, have their feature response area in the odor map [21]. Although odor maps can be affected by the concentration of stimuli chemicals, their z-score patterns are similar [22]. It indicated that normalized odor maps are only depended on chemical's structure. Additionally, Mori et al. have summarized nine molecular-feature clusters that they found at stereotypical OB positions [23].

Research into the molecular biology of olfactory perception has revealed a close relationship between the structural features of odorants and their olfactory perception. For example, functional groups and carbon-chain length play central roles in odor perception [24]. Therefore, to understand the structure-odor relationship is very helpful for clarifying the mechanism of bio-olfaction.

## 1.2 Structure-odor relationship

Much effort has been focused on clarifying the structure-odor relationship using structural, topological, geometrical, electronic, and physicochemical parameters of odorant molecules. For measuring the similarity between two odorants, a vector containing 1,664 descriptors is applied to describe the structure or shape of molecules, and the physicochemical space (principal component space) is used to evaluate the “distance” between them [24]. Further, mass spectra and infrared absorption spectra are used to encode odors via artificial neural networks or self-organizing maps [25]. Sobel et al. related these two spaces to each other and find that the primary axis of perception (defined as odor pleasantness), reflects the primary axis of physicochemical features [26]. Kumar et al. developed a network based approach (smell network) which can be used to explore the perceptual universe and prove the underlying similarity of percepts [27, 28]. Keller et al. established a machine-learning algorithm using a large olfactory psychophysical data set, which can be used to predict odor intensity, pleasantness, and semantic descriptors from chemical features of odor molecules [29]. However, olfaction is extremely complex, and a complete understanding of the structure-odor relationship has yet to be realized. Despite efforts have been made to measure smell, none can describe all pertinent aspects of olfactory perception [30], and the results are difficult to explain without knowledge of biology [31].

Finding basic odors is difficult because the numbers of olfactory receptors and odorants are very large. While most studies focus on attempting to connect odorant physicochemical

properties to olfactory perception [32], objective information such as OB response patterns has rarely been considered. Nevertheless, studying the relationship between olfactory response patterns and the structural features of odorants can be helpful in understanding the mechanisms underlying olfactory perception and for predicting the structure-odor relationship [33].

## 1.3 Odor detection method

### 1.3.1 Gas chromatography/mass spectrometer

As the most excellent gas analytical method, Gas chromatography/mass spectrometer (GC/MS) can offer both high accuracy and sensitivity for the analysis of highly complex mixtures of compounds [34, 35]. In GC part, molecules can be separated from a mixture base on their relative affinity (retained time) for the stationary phase of the column. MS was used to capture, ionize, accelerate, deflect, and detect the ionized molecules separately. By breaking each molecule into ionized fragments to obtain the mass-to-charge ratio, molecules can be detected and distinguished. As illustrated in Figure 1.2, combined with solid phase micro extraction (SPME) method, gas/odor sample can be easily detected and analyzed by GC/MS. GC/MS has been widely used in several areas, such as medicine, food and fragrance analysis and biological analysis [36]. However, GC/MS is not suitable for on-line detection because of its high-cost, time-consuming and bulky size etc. [37]. Therefore, sensors need to be explored for odor sensing.

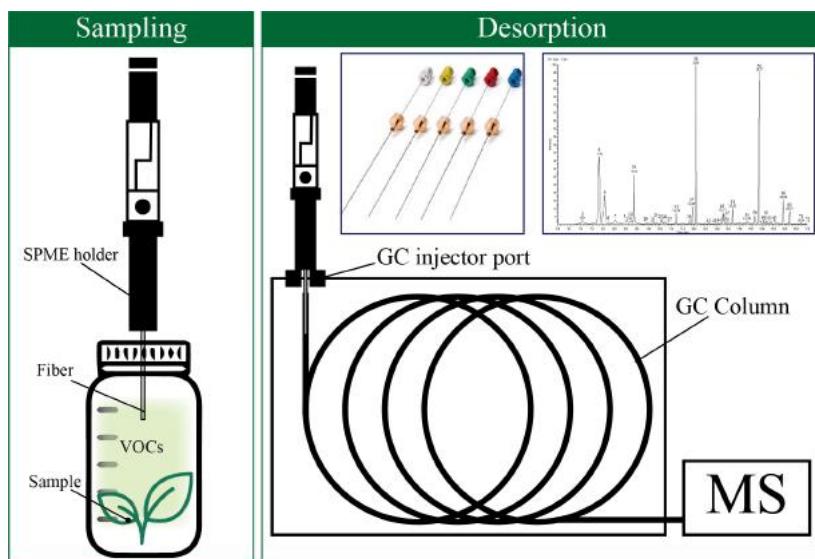


Fig. 1.2 Schematic diagram of SPME-GC/MS.

### 1.3.2 Gas chromatography-olfactometry

Gas chromatography-olfactometry (GC-O) has been developed as a powerful tool in the field of odor research because of the coupled performance of gas chromatographic analysis with human panelist sensory detection (Figure 1.3) [38, 39]. GC-O can work not only as an instrumental analysis to identify and quantify complex odor mixtures, but also as a sensorial analysis to assess odor or odor-active compounds within the GC effluents [40]. In GC-O analysis, the eluted substances are perceived simultaneously by two detection systems; one is a MS system and the other is the human olfactory system [41, 42]. Evaluation by a human sniffer plays an important role because it can make up for deficiencies of GC (or GC/MS) in odor analysis [43]. For example, many of peaks detected by GC for an odor mixture may not actually contribute to our perception since they are present below our thresholds for detecting them. Conversely, some compounds may not show up as detectable GC peaks, but may have a low perception threshold and contribute substantially to a sample's profile. The sensory evaluation of smells by trained panelists can overcome such problems and represents a valid approach to odor assessment. Through sniffing GC effluent components, panelists can determine the odor characteristics ascribed to each individual component, which is important information for the overall odor analysis [44].

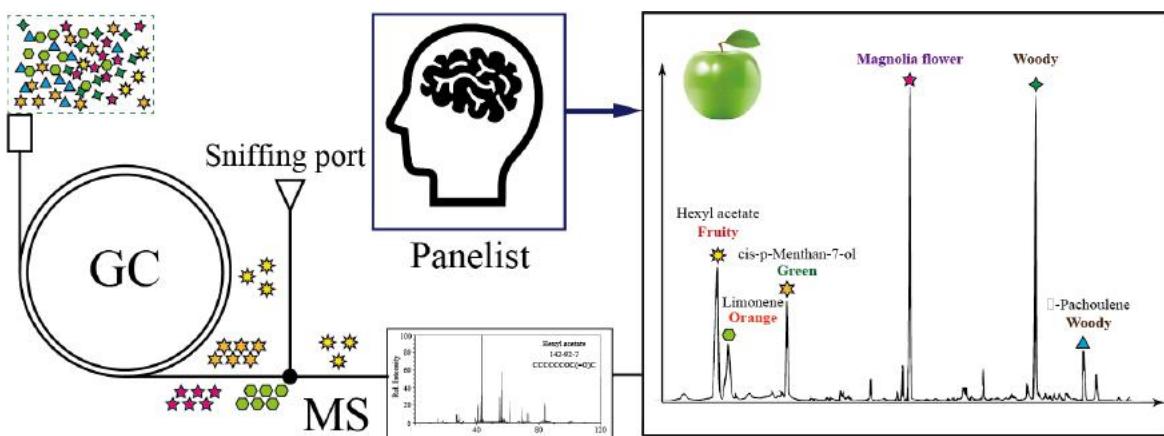


Fig. 1.3 Schematic diagram of GC-O.

A major problem of GC-O is the subjectivity of assessors at the intra- and inter-individual level. Sensory assessment of smells by panelists is influenced by many factors, such as the testing environment, experimental bias, assessor sensitivity, assessor selection, and training [45]. Experimental conditions should be well established to ensure accuracy and precision of the odor descriptor data collected by the panelist. Therefore, although GC-O has presented many challenges not considered on typical GC analysis, its application and promotion are hindered by the variability, high technical requirements, and high costs of the

trained panelist [46–48]. Some software modules have been designed and applied in GC-O as a supplement to odor and chemical analysis. For example, AroChemBase (Alpha MOS) consists of the most comprehensive chemical and sensory library ever, which is convenient for fast sensory profiling and detailed chemical/odor characterization. In this kind of software module, however, the number of compounds with odor descriptors (around 2000) is far less than that of the total compound pool (around 44000). Therefore, regardless of human assessment or software indexing, the question of how to effectively obtain sensory information for eluted compounds from GC is still unanswered for GC-O.

### 1.3.3 Gas sensors

Except instrument analyzation, odorants can also be detected by chemical sensors. Basically, chemical sensors are mainly consisted by transducers and sensing materials [49]. Based on the physico-chemical interaction between volatile compound molecules and sensitive material coated on a transducer, odor can be detected and recorded [50]. Variety of sensors are reported, such as metal oxide semiconductor (MOS), conducting polymer, quartz crystal microbalance (QCM), and semiconductors base gas sensors [51–54].

MOS sensors play a critical role in gas sensors [55]. As illustrated in Figure 1.4, gas molecules can be detected by the resistance change of dioxide (such as  $\text{SnO}_2$ ,  $\text{ZnO}$ ,  $\text{In}_2\text{O}_3$  and  $\text{WO}_3$ ) upon exposure to a trace concentration of reducing or oxidizing gases [56]. In air, donor electrons can be attracted toward oxygen on the surface of the sensing film to prevent electric current flow. When the dioxide exposed in gas, the surface density of oxygen is decreased by reacting with the gases. Therefore, the electrons can be released into the dioxide layer to increase current flowing through the sensor. Because nanostructures with high surface area are employed to enhance the sensitivity, various oxide nanostructures have been explored for MOS sensor development [57]. MOS sensors have been widely used to detect  $\text{H}_2$ ,  $\text{O}_2$ , alcohol and harmful gases, such as CO and  $\text{Cl}_2$  [58, 59]. Liu et al. demonstrated the use of  $\text{SnO}_2$  nanomembranes for high sensitivity and fast response-recovery for acetone vapor detection [60]. Chen et al. recently developed a new  $\text{Pd}/\text{HfO}_2/\text{GaN}$  MOS-type sensor for hydrogen sensing with good sensing performance and lower detection limit [61]. Li et al. reported a successful decoration of ZTO NPs on the surface of RGO nano sheet for ethanol-sensing. The sensitivity and response to ethanol vapor were observably enhanced [62]. The main advantages of MOS sensors are fast response and recovery speed, inexpensive, and easy to fabricate [63, 64]. However, high operating temperatures (200–400 °C) are limited their applications. Besides, sulphur poisoning and ethanol would be generated by MOS sensors.

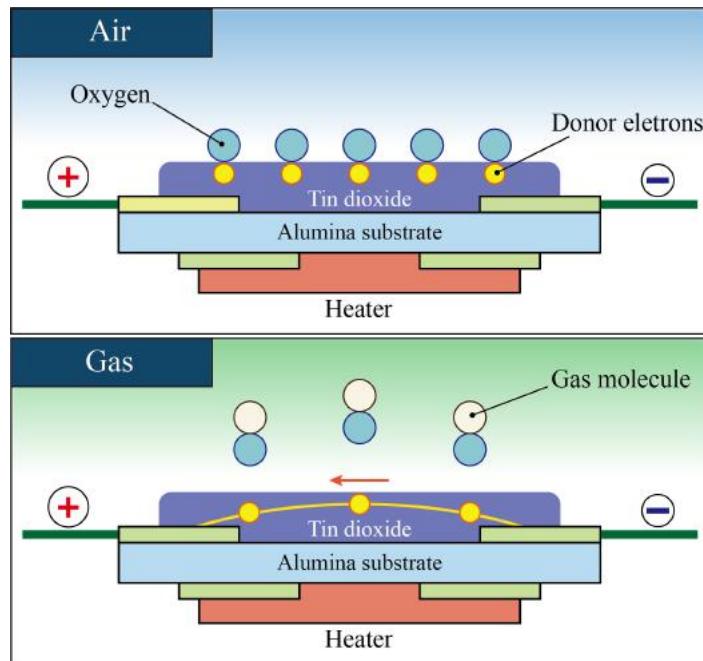


Fig. 1.4 Schematic diagram of MOS sensors.

Chemiresistors are the most common type of sensors for gas sensing, which can be easily fabricated through a cheap and convenient process [65–67]. Conducting polymers have been used as the active layers for gas sensor developing [68, 69]. As illustrated in Figure 1.5, by spin coating, drop-coating, dip-coating or vapor deposition polymerization, conduct polymer film can be established on the surface of interdigitated electrodes [70]. In addition, conduct polymer composites consist of conducting particles, such as polypyrrole or carbon black, have been used for enhancing sensitivity and selectivity [71]. When the sensor is exposed to gases, molecules permeate into the polymer layer and causes the film to expand, which induced an increase in the electrical resistance [72]. Jia et al. developed a flexible gas sensor with PPy for in situ detections for ammonia with limit of detection as 1.2 ppm [73]. Khalil et al. reported a transparent conducting polymer AuNPs nanocomposite thin films for organic gases detection [74]. Ghoorchian et al. presented a sensitive chemiresistor gas sensor modified with nanostructured PPy to detection of TNT with a high sensitivity and good selectivity [75]. Conducting polymers can be synthesized through chemical or electrochemical processes easily and their molecular chain can be modified with copolymerization or structural derivations [76]. They offered many advantages over other materials, such as high sensitivities, short response time and stable at room temperature [77, 78]. However, aging problem limited their applications.

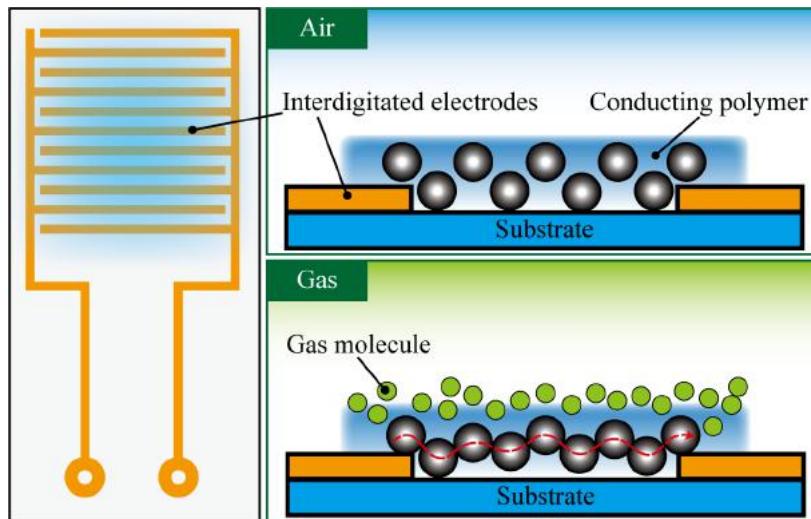


Fig. 1.5 Device configuration of conducting polymer films coated chemresistor sensor.

The other widely used gas sensors are QCM sensors. As shown in Figure 3.5, their sensing principle is the variation of frequency of a crystal based oscillator circuit due to the mass deposition on the crystal [79]. Sorbent coating materials (such as polymeric materials) are necessary for QCM sensors. By designing these coating materials, the vapors of interest and capable to detect in part per billion (ppb) level [80, 81]. Liu et al. demonstrated QCM sensor array coated with molecularly imprinted nanobeads for detection of typical carboxylic acid vapors from human body odor [82]. Lal et al. recently developed a QCM chemical sensor modified nanoclay doped polymeric films for sensing of toxic chemicals in environments. Their sensors shown large and stable response to target chemicals [51]. Ayankojo et al. designed and fabricated a QCM sensor based on based on molecularly imprinted polymer to detect amoxicillin antibiotics in aqueous samples [83]. QCM sensors are as highly sensitive and accurate mass sensors for the detection of chemical and biological warfare agents [84, 85]. However, their response signals are easily effected by noise and humidity, which limited their applications.

### 1.3.4 Electronic-nose

Electronic-nose (E-nose) systems have been regarded as the most useful instrument because of its significant features, such as relatively fast assessment speed, selectivity and stability [86]. In addition, they are cheap sensors which can be easily integrated in current production processes [87]. As illustrated in Figure 1.7, E-nose is composed by several of sensor elements, we can also call them sensor array [88]. Similar to the OBs in our noses, the sensor array is employed as odor molecular parameter capture system. Then, pattern recognition

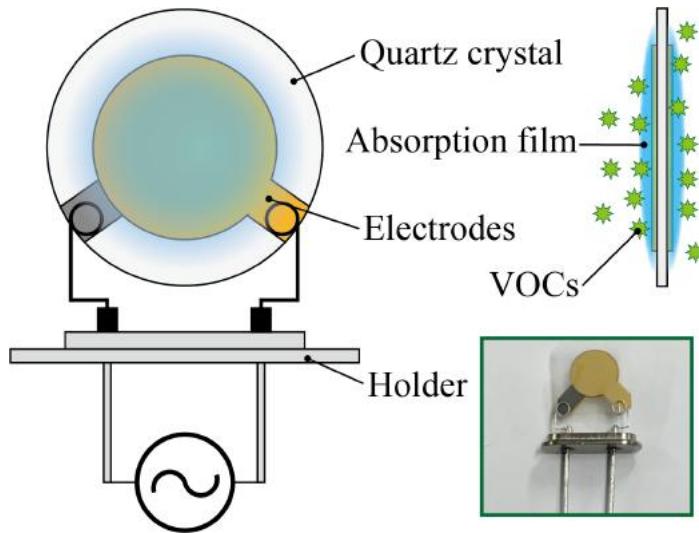


Fig. 1.6 QCM gas sensor and sensing principle.

algorithms are used for analyzing these high dimension responses, just as our brain. Nowadays, E-noses were commercially available in the market, and most of them are consisted by MOS sensors or conducting sensors [89]. However, there are still relatively few applications of electronic noses adopted in industry, which is attributed to difficulties in robustness, selectivity and reproducibility of the sensors. Besides, it is difficult to obtain more molecularly parameters based on existing sensing platform. Nonetheless, the application of E-noses is rapidly expanding in many areas, including the agricultural, biomedical, environmental, food, manufacturing, military, pharmaceutical, regulatory, and various scientific research fields [90]. Dutta et al. described a method for developing a novel, low-cost, hand-held E-nose for black tea flavor estimation with high classification efficiency [91]. Chen et al. successfully identified the characteristic aroma components from ten jujube varieties using E-nose combined with HS-SPME/GC-MS technology [92]. Furthermore, we now have a greater understanding of the genetics behind the ORs/OBs and the relationships between an odorant's molecular property and odor perception (such as sweet). Therefore, development of sensor element competed with ORs is the critical challenge.

## 1.4 Localized surface plasmon resonance

The phenomenon of localized surface plasmon resonance (LSPR), which results from the plasmonic response of nanoparticles by incident electromagnetic waves, can be applied in sensing of analytes [93–97]. As illustrated in Figure 1.8, LSPR involves the interaction of metallic nanoparticles (NPs) with electromagnetic waves to induce plasmon oscillations at

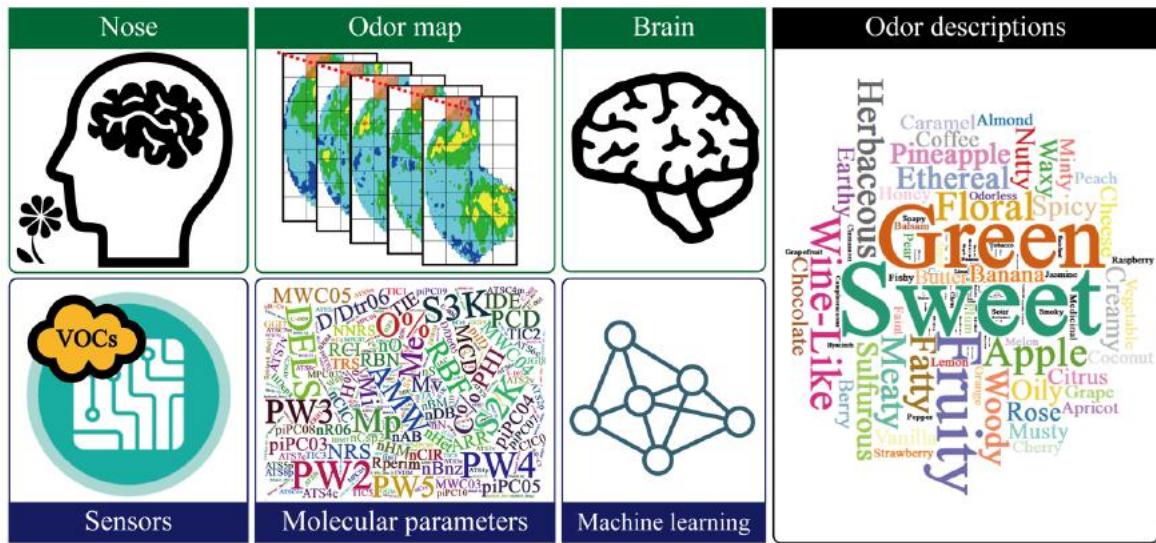


Fig. 1.7 Comparison of the mammalian olfactory system and the e-nose system.

NP surfaces [98–100]. To clarify the mechanism of LSPR sensors, the model was shown by Equation 1.1 [101, 102].

$$\Delta\lambda_{max} = m\Delta n(1 - e^{-2d/l_d}) \quad (1.1)$$

where,  $\Delta\lambda_{max}$  indicates the wavelength shift,  $m$  indicates the refractive index (RI) sensitivity,  $\Delta n$  is the change in RI induced by adsorption,  $d$  is the effective thickness, and  $l_d$  is the characteristic length of the electromagnetic-field decay.

Equation 1.1 indicated that the changes in RI of the surrounding media were induced when molecules entered in the sensing volume of the metal nanoparticles [103]. Therefore, it can be used as a transducer by converting changes in RI into spectral shifts through induced electromagnetic fields [104, 105]. Because of its rapid response and high sensitivity, LSPR has led to the development of optical sensors for various analytes, such as polyphosphates, blood plasma, and wine [106–110]. Shrivastava et al. reported a simple and selective method for detection of vitamins B-1 and B-6 using LSPR as a chemical sensor colorimetrically. The sensor showed remarkable abilities in terms of the stability, reproducibility and sensitivity [111]. Lin et al. recently developed a fast, sensitive and high-resolution colorimetric sensor for benzoyl peroxide detection by using orange sheath-like Au@Ag nanorods as the LSPR transducer [112]. Li et al. described a simple, rapid and reliable LSPR biosensor for erythrocyte counting and ABO blood group typing, which fabricated by immobilization of antibodies on the gold nanoprisms surface [113]. Alula et al. recently developed a simple, rapid and low cost LSPR sensor to sensitive and selective quantitation detection of creatinine

in urine with a low detection limit [114]. Compared with other transducers, such as MOS sensors and QCM sensors, the superiorities of LSPR sensors are high-speed response and rapid recovery, which had been proved by our previous to be suitable in gas and odor sensing [115].

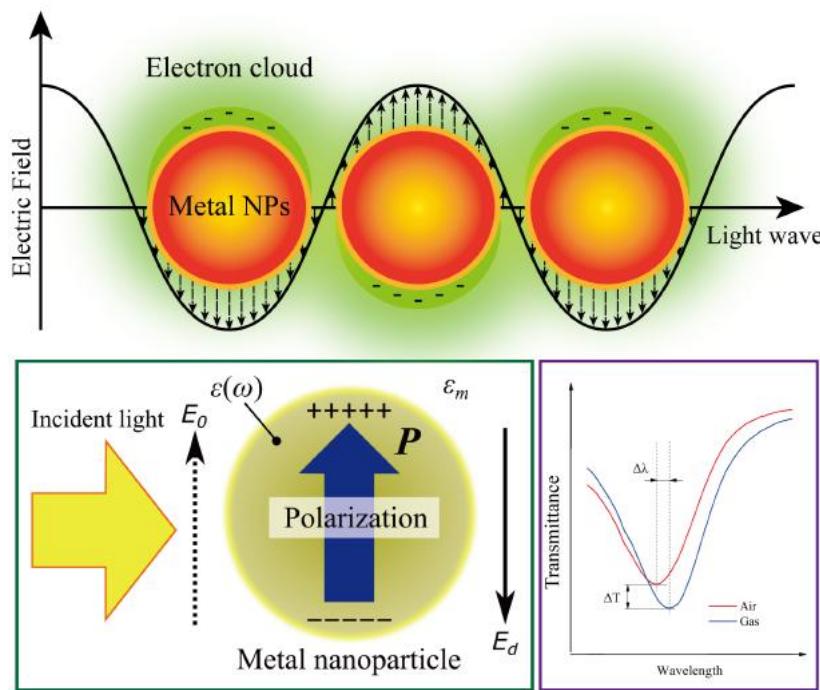


Fig. 1.8 Schematic diagrams representing localized surface plasmon resonance.

## 1.5 Molecular imprinted sol-gel

Because responses of LSPR sensors depend on changes in the media surrounding the NPs, there is a low selectivity for target analytes [116, 117]. Therefore, absorbing materials or antibodies are always coated on the NP surfaces to enhance sensitivity and selectivity [118, 119]. Molecular imprinting is an effective approach for creating recognition patterns of diverse length scales for molecules [120, 121]. By adding templates in the starting material, recognition sites would be generated in these polymer materials [122]. When the templates were removed from the polymer matrix by washing or heating, nano-scale cavities similar to the template molecules were generated [123]. Through these high specificity cavities, molecularly imprinted polymer (MIP) had been applied as a highly selective sensitive layer to some transducers, such as QCM sensors [124–127]. Babamiri et al. developed a MIP-modified nickel nanoclusters biosensor for creatinine determination with a large linear

range and low detection limit [128]. Zhang et al. recently demonstrated organic electrochemical transistors modified with MIPs as highly selective film for chiral recognition of D/L-tryptophan and D/L-tyrosine [129]. Liu et al. reported a new MIP electrochemical sensor with ultrahigh sensitivity and selectivity for the detection of 17-beta-estradiol in attomolar levels [130]. Huang described a biomimetic electrochemical sensor based on MIPs as recognition film for detecting glycoproteins selectively and sensitively [131]. Our previous work has demonstrated that the introduction of MIP as a sensitive layer is an effective approach to increase the selectivity of LSPR sensor, which has been used in the selective detection of terpene vapors [115, 132]. However, organic polymers have some limitations, such as poor site accessibility and chemical stability.

As one of the most promising tools in material science, sol-gel technique allows us to design desired materials at low temperatures [133]. The process of sol-gel is concerned with a transition of a system from liquid ‘sol’ into solid ‘gel’. The schematic representation of the sol-gel process is illustrated in Figure 1.9. By reacting with each other or with the un-hydrolyzed ones to undergo a condensation reaction, a cross-linked matrix can be formed [134–136]. In addition, by adding recognition element (templates) in sol-gel matrix, molecular imprinted sol-gels (MISGs) can be generated as sensitive materials for gas sensing. Recently, sol-gel technology has been applied in various fields, specially in biological and chemical sensors [137]. Different to traditional polymers, MISGs have high selectivity, low cost, long life cycles, and tailored physical-chemical properties [138]. Thus, MISGs have been used for chromatographic separations and chemical sensors [139]. Ostovan et al. developed superparamagnetic MISGs as bio-sorbents for selective dispersive solid phase extraction from urine samples [140]. Bou-Maroun et al. reported a microwave sensor coupled with molecularly imprinted silica for iprodione fungicide detection from hydroalcoholic medium [141]. Because MISGs are transparent in the visible region, the potential application of these thin sol-gel films is to develop optical sensors. For this reason, MISGs-coated LSPR would be expected as a smart strategy for odor detection. Besides, MISGs can be designed by functional monomers (FM) and template molecules selection. The interaction between FMs and target molecules plays an important role in cavities generation in MISGs. Additionally, the interactions also associated with the molecularly structure of odorants. Therefore, MISGs would be applied as smart absorption materials for developing bio-olfaction based sensing system.

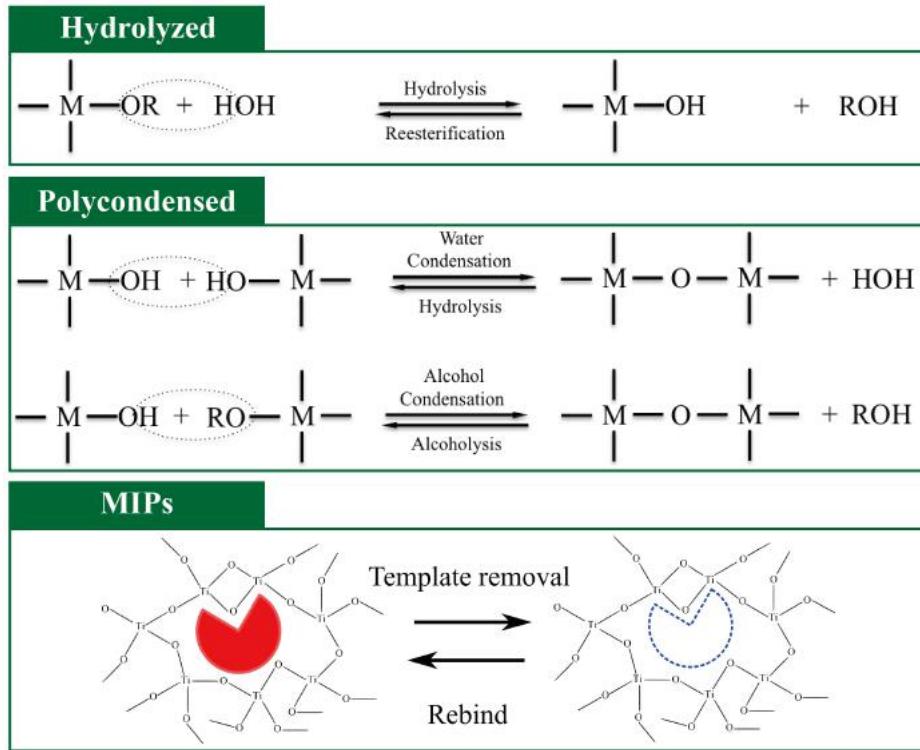


Fig. 1.9 Schematic diagram of MISGs.

## 1.6 Motivation and objectives

As illustrated in Figure 1.10, for developing olfaction inspired odor sensor, the relation between odor descriptions, odor maps and molecular parameters should be discussed. In addition, highly selectivity and sensitivity sensing platform should be explored and developed. MISGs was expected for capturing the molecularly structure information from odorants. Therefore, the primary objective of present study will focus on the following 3 aspects:

- The first objective is to explore the relationships between odor-induced patterns of activity (odor map) and the associated molecular parameters.
- The second task is to explore the relationships between odor perception and molecular parameters of odorants.
- The last objective is to explore a possibility to use LSPR and MISG for developing olfaction inspired odor sensors.

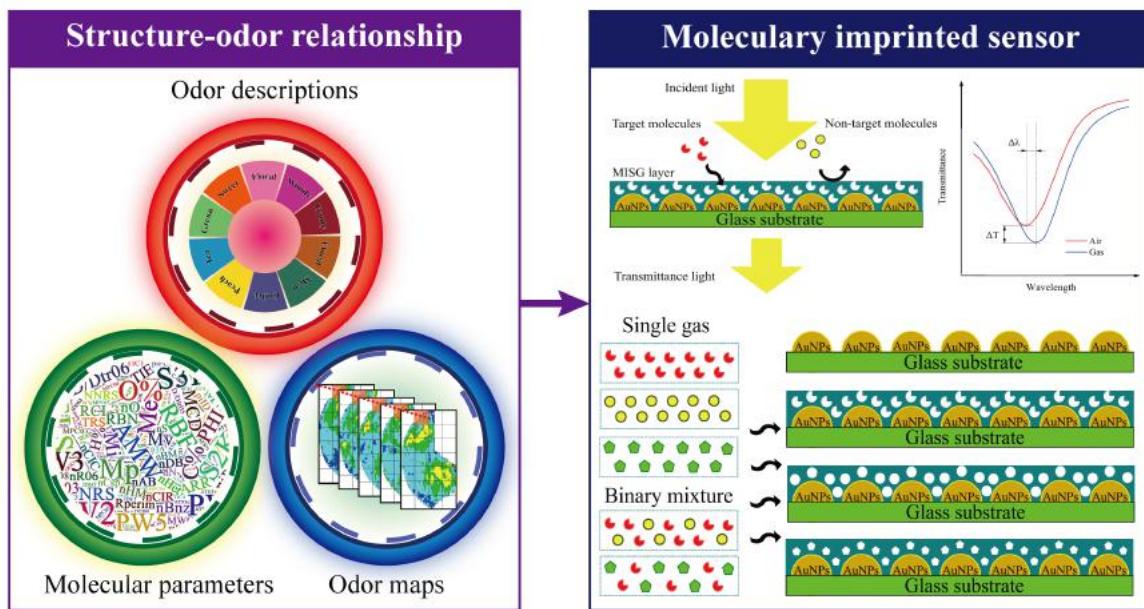


Fig. 1.10 Motivation and objectives.

## 1.7 Organization of dissertation

The organization of the dissertation is illustrated in Figure 1.11.

Chapter 1 composed the background of the present study. The general introductions on the mechanism of bio-olfaction model and odor sensors. In addition, the basic characteristics of LSPR sensors and MISGs are presented. Finally, the motivation and objectives of the dissertation are introduced.

Chapter 2 explores the correlation between odor maps (OMs) and the molecular parameters (MPs) of odorants [142]. Artificial neural network models based on the OM and MP feature values are proposed as a means to identify odorant functional groups. The feasibility of OMs and MPs for odorant function-group classification using each model is accessed.

Chapter 3 presents a proof-of-concept model by which odor information can be obtained by machine-learning-based prediction from molecular parameters of odorant molecules [143]. Different machine learning algorithms are employed and their prediction results were compared. Finally, golden delicious apple GC-MS data is used to prove the feasibility of the model established in present study.

Chapter 4 explores a possibility to use LSPR of AuNPs and MISGs as the sensitive layer to recognize typical organic acid [144]. MISG films generated by different organic acid templates are coated on Au nano-island layers for establishing a MISG-LSPR multichannel sensor platform. The feasibility of the sensor array is discussed and evaluated.

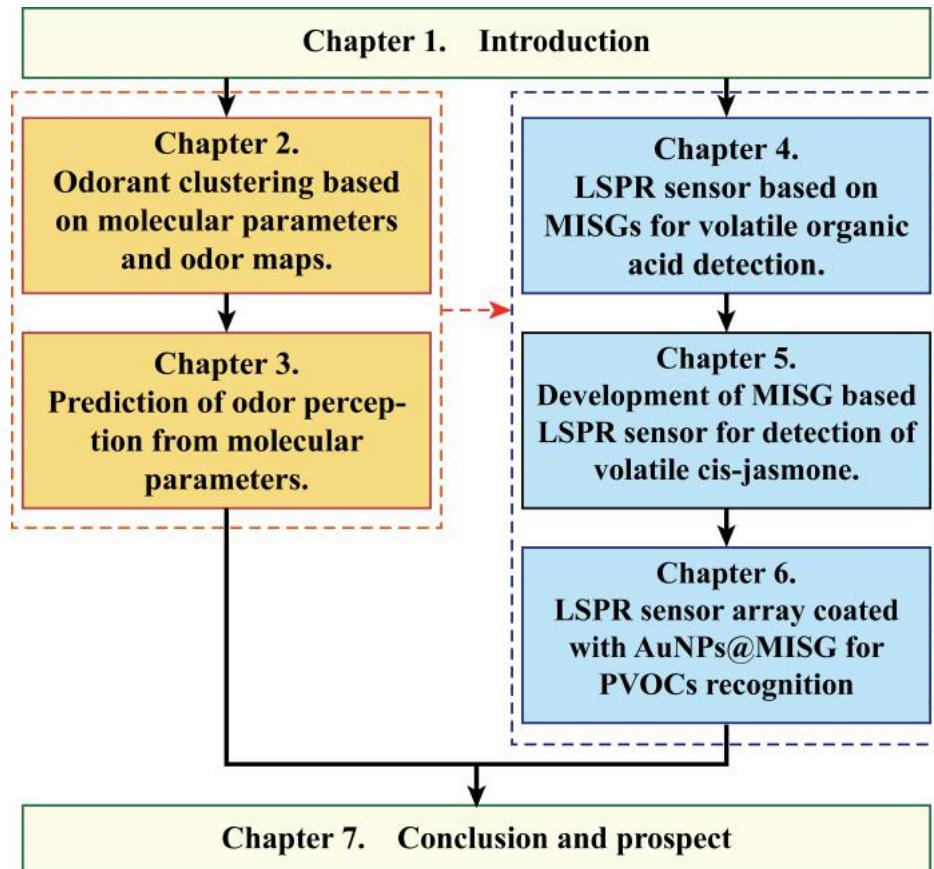


Fig. 1.11 Outline of the thesis.

Chapter 5 fabricates a sensitive and selective nanocomposite-imprinted, LSPR sensor for cis-jasmone vapor [145]. The functional monomer and the ratio of matrix materials to functional monomers in the MISGs are investigated and optimized.

Chapter 6 develops an AuNPs@MISG-coated LSPR sensor platform for detection and identification of PVOCs. AuNPs doped in MISGs are expected for enhancing the signal intensity by hot spot effect. Critical parameters for AuNPs@MISGs (AuNPs size, amount, and spin coating speed) are tuned for sensitivity optimization. The identification capacity of the sensor array for PVOCs is assessed and discussed.

Chapter 7 summarizes the experimental works and concludes the dissertation with recommendations for future work.



# **Chapter 2**

## **Odorant clustering based on molecular parameters and odor maps**

### **2.1 Introduction**

Studying the relationship between olfactory response patterns and the structural features of odorants can be helpful in understanding the mechanisms underlying olfactory perception and for predicting the structure-odor relationship. Consequently, the primary goal of this chapter was to explore the relationships between odor-induced patterns of activity and the associated molecular features.

Firstly, we obtained 2-deoxyglucose glomerular activity-pattern images for 178 odorants from the Johnson freely available odor-map database. For each map, the gray value of each pixel was extracted from the images to fabricate a  $178 \times 70329$  image matrix. Forty-six molecular feature parameters for the odorants were calculated using BioChem3D software. A schematic of the data-processing method is shown in Figure 2.1. Based on the characteristic variables extracted by principal component analysis (PCA), hierarchical clustering analysis (HCA) was performed on the Pearson correlation coefficient maps (PCC-maps) to investigate the effects of the molecular parameters. 2D artificial cluster maps based on the olfactory and molecular information were generated via t-distributed stochastic neighbor-embedding (t-SNE). Based on these datasets, three machine learning models-learning vector quantization (LVQ) network, support vector machine (SVM), and extreme learning machine (ELM)-were employed to establish odorant function-group discrimination models. We then assessed the feasibility of odor maps (OMs) and molecular parameters (MPs) for odorant function-group classification using each model.

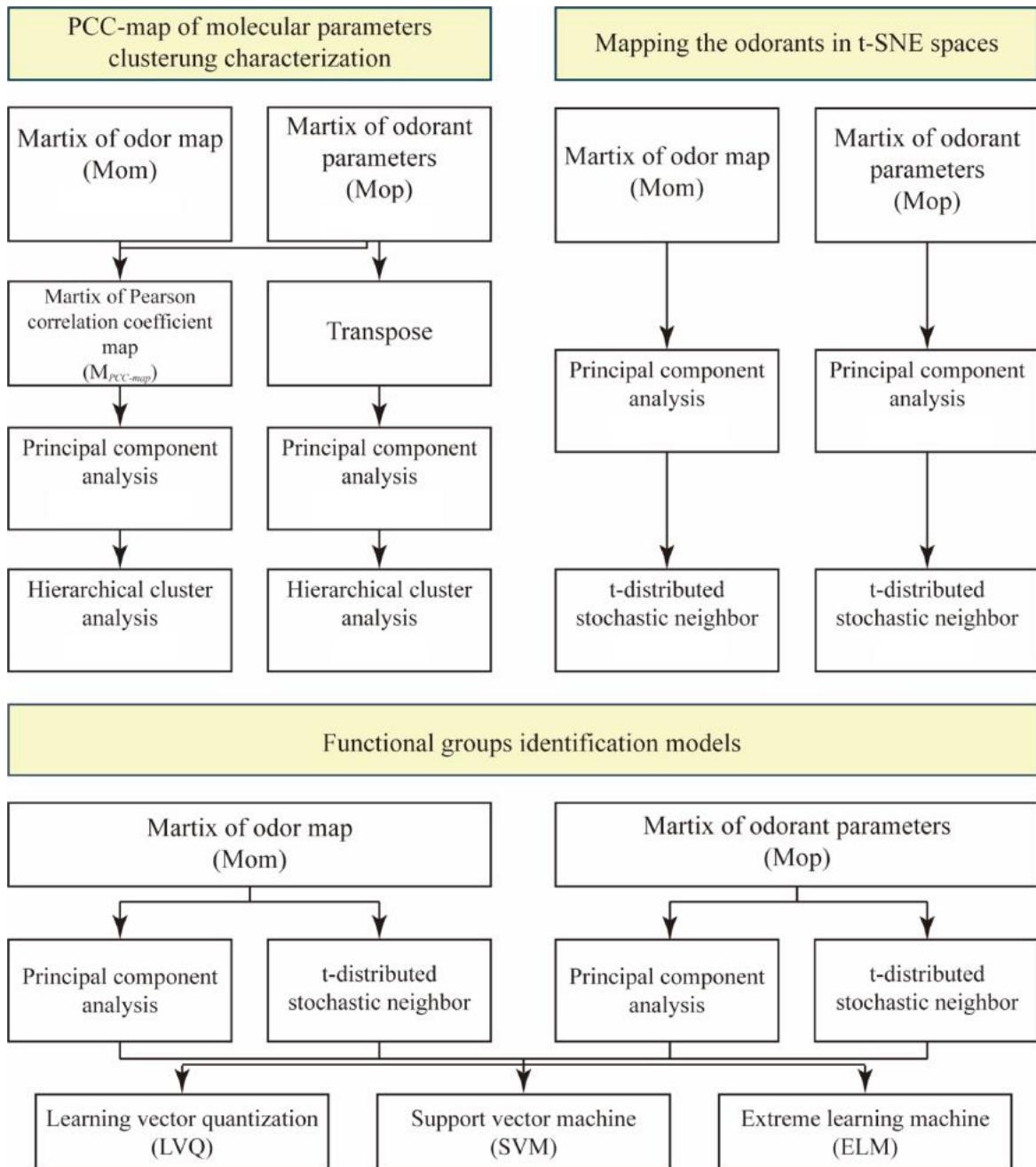


Fig. 2.1 Schematic diagram of data processing.

## 2.2 Materials and method

### 2.2.1 Glomerular activity patterns and molecular parameters

We used glomerular activity patterns (odor maps) from the dorsal part of rat olfaction bulbles. OMs (grey image), chemical abstracts service (CAS) numbers, and functional group

labels were extracted using semi-automatic and manual methods from the Johnson and Leon database. Forty-six MPs for 178 odorants (Table A.1) were determined using the MOPAC and GAMESS packages in ChemBio3D Ultra 11.0 (2008, Cambridge Soft, Massachusetts, USA) by establishing 3D models for odorants based on simplified strings of molecular input-line entry specifications (SMILES). All the parameters used in this study are listed in Table 2.1.

Table 2.1 46 types of molecular parameters extracted by ChemBio 3D.

No.	Molecular parameter	No.	Molecular parameter	No.	Molecular parameter
1	Boiling point	18	Total energy	35	Shape coefficient
2	Critical pressure	19	Dipole	36	Sum of degrees
3	Critical temperature	20	Number of Hbond acceptors	37	Sum of valence degrees
4	Critical volume	21	Number of Hbond Donors	38	Topological diameter
5	Gibbs free energy	22	Ovality	39	Total connectivity
6	Heat of formation	23	Principal moment	40	Total valence connectivity
7	Henry's law constant	24	Elemental analysis	41	Wiener index
8	Ideal gas thermal capacity	25	Molecular weight	42	Core-core repulsion
9	LogP	26	LogS	43	COSMO area
10	Melting point	27	Pka	44	COSMO volume
11	Mol refractivity	28	Balaban index	45	Electronic energy
12	Vapor pressure	29	Cluster count	46	Ionization potential
13	Water solubility	30	Molecular topological index		
14	Connolly accessible area	31	Num rotatable bonds		
15	Connolly molecular area	32	Polar surface area		
16	Connolly solvent excluded volume	33	Radius		
17	Exact mass	34	Shape attribute		

## 2.2.2 Construction of PCC-maps for molecular features

We used the following computational process to generate PCC-maps for investigating the relationship between OMs and MPs.

Step 1. Each  $357 \times 197$  OM was transformed into a  $1 \times 70329$  vector. We created a matrix of odor maps ( $M_{om}$ ) by combining the 178 OMs (gray images) (Figure 2.2 a, Formula 2.1).

$$M_{om} = \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,70329} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,70329} \\ \vdots & \vdots & \vdots & \vdots \\ P_{178,1} & P_{178,2} & \cdots & P_{178,70329} \end{bmatrix}_{178 \times 70329} \quad (2.1)$$

Step 2. We created a similar matrix of odorant parameters ( $M_{op}$ ), which contained 46 molecular parameters from 178 odorants (Figure 2.2 b, Formula 2.2).

$$Mop = \begin{bmatrix} V_{1,1} & V_{1,2} & \cdots & V_{1,46} \\ V_{2,1} & V_{2,2} & \cdots & V_{2,46} \\ \vdots & \vdots & \vdots & \vdots \\ V_{178,1} & V_{178,2} & \cdots & V_{178,46} \end{bmatrix}_{178 \times 46} \quad (2.2)$$

Step 3. We calculated the correlation coefficients between molecular parameters and the gray-level value of each OM pixel. The PCC matrix  $M_{PCC-map}$  was defined as follows:

$$M_{PCC-map} = \{R_{i,j} | R_{i,j} = Cor(Mom(i, :), Mop(j, :)), i = 1, 2, 3, \dots, 46, j = 1, 2, 3, \dots, 70329\} \quad (2.3)$$

where,  $Mom(i, :)$  and  $Mop(j, :)$  indicate the i-th and j-th row vector in Mom and Mop, respectively.  $Cor(x, y)$  was defined in Formula 2.4.

$$Cor(x, y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2.4)$$

where,  $\bar{x}$  and  $\bar{y}$  are the mean values of vector  $\vec{x}$  and  $\vec{y}$ , respectively.  $N$  is the dimension of vector  $\vec{x}$  or  $\vec{y}$  (here,  $N$  was 70329). Thus, the PCC matrix  $M_{PCC-map}$  was established as Formula 2.5.

$$M_{PCC-map} = \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,70329} \\ R_{2,1} & R_{2,2} & \cdots & R_{2,70329} \\ \vdots & \vdots & \vdots & \vdots \\ R_{46,1} & R_{46,2} & \cdots & R_{46,70329} \end{bmatrix}_{46 \times 70329} \quad (2.5)$$

Step 4. We performed HCA based on the Euclidean distances in the latent variables extracted by PCA, and applied Ward's method as a similarity criterion to cluster the 46 molecular parameters into homogeneous groups. The clustering results were then evaluated and analyzed to investigate the relationships between MPs and OMs.

Step 5. We reshaped each row vector of  $M_{PCC-map}$  as a  $357 \times 197$  matrix and obtained the PCC-map for each MP.

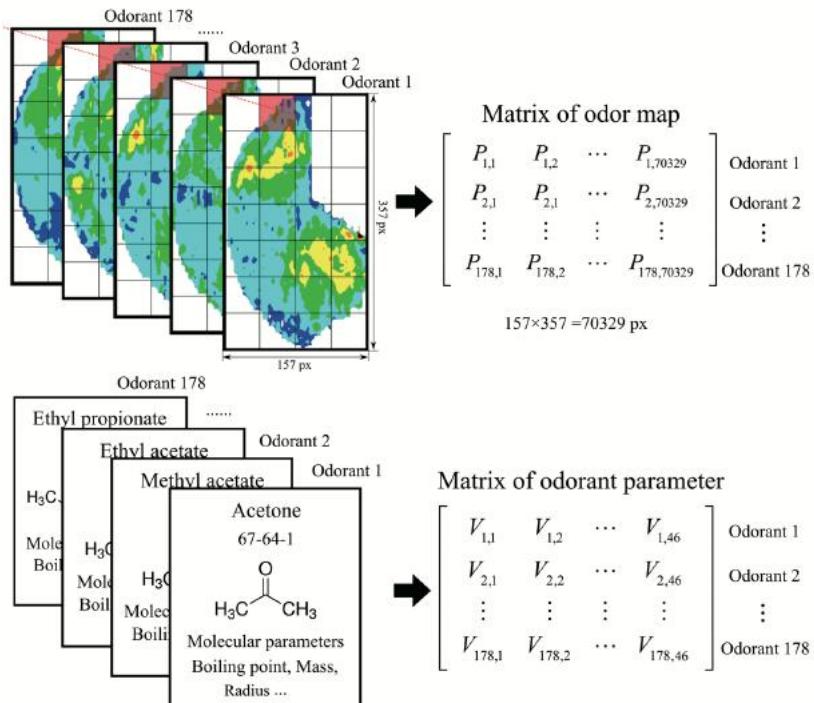


Fig. 2.2 Schematic diagram of obtaining matrix of odor maps (Mom) (a) and molecular parameters (Mop) (b).

### 2.2.3 Selection methods for characteristic variables

#### Principal component analysis (PCA)

PCA is generally used to extract characteristic variables from a high-dimensional data set [146]. PCA can remove linear and duplicated information by constituting principal components (PCs) from original data. The PCs listed first in the output are selected as characteristic variables according to the cumulative contribution rate, while those listed at the end of the output are removed because of noise [147].

#### Barnes-Hut t-distributed stochastic neighbor embedding (t-SNE)

t-SNE is a novel, unsupervised embedding method that has been used to visualize high-dimensional data at a lower dimension [148], and that allows dataset embeddings to be learned. The computational process is as follows.

Step 1. Given a high dimension dataset  $L = \{x_i | x_i \in R^m, i = 1, 2, \dots, N\}$ , where  $x_i$  is a vector ( $1 \times m$ ) of the  $i$ -th sample,  $m$  indicates the total variable number of each sample, and  $N$  indicates the total sample number in the dataset, the function  $d(x_i, x_j)$  is the computed

distance between a pair of samples. Here, we calculated the Euclidean distance,  $d(x_i, x_j) = \|x_i - x_j\|$ .

Step 2. We used Formula 2.6 to calculate the pairwise similarities ( $p_{ij}$ ) between the samples.

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2/2\delta^2)}{\sum_{k \neq l} (-\|x_l - x_k\|^2/2\delta^2)} \quad (2.6)$$

where  $\delta$  is the variance parameter of the Gaussian function.

Step 3. We calculated the similarity ( $q_{ij}$ ) between target values,  $T = \{y_i | y_i \in R^m, i = 1, 2, \dots, N\}$ , in a low-dimensional space ( $m = 2$  or  $3$ ) using a normalized Student-t kernel with one degree of freedom (Formula 2.7).

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|y_k - y_l\|^2)^{-1}} \quad (2.7)$$

Step 4. Based on Kullback-Leibler divergence measuring, we determined the locations of the embedding points ( $Y$ ).

$$C(Y) = KL(P||Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (2.8)$$

The optimal low-dimensional representation  $Y$  was calculated by minimizing  $C(Y)$ . Barnes-Hut t-SNE was employed to establish 2D artificial odor maps. Here, t-SNE was calculated using the R (version 3.2.2) package named "Rtsne" (version 0.1). More detailed explanations of t-SNE processing can be found elsewhere [149].

## 2.2.4 Sample division method

Rational division of sample sets is crucial for improving the test accuracy of models [150]. The test set should include the utmost main information from the original samples. Another benefit of rational sample-set division is that it avoids overlapping in machine-learning models. Here, the Kennard-Stone (KS) algorithm was selected for the sample partition [151]. The details of this process have been described by other research [152, 153]. In the current study, 178 odorants were divided into train and test sets via the KS method. The ratio of samples between train and test sets was 3 to 1. The train set therefore contained 135 samples and the test set contained 43.

### 2.2.5 Modeling methods

#### Learning vector quantization (LVQ)

As a supervised learning algorithm for classification, LVQ is mostly applied for pattern recognition or qualitative analysis based on self-organizing maps [154]. An LVQ network can be optimized by confirming the decision boundaries between neighboring groups. The network contains three layers: an input layer, a competitive layer, and a linear output layer [155]. In this study, LVQ1 was applied to establish the classification models. Additional information about this type of LVQ network can be found elsewhere [156].

#### Support vector machine (SVM)

SVM is a powerful classification model based on statistical learning theory, which has been widely applied in machine vision, image processing, and pattern recognition [157, 158]. By establishing a hyperplane as a decision surface, the positive-examples and counter-examples can be divided such that they are separated by the greatest possible distance. Details regarding SVM have been published elsewhere [159, 160]. In the current study, a radial basis function (RBF) was selected as the kernel function for the SVM model, which was established using the Libsvm (version 2.81) package [161].

#### Extreme learning machine (ELM)

ELM is an efficient single-hidden-layer feed-forward neural network that is widely used for establishing non-linear relationships because of its good performance at generalization [162]. ELM can also overcome some difficulties in traditional learning methods, such as learning rate and epochs [163]. If the number of hidden layer nodes is assigned, the weights between input neurons and hidden neurons can be chosen and fixed randomly [164]. Details regarding the computational process for ELM can be found elsewhere [165].

## 2.3 Results and discussion

### 2.3.1 Molecular parameters and functional group labels

In our previous study, PCA was carried out on the basis of 79 molecular parameters of odorants from the odor-map database [21]. The result indicated that the number of parameters that are important for generating odor maps was not large. Indeed, only 15 parameters showed strong relevance to the first 6 PCs. Based on the above results and limitation of the software (Chembio 3D), in this study we only calculate 46 parameters to carry out the analysis.

The research of Mori reveal that in the rat OB there exist nine independent zones which is corresponding to different functional groups [166]. Other evidence has shown that functional groups affect odor sensation. For example, odorants with the functional group '-COOH' are perceived as smelling like sweat [167]. The above results indicate that functional groups may be a good labels for odorant classification, and the identification of functional groups is a better way to understand odorant sensation. Analysis of variance (ANOVA) was used to search the feature response area for six mainly functional groups firstly ( $p < 0.05$ ). The result was shown in Figure 2.3. It indicated that each functional group has their special region. However, we can also find some overlapped between some functional groups. Therefore, functional group would be a feasible label for odor maps analyzation.

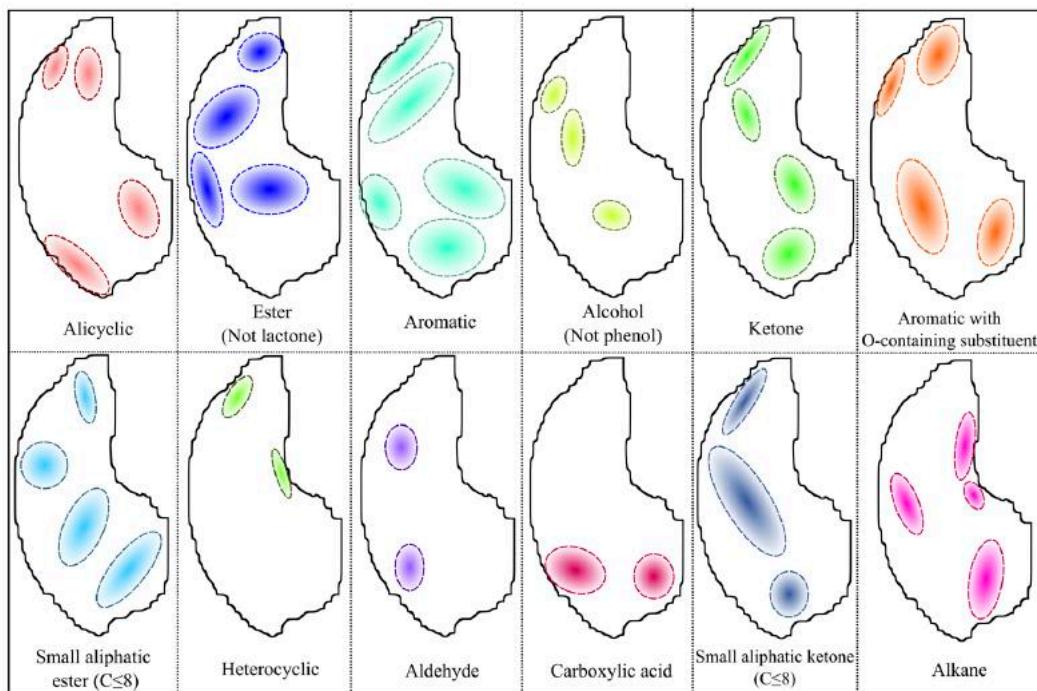


Fig. 2.3 Response regions for 12 types of functional groups are extracted by ANOVA.

In addition, we summarized the functional group label that appeared in the OdorMapDB. The 300 odorants are labeled by seven functional groups in which most show crossed information, especially for complicated odorants with high molecular weights. For odorant classification, the labels should not intersect each other. To simplify the model and to improve prediction accuracy, 14 non-intersection labels collected from 178 odorants were chosen to test our hypothesis (Table 2.2).

Table 2.2 The list of 14 functional group labels considered in this study.

Cluster No.	Functional group description	Sample number
1	Small aliphatic ester (not alicyclic, $C \leq 8$ )	11
2	Primary alcohol	26
3	Alcohol (not phenol)	13
4	Aromatic ester (not lactone)	10
5	Aliphatic ester (not alicyclic)	5
6	Aliphatic or alicyclic hydrocarbon alkane	20
7	Aromatic aldehyde	11
8	Aliphatic or alicyclic hydrocarbon	11
9	Aromatic (only)	15
10	Aldehyde (only)	12
11	Aliphatic ketone (not alicyclic, $C \leq 8$ )	12
12	Carboxylic acid	21
13	Aliphatic or alicyclic ketone ( $C \leq 8$ ) with multiple O-containing function groups	5
14	Aliphatic ester (not lactone) with multiple O-containing functional groups	6

### 2.3.2 Clustering characterization for PCC-maps of molecular parameters

A correlation heat map for the 46 molecular parameters (Figure 2.4) shows that some parameters are linearly related and that some redundant information is included in the molecular parameter matrix. Therefore, before clustering, we performed PCA to reduce the dimensions of the PCC-maps and MPs. Based on Euclidean distances between the first four latent variables extracted by PCA (accumulative contribution = 90.9%, Figure 2.5 a), HCA was performed to investigate the relation between MPs and OMs. The results were organized and depicted by a heat map shown in Figure 2.6 a. The horizontal dendrogram of the heat map shows the 46 MPs are clustered into 7 groups. All the PCC-maps were provided in Figure 2.7. Similar response pattern is shown in each group. It indicated that the MPs in the same group could contribute the similar information to OMs. For example, most of parameters contained energy information are clustered in group 1, and parameters contained polarity information are clustered in group 2. Further, low correlation coefficients are observed for each molecular feature. This finding indicates that the relationships are non-linear, and that one odor receptor could be linked to multiple physicochemical odorant features. These conclusions are supported by Kaeppeler's and Johnson's work [168].

Through the same procedure, HCA was applied to the first 11 PCs (cumulative contribution rate = 91.4%) of the Mop (Figure 2.5 b), and the result is shown in Figure 2.6 b. Groups visualized by these heat maps shared some similarities to the PCC maps, such as cluster 2 and cluster G, cluster 5 and cluster D, and cluster 6 and cluster F. However, some parameters are clustered differently between the two heat maps. This indicates that these parameters are sensitive to olfactory information.

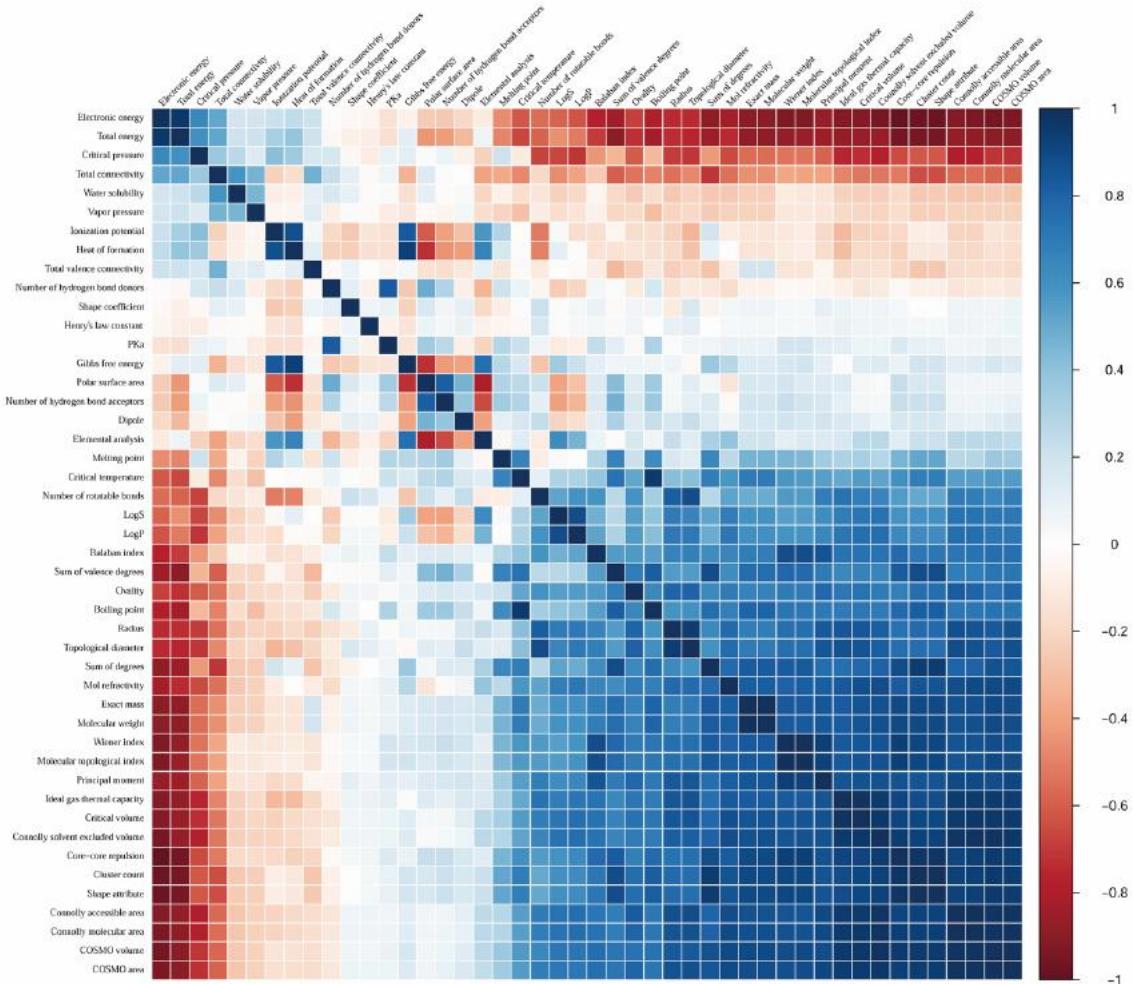


Fig. 2.4 Correlation map for 46 types of molecular parameters.

### 2.3.3 Mapping the odorants in t-SNE space

Next, we investigated the possibility of mapping odorants in 2D space. Before using t-SNE, we always employed PCA to extract vital information from the original matrixes. Generally, when PCs have more than 85% cumulative contribution from the original dataset, these PCs can be used to replace the originals [169]. Here, the first 80 PCs (cumulative contribution rate = 86.0%) for Mom and the first 23 PCs for Mop (cumulative contribution rate = 99.0%) were used as the inputs to the t-SNE analysis (Figure 2.8). Next, the Barnes-Hut t-SNE algorithm was utilized to plot the odor maps in 2D space. In total, 178 odorants from odors with 14 functional groups were mapped into 2D t-SNE space. The initial dimensions were set to 80 for Mom and 23 for Mop, and the perplexity and maximum number of iterations

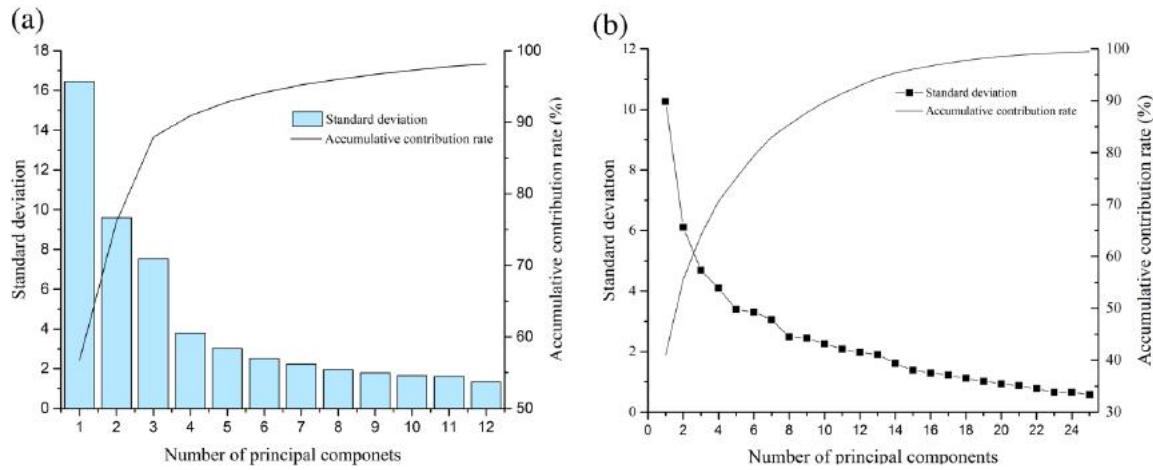


Fig. 2.5 Contribution rates and standard deviations of the first 12 PCs for PCC-maps (a) and the PCC-maps for 7 clusters (b).

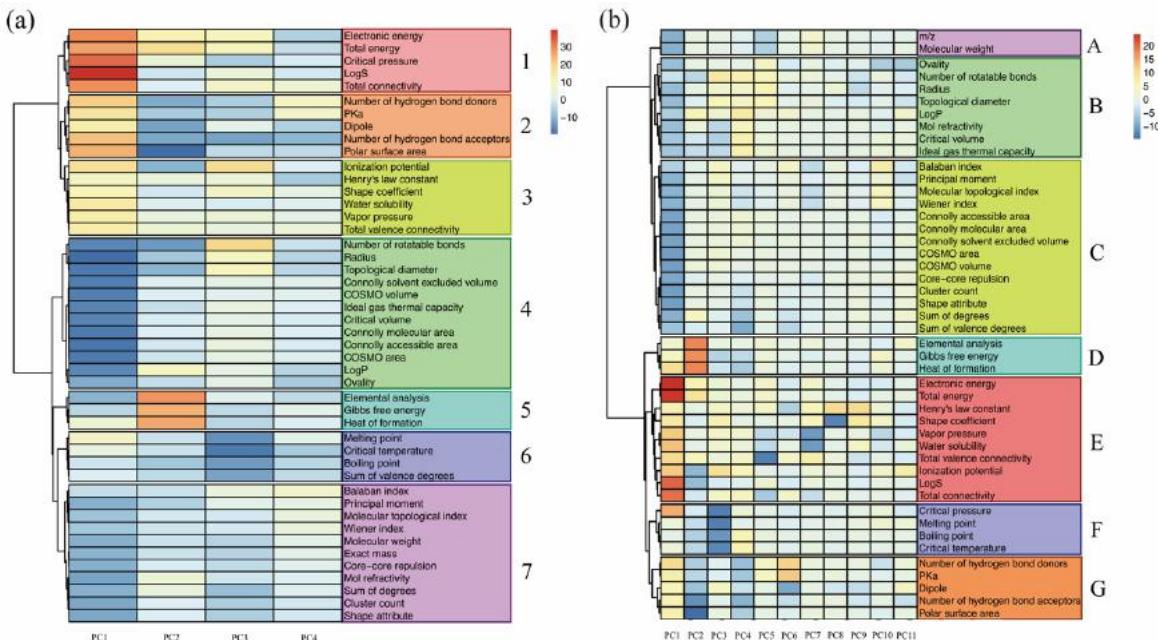


Fig. 2.6 Heat map and hierarchical dendrograms of the PCC-maps (a) and parameters (b) for 46 molecular parameters. Cluster analysis was performed by Ward's method on Euclidean distance of the first 4 PCs for R-maps. Each row indicated one type of molecular parameter, and each column indicated a PC.

were set to 50 and 10000, respectively. Table A.1 shows the details for the data calculated by t-SNE.

The artificial map generated from the olfactory information is shown in Figure 2.9 a. It indicates that 64.04% of samples within the same category are clustered together (Table 2.3).

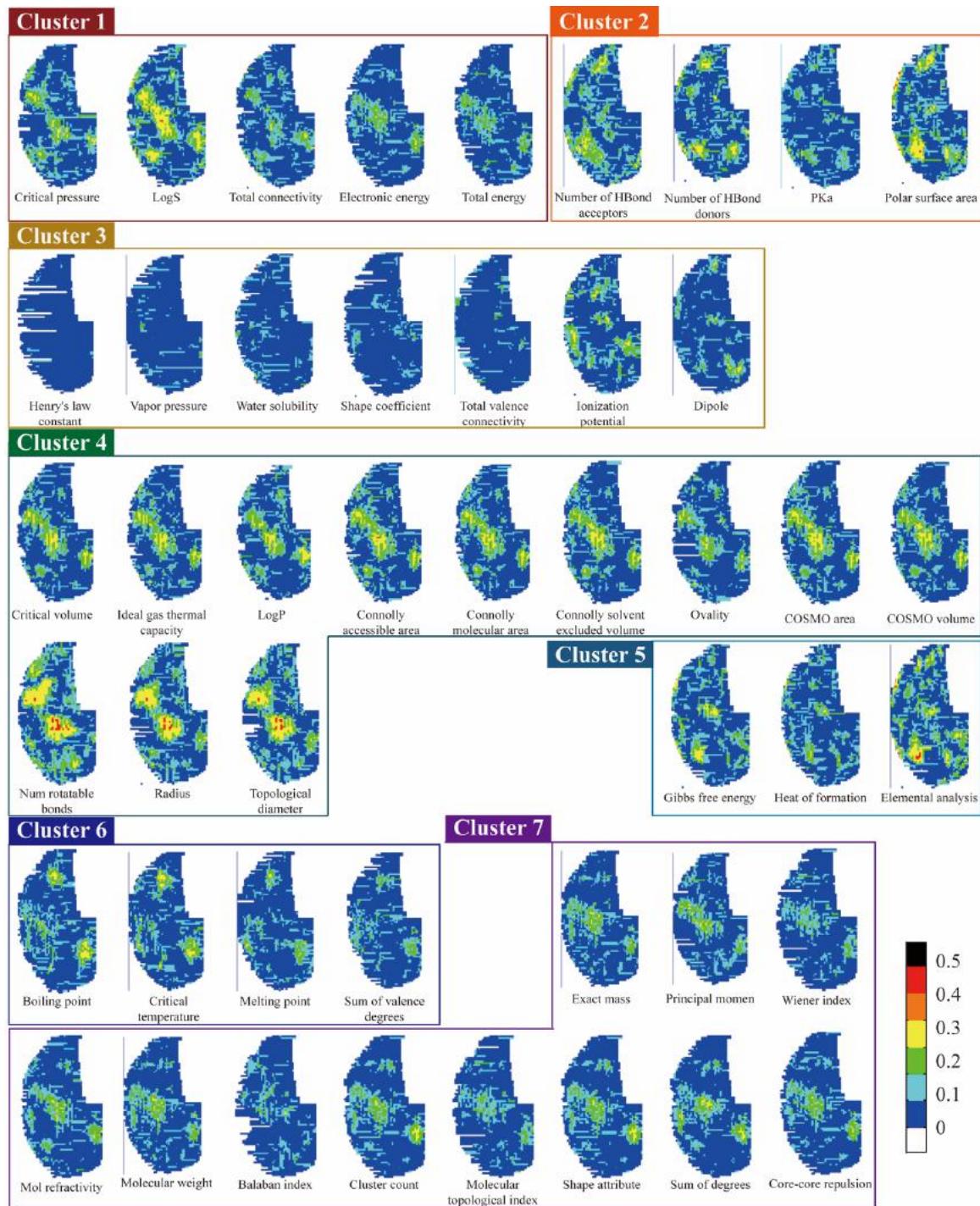


Fig. 2.7 The PCC-maps for 46 molecular parameters. The value on each pixel of a PCC-map indicates the correlation coefficient between a pixel and a molecular parameter. All PCC-maps are clustered seven groups.

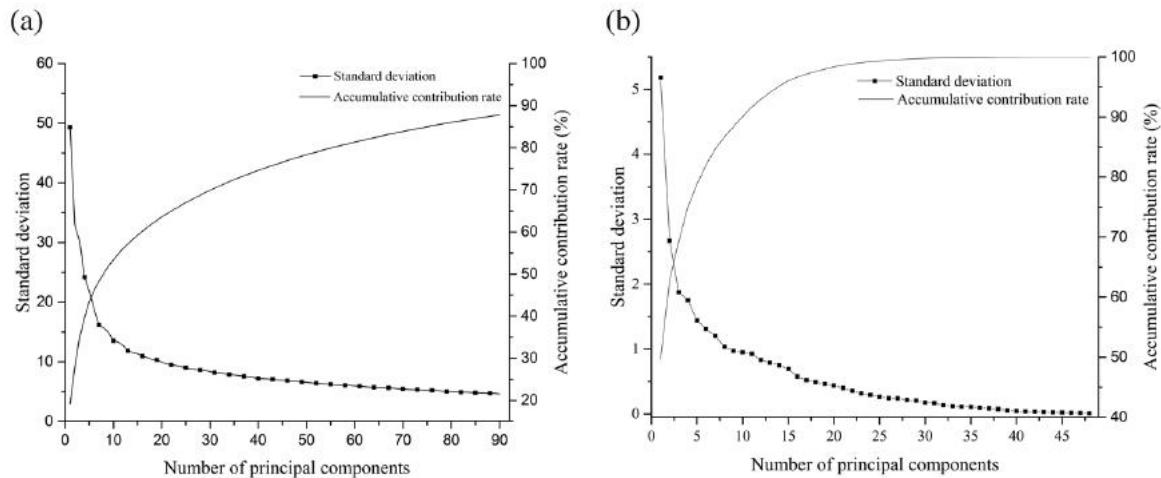


Fig. 2.8 Contribution rates and standard deviations of PCs for olfactory image matrix (a) and molecular parameter matrix (b).

Additionally, samples from the chemical categories small aliphatic ester ( $C \leq 8$ ), aliphatic or alicyclic hydrocarbon, aromatic, and carboxylic acid are clustered in multiple groups. This indicates that carbon-chain length is a vital factor for aliphatic esters (cluster 1) and alcohols (cluster 9), and that branched chains play a role in distinguishing odorants with hydrocarbon alkanes (cluster 6) from those with carboxylic acids (cluster 12). Most odorants classified as aromatic (cluster 3, 4, and 7) are mapped in the left of t-SNE space. However, some aromatic odorants are scattered and not clustered together. We attribute this result to insufficient numbers of samples which cannot completely identify hidden patterns of the molecular structure. Some clusters, such as clusters 4, 7, 6-1, and 10, overlapped in t-SNE space, demonstrating that they would be smelled similarly by a rat. Compared with olfaction information maps, cluster overlapping is observed more in the molecular information maps (Figure 2.9 b). Interestingly, some clusters, such as clusters 1, 9, and 12, are clustered into multiple groups in the olfactory map, but are clustered into a single group in the molecular information map. For example, the olfactory images for odorants in cluster 12-1 are different from those in cluster 12-2 because of the distance between these two groups in the olfaction information map. However, in the molecular information map, odorants in cluster 12 are clustered into only one group. This demonstrates that 46 parameters are not enough for ideal mapping, and that some vital functional group parameters could have been missing from the analysis.

Table 2.3 Summary for the cluster accuracies for 2D artificial maps established by OMs and MPs.

Cluster No.	Sample number	Olfactory images		Molecular parameters	
		Correct number	Accuracy	Correct number	Accuracy
1	11	9	81.82%	8	72.73%
2	26	11	42.31%	12	46.15%
3	13	12	92.31%	13	100.00%
4	10	6	60.00%	7	70.00%
5	5	4	80.00%	5	100.00%
6	20	13	65.00%	15	75.00%
7	11	7	63.64%	7	63.64%
8	11	7	63.64%	8	72.73%
9	15	13	86.67%	9	60.00%
10	12	4	33.33%	7	58.33%
11	12	6	50.00%	8	66.67%
12	21	12	57.14%	17	80.95%
13	5	5	100.00%	4	80.00%
14	6	5	83.33%	4	66.67%
Total	178	114	64.04%	124	69.66%

### 2.3.4 Functional group-identification models

To assess the potential for OM and MP to classify odorants, we applied LVQ, SVM, and ELM classification methods to each and compared the results. The features acquired by PCA or t-SNE were set as the input data for 178 odorants, and the 14 types of functional group labels were set as the output of the models.

#### LVQ Models

To acquire optimized LVQ models, training epoch, learning rate, and learning goal were set to 500, 0.1, and 0.05, respectively. As a necessary parameter for LVQ networks, the number of hidden nodes is always set by trial and error. Here, the range of hidden layer nodes was 1 to 100 (Figure 2.10). We chose the smallest numbers because these were associated with greater accuracy. The numbers of hidden layer nodes for OM-PCA, OM-tSNE, MP-PCA, and MP-tSNE were 35, 15, 34, and 98, respectively (Table 3.2). The accuracy of functional group identification for train and test sets using the four types of datasets are listed in Table 3.3. The results suggest that models established by PCA higher accuracies than those established by t-SNE. Compared with t-SNE, LVQ thus more suitable for the dataset obtained by PCA. However, LVQ identification of odorants generally poor, with the highest accuracy being only 69.77%.

Table 2.4 The optimal training parameters of LVQ, SVM and ELM models.

Input data	LVQ			SVM			ELM		
	Pretreatment methods	Input layer nodes	Hidden layer nodes	Output layer nodes	c	g	Input layer nodes	Hidden layer nodes	Output layer nodes
Odorant descriptor	PCA	80	35	14	1.000	0.100	80	54	14
Odor maps	t-SNE	2	15	14	1.000	0.100	2	51	14
Molecular parameters	PCA	23	34	14	5.657	0.354	23	21	14
	t-SNE	2	98	14	22.627	11.314	2	47	14

Table 2.5 Identification accuracies of LVQ, SVM and ELM models.

Odorant descriptor	Pretreatment methods	Modeling approach	Accuracy (%)		Mean
			train set	test set	
Odor maps	PCA	LVQ	72.09	68.89	70.49
		SVM	29.63	18.60	24.12
		ELM	94.81	93.02	93.92
Molecular parameters	t-SNE	LVQ	14.81	13.95	14.38
		SVM	14.81	13.95	14.38
		ELM	72.59	90.70	81.65
	PCA	LVQ	65.93	69.77	67.85
		SVM	93.33	90.70	92.02
		ELM	89.63	93.02	91.33
	t-SNE	LVQ	21.48	23.26	22.37
		SVM	82.22	83.72	82.97
		ELM	86.67	95.35	91.01

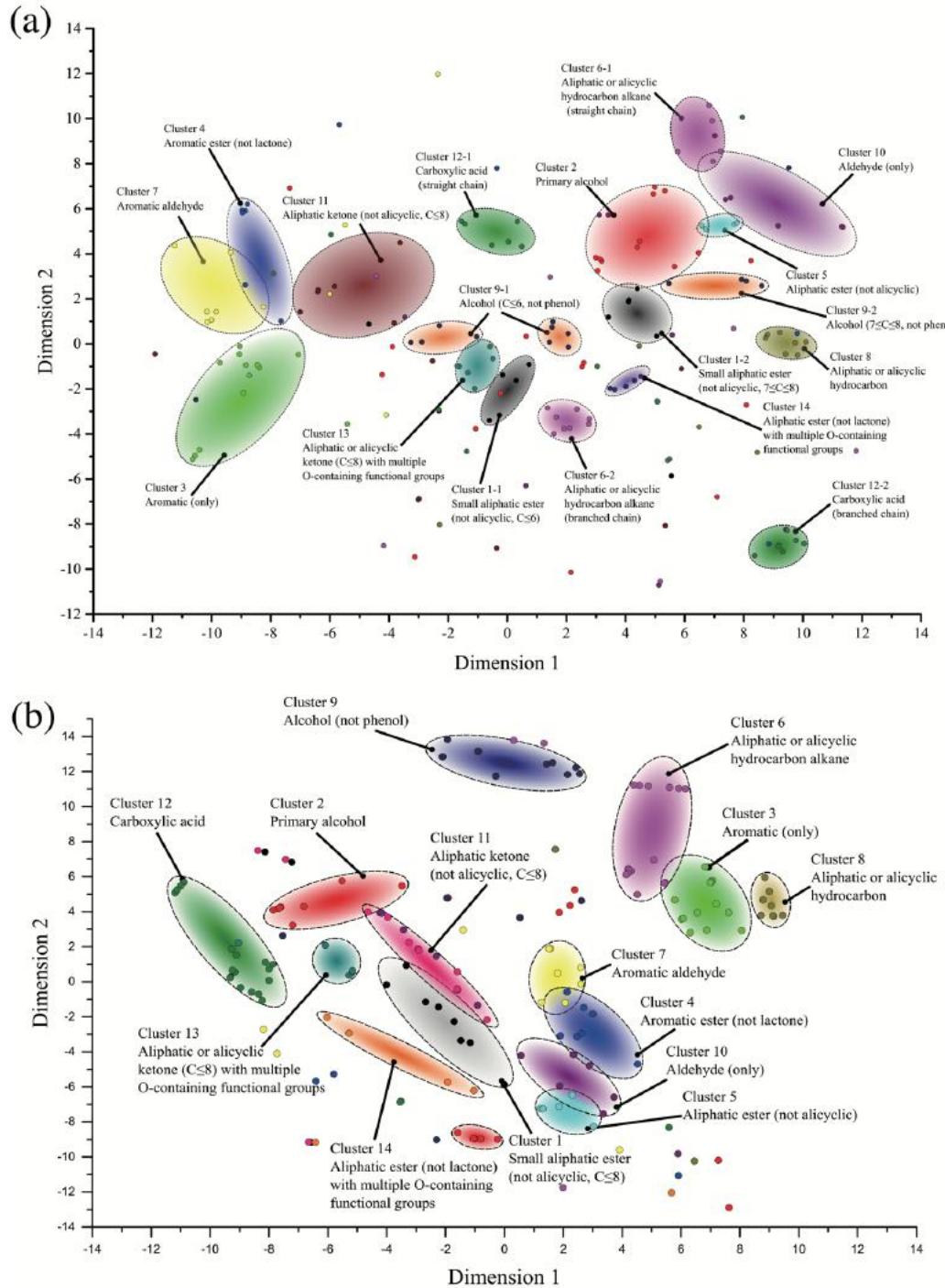


Fig. 2.9 Odorant clustering generated in OM (a) and MP (b) spaces by using t-SNE method. Dimension 1 and 2 indicates 2 values calculated by t-SNE. Each point indicates an odorant.

### SVM Models

The SVM kernel function was set to RBF, and 5-fold cross test was performed to obtain the penalty factor ( $c$ ) and the RBF parameter ( $g$ ). The parameters for the SVM models are listed

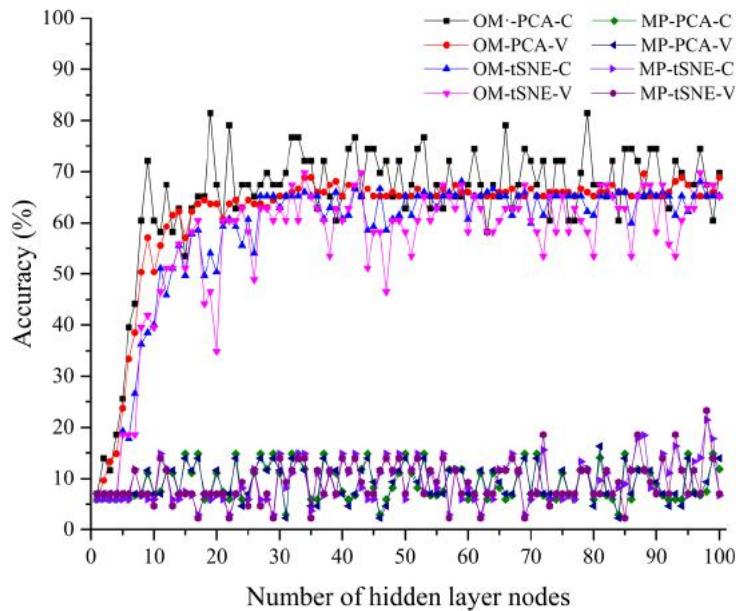


Fig. 2.10 The accuracies under different numbers of hidden layer nodes for LVQ networks in the train (C) and test (V) set. OM-PCA and OM-tSNE indicated that the models established by the PCA or tSNE of odor maps (OM) for odorants. MP-PCA and MP-tSNE indicated the models established by the PCA or tSNE of molecular parameters (MP) for odorants. The optimal number for OM-PCA, OM-SNE, MP-PCA and MP-tSNE are 35, 15, 34 and 98, respectively.

in Table 3.2, and the accuracy of each SVM model is shown in Table 3.3. For the train set, the accuracies for the MP-PCA-SVM and MP-tSNE-SVM models are 93.33% and 82.63%, respectively, which higher than those observed for the OM-PCA-SVM and OM-tSNE-SVM models (29.63% and 14.81%, respectively). Similar results are shown for the test sets. This demonstrates that the SVM models established by molecular information performed better than those established by olfactory information.

### ELM Models

We set the excitation function for the ELM models to "sig". The number of hidden layer nodes for ELM models were also determined by trial and error. Here, one model was trained 1,000 times to overcome the randomness of ELM models. The numbers of hidden layers were chosen based on the average accuracy across the 1,000 models (Figure 2.11), and equaled 54, 51, 21 and 47 for the OM-PCA, OM-tSNE, MP-PCA, and MP-tSNE models respectively (Table 3.2). The accuracy of functional group identification for train and test

sets using different datasets is shown in Table 3.3. The results show that for the train set, the OM-PCA-ELM model was more accuracy (94.81%) than the other models (OM-tSNE-ELM, 72.59%; MP-PCA-ELM, 89.63%; MP-tSNE-ELM, 86.67%). For the test set, all models were more than 90% accuracy. The MP-tSNE-ELM model was the most accuracy (95.35%), followed by the OM-PCA-ELM and MP-PCA-ELM models (93.02%), and the OM-tSNE-ELM model (90.70%). Thus, the OI-tSNE-ELM, MP-PCA-ELM, and MP-tSNE-ELM models were more accuracy for the test set than for the train set. Although test-set accuracy is generally lower than train-set accuracy, the reverse is possible if most of the represented samples were chosen in the train set that established the model [170, 171].

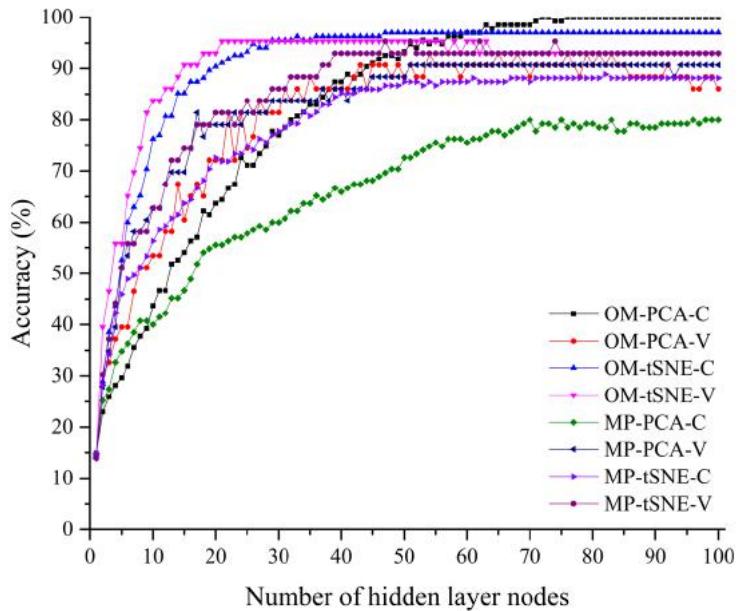


Fig. 2.11 The accuracies under different numbers of hidden layer nodes for ELM networks in the train (C) and test (V) set. OM-PCA and OM-tSNE indicated that the models established by the PCA or t-SNE of odor maps (OM) for odorants. MP-PCA and MP-tSNE indicated the models established by the PCA or t-SNE of molecular parameters (MP) for odorants. The optimal number for OM-PCA, OM-SNE, MP-PCA and MP-tSNE are 54, 51, 21 and 47, respectively.

### Identification performance for different models

Comparing the three types of models, ELM was the best at functional group identification. This was likely because ELM does well in generalization. This is consistent with other studies showing good prediction performance by ELM [172, 173]. We also found that SVM

performed well with MP-PCA (93.33% and 90.70%) and MP-tSNE (82.22% and 83.72%) datasets. However, poor results were observed for OM-PCA-SVM (29.63% and 18.60%) and OM-tSNE-SVM (14.81% and 13.95%) models. This suggests that SVM models are more suitable for establishing functional group-identification models based on molecular parameters. Although MP-tSNE-ELM had the highest accuracy (95.35%) for the test set, its accuracy in the train set was only 86.67%. The model established by OM-PCA-ELM presented acceptable identification accuracies for both the train (94.81%) and test (93.02%) sets. Therefore, we suggest that OM-PCA-ELM is the optimal model for identifying functional groups of odorants. Compared with other datasets, for functional group identification in odorants, the features extracted from odor maps via PCA contained more information than the 46 molecular parameters.

### 2.3.5 Discussion

An odorant can be described by multiple molecular parameters or by a neuronal response pattern in the mammal OB. Investigating the relationship between molecular features and OB-derived images of neuronal activity is a challenge in developing sensor-based machine olfaction. Many analyses have been carried out focusing on the classification of odor descriptors or molecular feature [28, 174, 175]. The importance of the olfactory information in OB is not fully understood and thus less attention has rarely been paid to the image analysis of OB. In our previous study that was based on PCA, 15 key parameters were obtained by evaluating the correlations between the molecular parameters and PCs [21]. However, only six PCs were analyzed which might not be enough for describing all odorant features due to the complexity and nonlinearity of the dataset. In the present study, t-SEN was applied for the high-dimensional data analysis. t-SNE was considered to be able to provide a competitive performance in dimensionality reduction if compared with conventional methods such as PCA and multi-dimensional scaling. The result shown in this study confirm that t-SNE can be used as an effective approach to establish relationship between molecular parameters, odor map and functional groups. In addition, some molecularly information include size, polarity or carbon chain, shown higher correlations to odor maps. Therefore, these molecular parameters should be considered firstly for developing odor sensors.

## 2.4 Conclusion

In this chapter, we accumulated and analyzed 178 odor maps from the LJ database and their 46 types of molecular features. PCC-maps for molecular parameters turned out to be clus-

tered in seven groups, and the parameters in each group had a similar effect on the images of olfactory responses. Low correlation coefficients indicated that the relationship between molecular features and the odor map responses was not linear. All odorants were mapped in 2D space, and similar odorants were clustered together. Compared with the cluster map generated by molecular parameters, olfactory images contained more detailed information, such as the lengths of carbon and branched chains. We tested how well different models could identify functional groups when the models were established based on olfactory information or molecular parameters. The results showed that OM-PCA-ELM was the optimal model. Although classifiers with molecular parameters as feature quantity were weaker than those with odor map as feature quantity, a comparative model could be established if it was based on enough molecular features. This research would be applied in developing biology olfaction based odor recognition sensors.

# **Chapter 3**

## **Prediction of odor perception from molecular parameters**

### **3.1 Introduction**

In this chapter, the relation between molecular parameters (MP) and odor descriptors (OD) was discussed. In addition, we test the possibility that a machine-learning based prediction model could be used to replace the human panelist in gas chromatography-olfactometry (GC-O) (Figure 3.1). As illustrated in Figure 3.2, after the GC effluent is identified by mass spectrometry, its MPs can be transferred by a cheminformatics software and inputted into a classifier system in which each classifier is labeled by a specific ODs (a word like "sweet", "green", "fruity", and "herbaceous", etc.). After the true or false classification, the system can output the sensory information of the GC effluent, which may consist of single OD (such as sweet) or multiple ODs (such as sweet and green). These ODs predicted by these models would be regarded as references for odor sensory information evolution.

A flavor and fragrance database (Sigma-Aldrich, 2016) that includes 1026 odorants and 10 ODs was considered in this study. The physicochemical MPs were acquired via a cheminformatics software. The features of the MPs were extracted using either unsupervised (principle component analysis: PCA) or supervised (Boruta: BR) approaches. Ten typical ODs with high occurrence frequency in the database were selected to establish the models. Different machine learning algorithms, including support vector machine (SVM), random forest (RF), and extreme learning machine (ELM), were used and their prediction results were compared. Finally, golden delicious apple GC-MS data was employed to prove the feasibility of the model established in present study. A Boruta-SVM model showed high accuracy in ODs prediction, which indicates the possibility for machine-learning based GC-O.

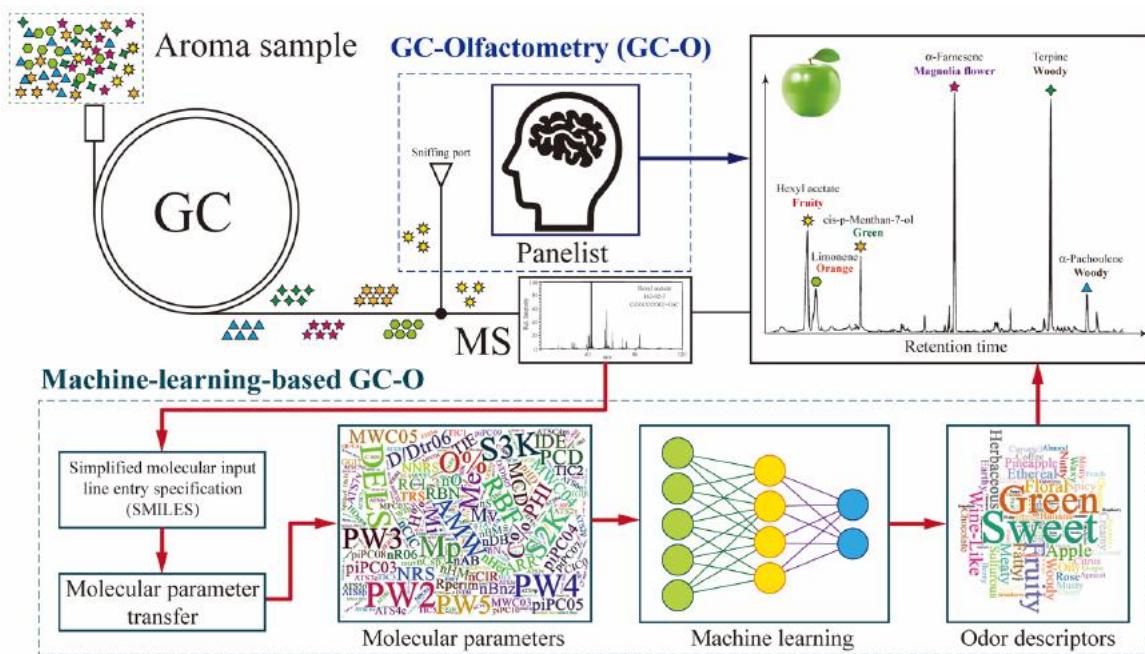


Fig. 3.1 Machine-learning-based olfactometer

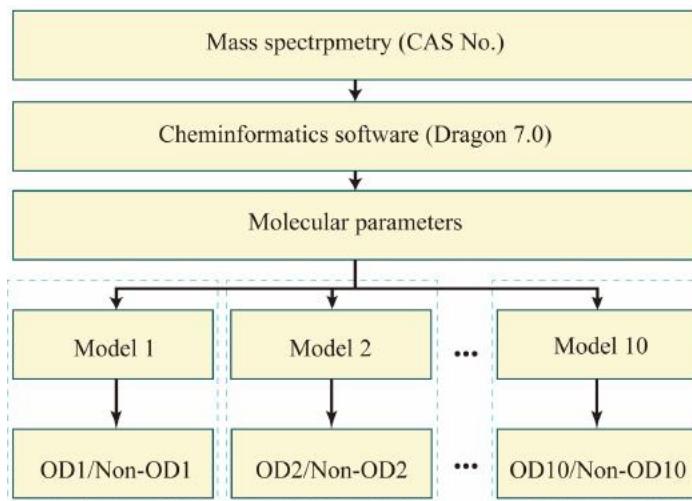


Fig. 3.2 Concept diagram to predict odor descriptors using molecular parameters

## 3.2 Materials and method

### 3.2.1 Odor data collection

Simplified molecular input line entry specification (SMILES) were obtained by both semi-automatic and manual methods from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) according to the CAS number of the odorant molecules recorded in the Flavors & Fragrances

database (Sigma-Aldrich, 2016). The SMILES strings were imported into the Dragon chemoinformation software (version 7.0, Kode, Italy) to compute the physicochemical parameters. The calculation afforded 5270 parameters with various values for each odorant molecule. It was found that most of parameters (around 4200) were assigned with not applicable (NA). We removed these parameters with "NA", and finally got a parameter matrix with 1006 MPs. All MPs were normalized and centered for further processing.

### 3.2.2 Data analysis

The data analysis process is shown in Figure 3.3. The dataset for odor prediction is a typical imbalanced dataset because the class distribution of the positive samples (minor samples with specified ODs labels) and the negative samples (major samples with non-specified ODs labels) is not uniform. Here, synthetic minority oversampling technique (SMOTE) was employed to overcome the imbalance problem [150]. The minority class was over-sampled at 300% of its original size and the majority class was under-sampled to obtain a balanced dataset. Afterwards, the sample pool was divided into training and test sets with a 3:1 ratio using the Kennard-Stone (KS) algorithm [151, 176]. Sample size details for each OD are listed in Table 3.1. The unsupervised feature combination method (PCA) and supervised feature selection method (Boruta) were performed to extract kernel information to enhance the performance of the classification frameworks [177, 178]. SVM, RF, and ELM classification algorithms were applied to predict ODs. The optimal model was determined by considering the accuracies of the training and test sets. As the last step, the F1 score based on precision and recall was used to verify the performance of the optimal model [179].

### 3.2.3 Model validation

For validation the OD prediction models developed in present study, volatile compounds identified from actual sample analyzed were employed. Arvisenet and his team studied primary compounds from Golden Delicious apples' VOCs by a solid-phase micro extraction (SPME) fiber and GC/MS analyzation [180]. Their results indicated that thirty compounds include 13 esters, 9 alcohols, 5 aldehydes, 1 ketone, 1 phenol, and (E,E)-R-farnesene, which would be considered as primary. Based on the MPs calculated by Dragon 7.0, the optimal model (BR-C-SVM) established in present study was applied for predicting their ODs. Compared with the ODs reported in other papers or databases, the OD prediction models developed in this study would be evaluated.

Table 3.1 Data sets, division of samples, principal components, and molecular parameters <sup>a</sup>.

Odor descriptor	Original dataset			SMOTE processed dataset		Division of samples by KS <sup>b</sup>		Number of PCs	MPs labeled by BR method <sup>c</sup>		
	P sample	N sample	F:T	P sample	T sample	Train set	Test set		Confirmed	Tentative	Rejected
Sweet	198	828	4.18:1	792	891	1262	421	260	180	206	620
Green	192	834	4.34:1	768	864	1224	408	267	211	532	532
Fruity	133	893	6.71:1	532	598	847	283	238	214	205	587
Floral	81	945	11.67:1	324	364	516	172	201	118	136	752
Meaty	80	946	11.83:1	320	360	510	170	201	141	208	657
Wine-like	81	945	11.67:1	324	364	516	172	200	95	126	785
Apple	76	950	12.50:1	304	342	484	162	188	133	109	764
Fatty	75	951	12.68:1	300	337	477	160	199	116	155	735
Woody	74	952	12.86:1	296	333	471	158	189	129	132	745
Herbaceous	72	954	13.25:1	288	324	459	153	188	122	177	707

<sup>a</sup> Original and synthetic minority oversampling technique (SMOTE)-processed data sets are described. P (positive sample) indicates the number of samples with the specific OD label. N (negative sample) indicates the number of samples with the nonspecific OD label.

<sup>b</sup> Divided by use of the Kennard-Stone (KS) algorithm.

<sup>c</sup> Molecular parameters labeled by the Boruta method.

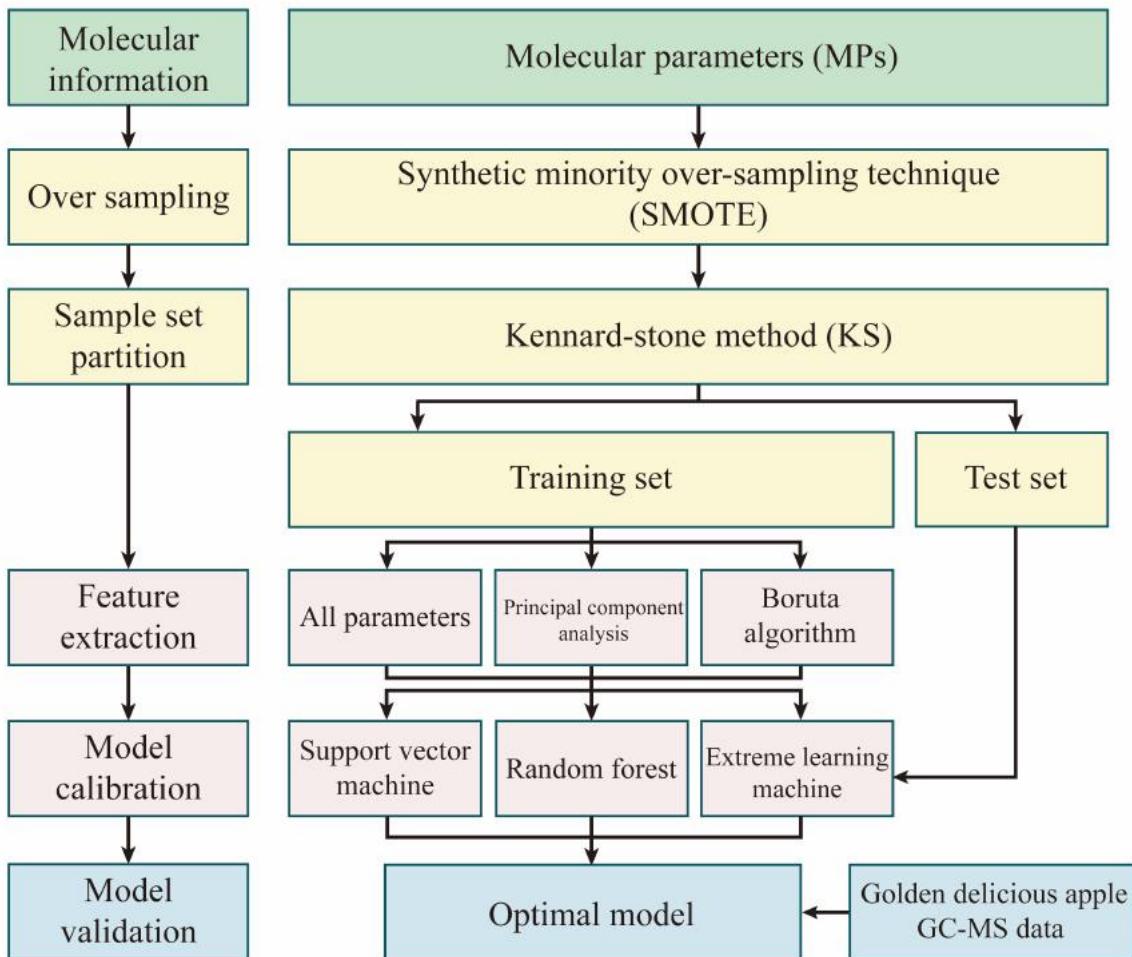


Fig. 3.3 Data processing diagram of prediction models calibration and validation

### **3.3 Results and discussion**

### 3.3.1 Odor descriptors

It is well known that for machine learning, the larger the sample size is, the higher the model accuracy tends to be. For odor prediction, an optimum database should have an appropriate number of odorant molecules and ODs. Up to now, there are a number of odor databases that have been reported and analyzed. Recently, Kumar et. al., carried out a comprehensive statistical analysis of 5 main odor databases which include Flavornet, GoodScents, Leon & Johnson, Sigma-Aldrich, and SuperScent. One problem of these databases is the sparseness of the data distribution because an odorant molecule can be described by a varying number of ODs, but very few molecules are described by a large number of ODs in the databases. The statistical results of Kumar indicate that the Sigma-Aldrich database possesses both a rela-

tively larger number of odorant molecules and larger average number of ODs per molecule, and thus leads to the highest average occurrence of ODs. In view of this characteristic, the Flavors & Fragrances database of Sigma-Aldrich (2016), which has been upgraded to 1026 odorant molecules and 160 ODs, was adopted and analyzed in the present study. Detailed information about the 160 ODs is listed in Table B.1. Figure 3.4 summarizes the 20 ODs that occurred most frequently in the database. The descriptor of "Sweet" is represented by approximately 200 odorants while the descriptor of "Rose" is represented by approximately 50 odorants. Considering that a badly-established model could result from insufficient sampling, only the top 10 ODs were used to establish our prediction models. The 10 descriptors include sweet, green, fruity, floral, meaty, wine-like, apply, fatty, woody, and herbaceous. The minimum number of samples (herbaceous) is over 70. This sample size may help ensure accuracy of the prediction models.

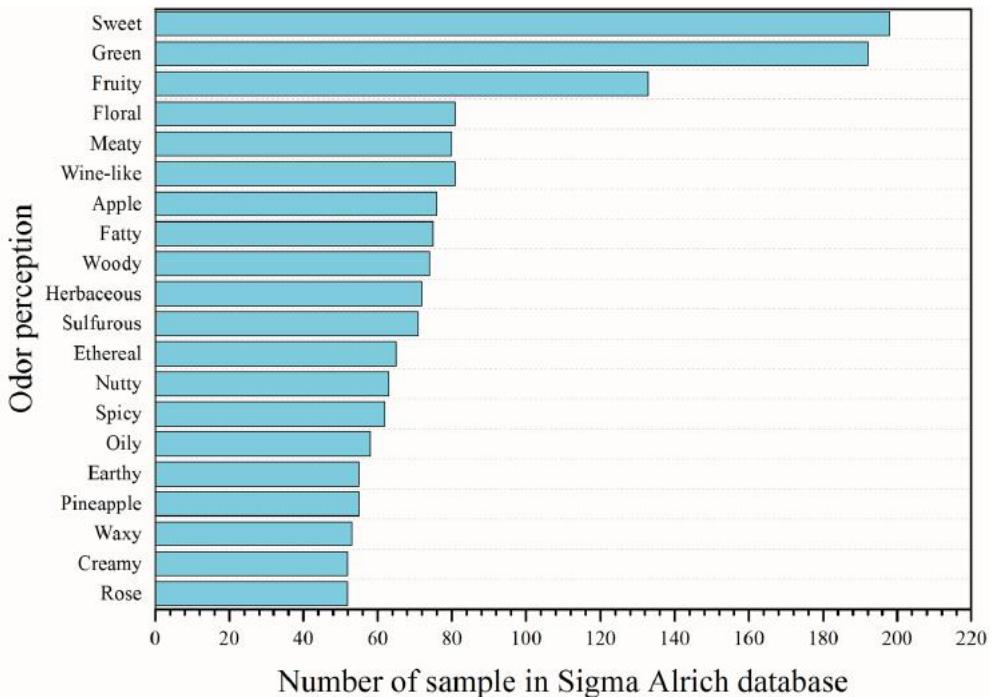


Fig. 3.4 The most 20 frequent ODs in Sigma-Aldrich database. Considered the sample size for model calibration, the firstly ten odor descriptors were considered in this study.

### 3.3.2 Feature extraction

As machine learning aims to deal with larger, more complex questions, the extraction of relevant features for data representation data is a critical problem within model calibration [181]. It has been reported that machine learning algorithms exhibit a decrease of accuracy

when the number of variables is significantly higher than an optimal number [182]. Consequently, before model calibration, PCA and BR were employed using unsupervised and supervised methods, separately, to extract features from all the MPs, and their effects were evaluated.

PCA was first performed to remove redundant information (Figure B.1). To avoid loss characteristic information from the original data set, PCs with accumulative contributions of 99.99% were selected. Table 3.1 lists the number of PCs for 10 ODs. BR was used to find useful features of each OD. By BR, 1006 MPs were labeled as 'confirmed', 'tentative', or 'rejected' (Table 3.1). In this research, MPs labeled 'confirmed' or 'tentative' (BR-CT), and labeled 'confirmed' only (BR-C) were used in further processing. Features selected by BR for the 10 ODs are shown in Figure 3.5. This illustrates that although an MP may be labeled as 'confirmed' for one OD, the MP could be regarded as a useless feature for other ODs. This indicates that ODs could be used to describe various dimensions for an odorant. It can be interpreted that some MPs are associated with some appointed functional groups of a molecule, and functional groups are related to ODs.

### 3.3.3 Model calibration

#### Support vector machine model

SVM Model. The 5-fold cross validation was applied to select penalizing factor ( $c$ ) and RBF kernel parameter ( $g$ ) for developing SVM models in this research. By multiple attempts, the range of  $c$  was set from  $2^0$  to  $2^{-20}$  (the range of  $g$  was set from  $2^{-10}$  to  $2^{10}$ ) with an increment of  $2^{0.5}$ . The optimal values for SVMs were confirmed by the highest accuracy in all combination of  $c$  and  $g$  (Figure B.2, B.3, B.4, B.5). The selected results for ten ODs are listed in Table 3.2. The ODs identification accuracies for calibration set and validation set of AP-SVM, PCA-SVM, BR-CT-SVM and BR-C-SVM are shown in Table B.2, B.3. The results indicated that all the models established by SVM had a similar average accuracy:  $96.83 \pm 1.7\%$  for AP-SVM,  $96.89 \pm 1.9\%$  for PCA-SVM,  $97.19 \pm 0.93\%$  for BR-CT-SVM and  $96.10 \pm 2.8\%$  for BR-C-SVM. However, considered the complexity of model, the BR-C-SVM would be regarded the optimal model than others in present study.

#### Random forest model

To establish random forest (RF) models, two parameters: the number of trees ( $n_{tree}$ ) and the number of features ( $m_{try}$ ) need to be optimized. Although adding more trees will not cause over-fitting, it will increase the model complexity. Therefore, a sufficient number of trees was needed to establish RF models. Considered the out of bag error and test error, the optimal

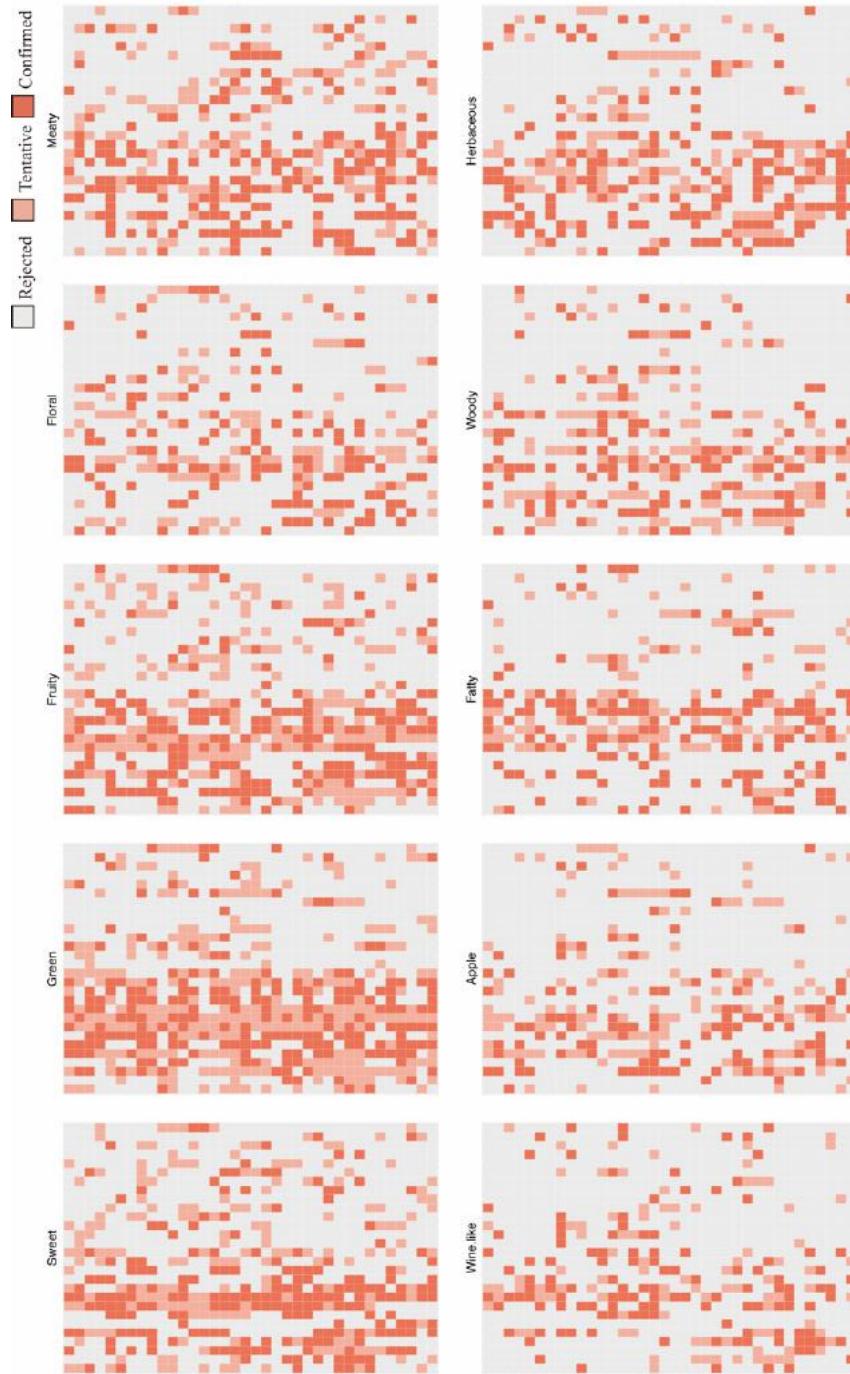


Fig. 3.5 MPs selection based on BR method for the ten ODs. 1006 MPs were arranged as a matrix (36 × 28). Each grid indicated one MP. MPs labeled as 'rejected', 'tentative' or 'confirmed' were colored. Here, the features labeled 'tentative' and 'confirmed' (BR-CT) and only labeled 'confirmed' (BR-C) were used for calibrating models. The number of MPs for each label were listed in Table 3.1.

Table 3.2 Modeling parameters for SVM, RF, and ELM calibration<sup>a</sup>.

Odor descriptor	AP				PCA				BR-CT				BR-C							
	c	g	$m_{try}$	$n_{tree}$	c	g	$m_{try}$	$n_{tree}$	c	g	$m_{try}$	$n_{tree}$	c	g	$m_{try}$	$n_{tree}$				
Sweet	1.414	0.006	46	187	660	1.414	0.008	160	126	794	1.414	0.011	96	39	575	1.414	0.022	90	52	387
Green	2.828	0.004	186	34	575	2	0.008	27	97	796	2	0.008	124	56	633	5.657	0.022	103	32	440
Fruity	2.828	0.008	86	90	537	4	0.008	128	91	626	1.414	0.011	99	32	508	256	0.022	54	25	399
Floral	1.414	0.006	36	93	320	1.414	0.006	21	133	792	1	0.022	34	53	230	1.414	0.011	28	218	204
Meaty	1	0.001	346	75	326	1	0.001	21	59	795	2	0.006	109	31	321	2.828	0.004	91	24	271
Wine-like	4	0.006	26	32	347	4	0.006	130	170	786	2	0.022	11	36	309	1.414	0.088	75	66	322
Apple	1.414	0.011	196	54	312	1.414	0.011	128	37	795	1.414	0.022	172	57	244	1.414	0.088	123	36	335
Fatty	8	0.002	666	27	255	8	0.002	129	106	786	4	0.006	51	42	220	1.414	0.031	96	34	180
Woody	4	0.008	46	92	295	2.828	0.008	69	107	797	1.414	0.022	121	61	270	1	0.044	79	64	217
Herbaceous	4	0.008	36	56	328	8	0.011	78	87	768	2	0.044	29	74	305	2.828	0.063	82	20	724

<sup>a</sup> AP, all parameters; PCA, principal component analysis; BR-CT, confirmed and tentative from Boruta method; BR-C, confirmed only from Boruta method; c, penalizing factor for SVM models; g, RBF kernel parameter for SVM models;  $m_{try}$ , number of features for RF models;  $n_{tree}$ , number of trees for RF models;  $n_{hidden}$ , number of hidden nodes for ELM models.

$n_{tree}$  (Figure B.6, B.7, B.8, B.9) and  $m_{try}$  (Figure B.10, B.11, B.12, B.13) were determined. The optimal modeling parameters of the RF models are listed in Table 3.2. The overall accuracies of the best RF models under AP, PCA, BR-CT, and BR-C datasets are shown in Figure 3.6. In summary, PCA-RF showed a better average accuracy ( $92.79 \pm 1.63\%$ ) than the AP-RF ( $90.62 \pm 1.26\%$ ), BR-CT-RF ( $90.50 \pm 1.21\%$ ), and BR-C-RF ( $90.61 \pm 1.85\%$ ) models.

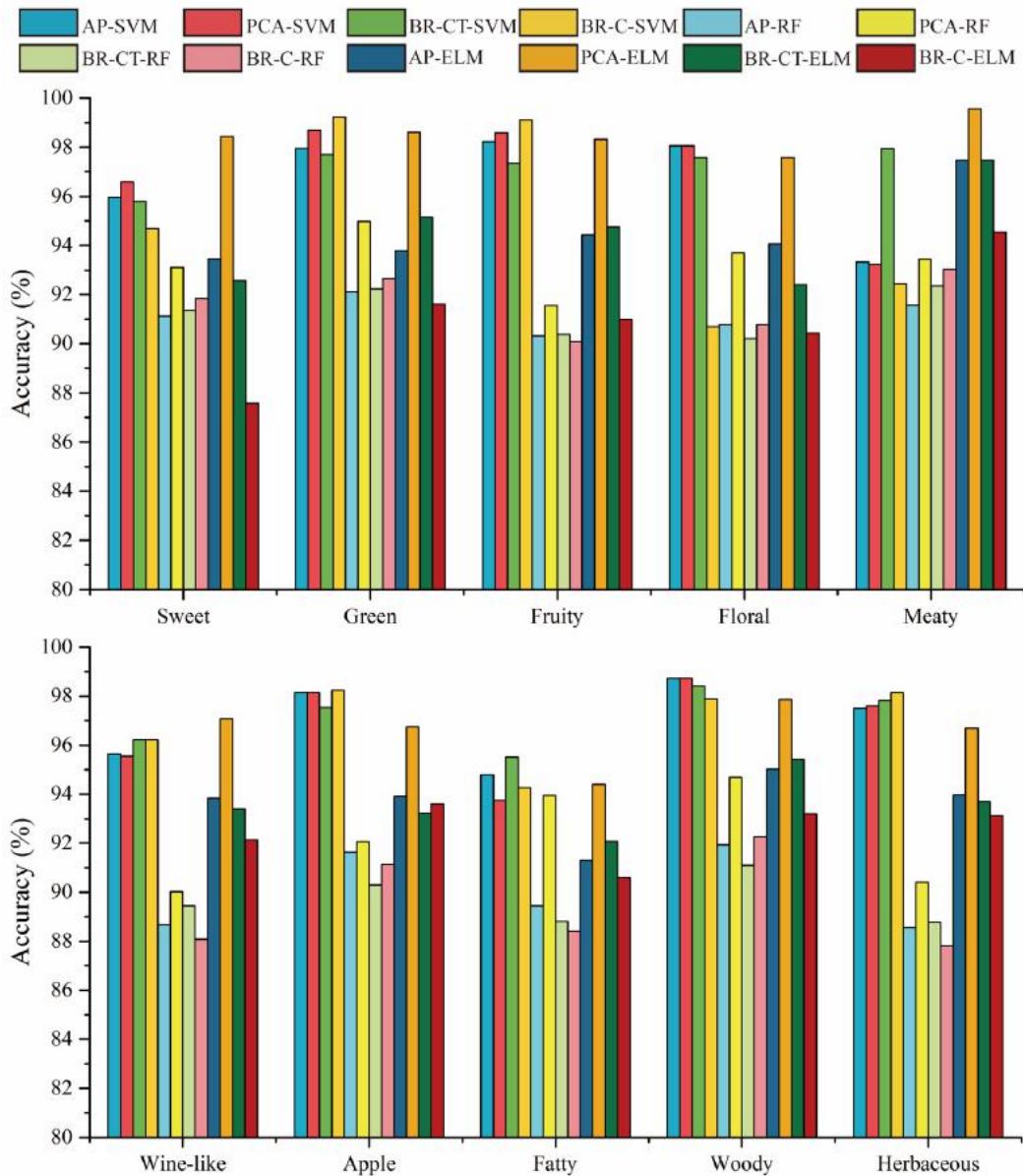


Fig. 3.6 Identification average accuracies of train and test sets for ten odor descriptors by SVM, RF and ELM.

### Extreme learning machine model

For ELM models, the parameter need to be tuned is the number of hidden layer nodes. In this study, the parameter was obtained by a trial and error method. The range of number of hidden nodes was set from 1 to 800. To avoid the randomness of ELM models, each ELM model was repeated 200 times and the average accuracy was employed to finish the parameter's selection (Figure B.14, B.15, B.16, B.17). Based on the highest average accuracies of calibration set and validation set, the optimal parameter was determined. Selected results are provided in Table 3.2. The ODs identification accuracies for the training and test sets of ELM are shown in Figure 3.6. It shows that the accuracy of PCA-ELM ( $97.53\pm1.35\%$ ) is higher than AP-ELM ( $94.13\pm1.44\%$ ), BR-CT-ELM ( $94.02\pm1.59\%$ ), and BR-C-ELM ( $91.78\pm1.91\%$ ).

#### 3.3.4 Model comparison

The average accuracies of the training and test sets for ODs are shown in Figure 3.6. For 'green', 'fruity', 'wine-like', 'apple', and 'herbaceous' identification, BR-C-SVM shows better results than other models. However, PCA-ELM did a better job identifying the 'sweet' and 'meaty' ODs. When the tree modeling methods were compared (Figure 3.7), it was found that the ELM had the best identification accuracy ( $97.53\pm1.35\%$ ), followed by SVM ( $97.19\pm0.93\%$ ), and RF ( $92.79\pm1.63\%$ ). Dealing with large variables slows down machine learning algorithms and requires more resources [149]. Here, PCA and BR were employed to extract kernel information from a large feature set. The results show that PCA did a better job than Boruta in the RF and ELM models. However, PCA is an unsupervised feature combination method; the PCs are computed based on original data set. Considering the amount of input information, BR is more suitable for feature extraction from MPs. It was confirmed that the training time increases with the number of features. Here, by the BR-C method, only 15.01% information was extracted instead of all MPs. Therefore, considering the accuracies and modeling time comprehensively, it is suggested that SVM combined with features extracted by BR-C, whose average accuracy was higher than  $96.10\pm2.8\%$ , is the optimal model in identifying perceptual descriptors based on MPs. Besides, the recall, precision, and F1 score of BR-C-SVM were  $94.83\pm5.61\%$ ,  $86.88\pm3.04\%$ , and  $95.74\pm3.52\%$ , respectively, which yields the model with an acceptable generalization ability to predict odor descriptors based on physicochemical parameters.

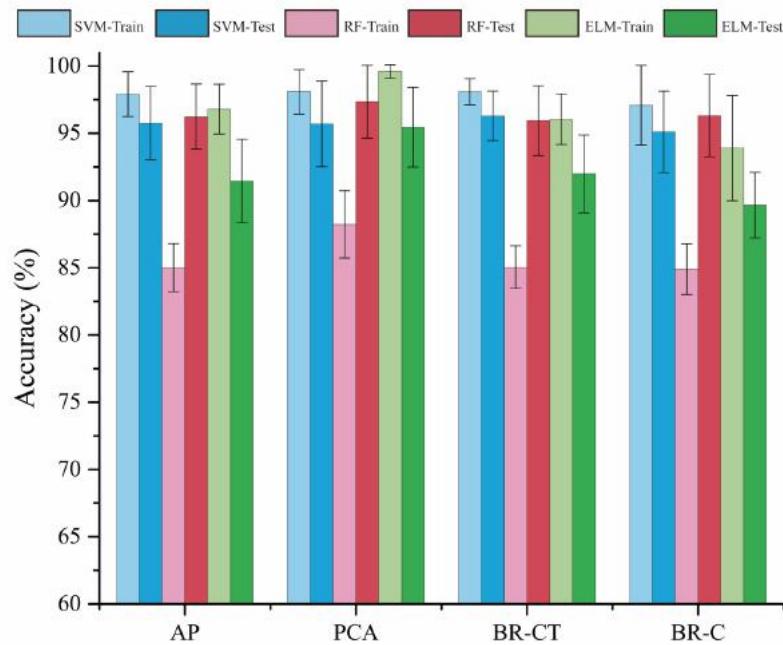


Fig. 3.7 Comparison of average identification accuracies by SVM, RF and ELM models under AP, features extracted by PCA, BR-CT ('confirmed' or 'tentative') or BR-C ('confirmed' only).

### 3.3.5 Model validation

Thirty primary VOCs identified from Golden Delicious apples analyzed by GC/MS and their ODs from databases and predicted by the models in this study were summarized in Table 3.3. It indicated that 70% (21/30) of compounds were predicted accurately. Besides, the other 7 compounds were shown unpredictable, which can be explained by the insufficient of OD models establish in presented research. Some ODs, such as 'peanut' and 'balsamic' etc., were not considered in present study because of their less samples. Although 70% would not enough to instead of panelists for GC-O, the ODs predicted by models can apply references for the panelists to enhance their work efficiency. Additionally, the predict accuracy would be increased by more odorant samples consideration and enough OD models establishment.

### 3.3.6 Discussion

In this chapter, the relation between MP and OD was discussed. Based on BR-C method, meaningful molecular information for 10 ODs were selected from 1006 MPs (Table B.4). Except constitutional MPs (such as molecular weight, number of atoms/bonds, number of sp hybridized Carbon atoms, etc.), information associated with molecular structure, include ring descriptors (cyclomatic number, number of circuits, total ring size, etc.), functional

Table 3.3 Models validation by golden delicious apple sample<sup>a</sup>.

No.	Volatile Organic Compound	Odor descriptor from database <sup>b</sup>	Predicted odor descriptor
1	2-propanol	Alcohol; butter	-
2	1-propanol	Alcohol; apple; musty; earthy; peanut; pear; sweet	Apple
3	1-butanal	Apple; chocolate; creamy; green; meaty; ethereal	Green, fruity
4	Ethyl acetate	Solvent-like; fruity; anise; ethereal; pineapple	-
5	2-methyl-1-propanol	Fruity; whiskey; wine-like; solvent-like	Fruity, wine-like
6	1-butanol	Banana; vanilla; fruity	-
7	Propyl acetate	Fruity, floral	Fruity
8	2-methyl-1-butanol	Onion; malty	-
9	1-pentanol	Sweet; vanilla; balsamic	-
10	Isobutyl acetate	Apple; banana; ethereal; pear; pineapple	Apple
11	1-hexanal	Fatty; green	Green, fatty
12	Butyl acetate	Banana; green; sweet	Green
13	(E)-2-hexen-1-al	Almond; apple; green; vegetable	Green, apple, fatty
14	1-hexanol	Green; herbaceous; woody	Green, fatty, woody, herbaceous
15	2-methyl-1-butyl acetate	Banana; peanut; fruity, apple-like	-
16	Butyl propanoate	Banana; ethereal	Apple
17	Amyl acetate	Fruity; banana; earthy; ethereal	Fruity, apple
18	(E)-2-hepten-1-al	Fruity; rose; fatty; almond-like	Green, fruity, apple, fatty
19	6-methyl-5-hexen-2-one	Fruity; citrus-like; strawberry	-
20	Butyl butanoate	Apple; banana; berry; peach; pear	Apple
21	Hexyl acetate	Apple; banana; cherry	Apple, fatty
22	2-ethyl-1-hexanol	Oily; rose; sweet	Woody, herbaceous
23	Butyl 2-methyl butanoate	Apple; chocolate	Apple
24	1-octanol	Fatty; citrus; waxy; woody	Fatty; woody
25	1-nonanal	Apple; coconut; fatty; fishy	Fatty
26	Hexyl butanoate	Green; fruity; apple; waxy	Fruity, wine-like, apple, fatty
27	P-allylanisole	Alcohol; green; minty; sweet; vanilla	Sweet; green; floral
28	Hexyl 2-methyl butanoate	Green; fruity; apple; grapefruit-like	Green; fruity; apple; herbaceous
29	Hexyl hexanoate	Green; vegetable; fruity; apple; cucumber-like	Green; fruity; fatty
30	(E,E)- $\alpha$ -farnesene	Green; herbaceous	-

<sup>a</sup> Boldface type indicates correctly predicted ODs.<sup>b</sup> The odor databases included Flavornet, Sigma-Aldrich, GoodScents, and SuperScent.

group counts (number of aromatic Carbon, umber of unsubstituted benzene Carbon, number of esters/primary amides/ketones, etc.), atom-centred fragments ( $\text{CH}_3\text{R}/\text{CH}_4$ ,  $\text{CH}_2\text{R}_2$ ,  $=\text{CH}_2$ , R-CH-R, H attached to  $\text{C}_1(\text{sp}3)/\text{C}_0(\text{sp}2)$ , etc.), 2D atom pairs (sum of topological distances between N..N/N..O/ N..S/ O..O, presence/absence of N-S, etc.), played critical roles on ODs. Additionally, 2D auto correlations (topological charge, mean topological charge, Geary autocorrelation, etc.), burden eigenvalues (largest eigenvalue of Burden matrix weighted by mass/van der Waals volume/ionization potential, etc.) also contributed on ODs prediction. The result indicated that if the molecular structure can be detected by sensor, ODs would be predicted. Therefore, molecularly recognized material, such as molecularly imprinted materials should be considered for developing olfaction inspired odor sensors.

### 3.4 Conclusion

The contribution of this chapter is to present an approach to predict odor perceptions based on physicochemical descriptors. After processing by the SMOTE and KS methods for balancing dataset and subset partitioning, two feature extraction methods (PCA and BR) were used to extract kernel information from 1006 MPs. Three machine learning approaches (SVM, RF, and ELM) were employed to establish odor descriptor classifier models. The results showed that SVM models presented better accuracies than others. Although the accuracy of the BR-C-SVM model is lower than AP-SVM, PCA-SVM, and BR-CT-SVM, when considering the complexities of the models, BR-C-SVM would be the optimal model in this study. Therefore, BR-C-SVM has a good potential in predicting odor perceptions rapidly and precisely. This study demonstrated that MPs associated with machine learning models can be adopted for odor perceptual senses identification. The research is expected to offer a novel approach for developing olfaction inspired odor sensor system.

# **Chapter 4**

## **LSPR sensor based on MISGs for volatile organic acid detection**

### **4.1 Introduction**

Human body odors emitted from skin and body parts are caused by gender, age, heredity, physiological condition and food habits etc. [183, 184]. Based on those odors alone, people can assess various personal features of others accurately [185]. Hence, human body odor would be applied in medical diagnosis and forensic expertise [186]. It has been demonstrated that human body odor is comprised by diverse VOCs, such as low molecular weight fatty acids, aldehydes, ketones, amines, alcohols, esters, etc. [187–189]. Among these odorants, organic acids ( $C_5$ - $C_{11}$ ) are considered as typical odorants from difference human body parts, such as foot odor and underarm odor [190]. Traditionally, human body odor has been analyzed by gas chromatography/mass spectrometer (GC/MS) method [34, 35]. However, GC/MS is not suitable for on-line detection because of its high-cost, time-consuming and bulky size etc. [37]. Therefore, novel sensors need to be explored for detecting organic acid from human body odor.

In chapter 2 and 3, we had founded that the detection of molecularly structure would be applied for developing odor sensors. In this chapter, molecular imprinted sol-gels (MISGs) were expected as a molecularly structure recognized layer for localized surface plasmon resonance (LSPR) sensors. The schematic diagram of MISG coated LSPR sensor is shown in Figure 4.1a. Based on the unusual cavities generated in sol-gel matrix, the target organic acid vapor would be absorbed selectively. And it would induce the change of surface plasmon peak position ( $\lambda_{\min}$ ) and the transmittance variation in spectrum. By detecting these variations, an optical sensor for organic acid vapors detection would be developed. In ad-

dition, human body odor is always composed by diverse of organic acid vapors. Hence, to detect the mixture of organic acid vapors is another topic in body human odor determination. Here, three MISG films generated by different organic acid templates, hexanoic acid (HA), heptanoic acid (HPA) and octanoic acid (OA), were coated on Au nano-island layers for establishing a MISG-LSPR multichannel sensor platform (Figure 4.1b). The response matrix was obtained by measuring for organic acid vapors: propanoic acid (PA), HA, HPA and OA, in single and their binary mixtures. Finally, the response matrix was processed and analyzed by principal component analysis (PCA) and linear discriminant analysis (LDA) for odor pattern recognition. The feasibility of the developed MISG-LSPR sensor array for determination of organic acid vapors was discussed and evaluated.

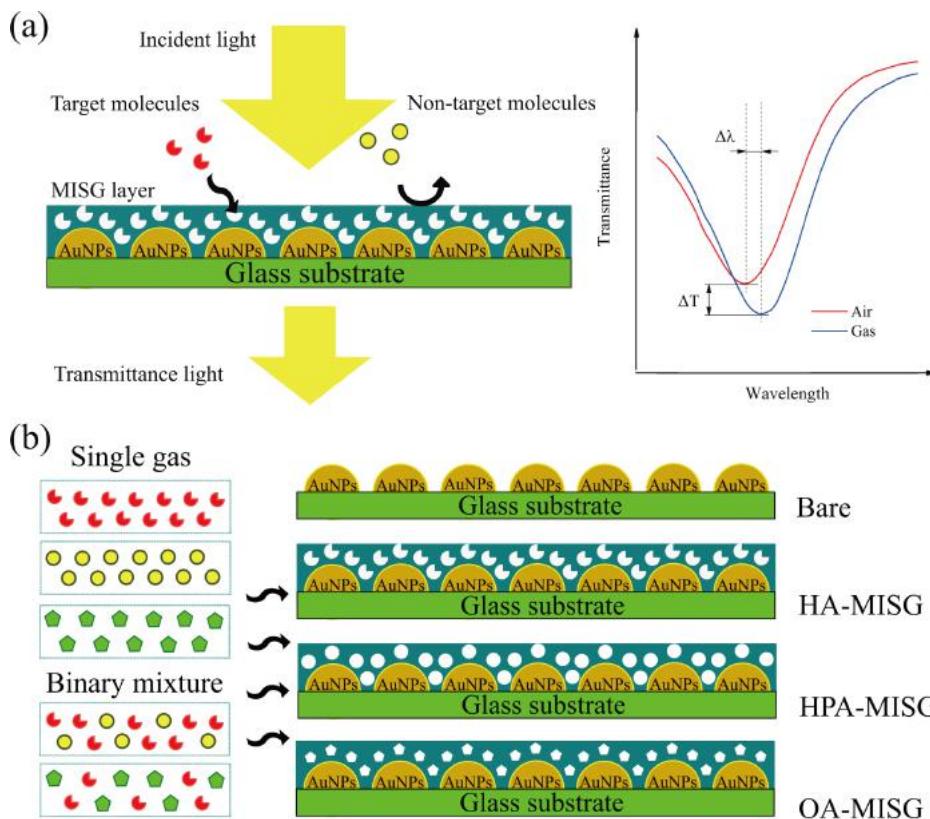


Fig. 4.1 The schematic graph of MISG-coated AuNPs film for selective organic acid detection (a) and MISG-LSPR multichannel sensor platform for organic acid vapors mixture detection (b).

## 4.2 Materials and method

### 4.2.1 Chemicals and reagents

Tetrabutoxy titanium (TBOT), iso-propanol, PA, HA, HPA, OA, titanium tetrachloride ( $\text{TiCl}_4$ ), acetone and ethanol were purchased from Wako Pure Chemical Industries, Japan. 3-aminopropyl triethoxysilane (APTES) was purchased from Shin-Etsu Chemical, Japan. All of the reagents were used as received.

### 4.2.2 Synthesis of MISG reaction solutions

MISGs reaction solution was prepared by dissolving 136  $\mu\text{L}$  TBOT as a precursor, 50  $\mu\text{L}$  of template molecules and 24  $\mu\text{L}$  APTES as a functional monomer in 2 mL of iso-propanol. Here, HA, HPA and OA were selected as the template molecules. Afterwards, 25  $\mu\text{L}$   $\text{TiCl}_4$  was added to initialize the reaction. Finally, the reaction solution was prehydrolyzed in a 70 °C water bath for 1 h.

### 4.2.3 MISG coated Au nano-island film preparation

Concisely, a TIO glass substrate was cleaned by ultrapure water, acetone and ethanol and dried with nitrogen flow, successively. After argon plasma cleaned for 5 min (PDC-001, Harrick plasma, USA), the substrate was immersed in a 1:10 (v:v) ethanol solution of APTES for 8 h. The substrate was cleaned with ethanol and drying with nitrogen flow, and put into a quick coater (SC-701 HMCII, Sanyu electron, Japan) for AuNPs deposition, the thickness was set as 3 nm by tuning the deposition current. Then, the sample was annealed in air atmosphere at 200 °C for 5 h in a muffle furnace (SSTS-13 K, ISUZU, Seisakusho, Japan) and cooled naturally till room temperature (25 °C). Afterwards, MISG layers were coated on the AuNPs film by spin coating 20  $\mu\text{L}$  of its reaction solution. As the last step, the sample was heated at 200 °C for 1 h for constructing the MISG layer and removing the template molecules.

### 4.2.4 Vapor generating system

The vapor generating system used in this study is shown in Figure 4.2a. It was consisted of an air pump (LV-125A, Linicon, Japan), an air-cleaning filter filled with molecular sieves and activated carbon, 2 mass flow controllers (MFC) (3660, Kofloc, Japan), a 3 way solenoid valve (FSM-0408Y, FLON Industry, Japan), a glass bottle (6mL) and a personal computer. Pure dry air was as the diluting gas in this study. All the gas flow paths were connected by

Teflon tubes. Through a NI DAQ (USB-6009, National instruments, Austin, USA) card and LabView software (National instruments, Austin, USA), two MFCs and the gas valve can be controlled by the personal computer. The concentration  $C$  (ppm) of single organic acid vapor can be calculated by Formula 4.1.

$$C = \frac{k \times D_r \times 10^3}{F} \quad (4.1)$$

where  $D_r$  ( $\mu\text{g}/\text{min}$ ) indicated the diffusion rate at the appoint temperature,  $F$  ( $\text{L}/\text{min}$ ) indicated the flow rate of diluent gas,  $k$  indicated the factor for converting gas weight to gas volume, which can be calculated as Formula 4.2.

$$k = \frac{22.4 \times (273 + t) \times 760}{M \times 273 \times P} \quad (4.2)$$

where  $M$  indicated the molecular weight of organic acid molecule,  $t$  is the gas temperature and  $P$  is the gas pressure (760 mmHg). In this work, single organic acid vapor was generated by injecting 2 mL of each of organic acids (PA, HA, HPA or OA) in the glass bottle. Binary mixture of acid vapors (A+B) was generated by injecting 1 mL of organic acid A and 1 mL of organic acid B in the glass bottle together. In this study, 3 types of binary mixtures (PA+HA, PA+OA, HA+OA) were considered. Cleaned glass bottles were used for each testing. The flow rates of diluent air were set as 0.6, 0.5 and 0.4  $\text{L}/\text{min}$ , respectively.

#### 4.2.5 Sensing system

SEM (SU8000, Hitachi, Japan) was employed to analyze the morphology characteristics of Au nano-islands before and after sol-gel deposition in this study. The schematic of transmittance spectra measurement system is show in Figure 4.2b. The system was included a light source (LS-1 tungsten halogen light source, Ocean optics, USA), a UV spectrometer (HR4000, Ocean optics, USA), a homemade sensing cell (Teflon), 2 optical fibers (Ocean optics, USA) and a personal computer. By the software named OPwave+ (Ocean optics, USA), the transmittance spectra in real-time were detected and recorded. The scanning range was set from 400 to 900 nm, and the wavelength resolution was 0.1 nm. The actual photo for the experimental system and MISG-LSPR electrode in sensing cell was provided in Figure 4.2c.

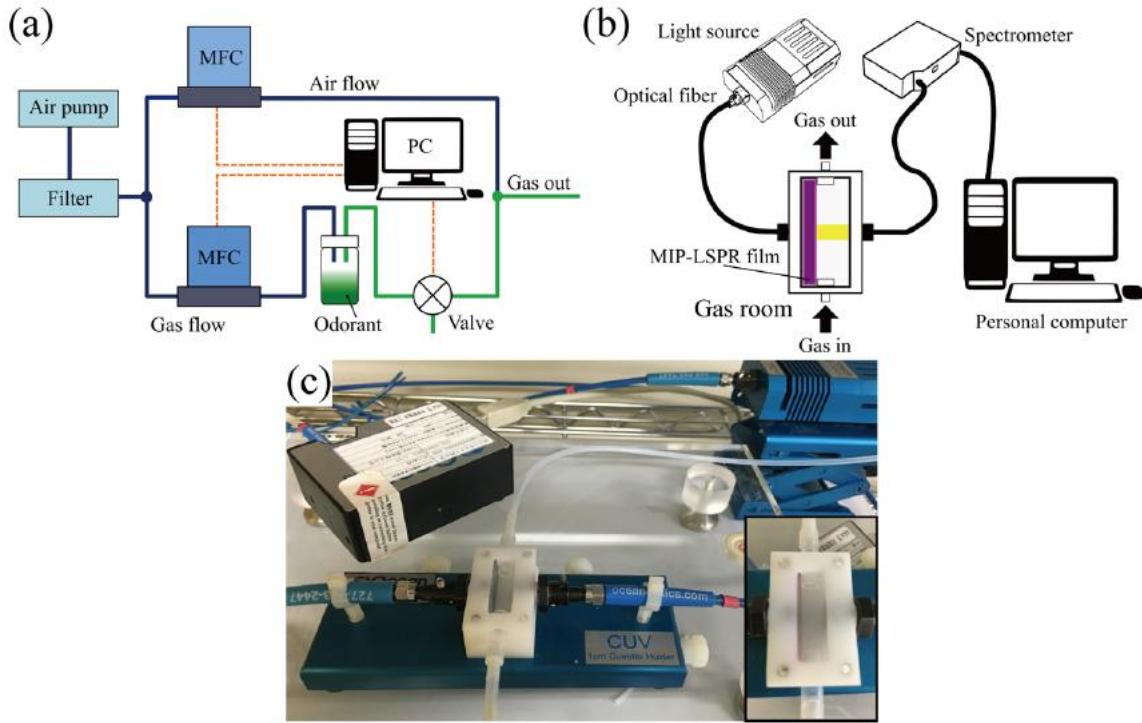


Fig. 4.2 Schematics of vapor generation system (a) and sensing system (b). The photo of the experimental system (c) and the MISG-LSPR electrode in the sensing cell.

## 4.3 Results and discussion

### 4.3.1 UV-vis spectra and vapor absorption characteristics

The thickness of MISG layer is a critical factor to its selective absorbability and it can be controlled by spin coating speed [115]. Firstly, the influence of the spin coating speed on optical characteristics of MISG/NISG coated Au nano-island films were investigated. The transmission spectra of bare, NISG and HA-MISG coated Au nano-island versus different spin coating speeds (1000 rpm, 3000 rpm and 5000 rpm) were shown in Figure 4.3a. It was demonstrated that the sol-gel layer makes the plasmon peak shift to the red and the transmittance decrease (Figure 4.3b). Besides, with the increase in spin coating speed, the transmittance was increased. In addition, the minimum transmittances of samples coated with NISG were lower than those coated with MISG at the same spin coating speed. These transmittance decrease and spectral position red shift showed a spin coating speed dependent feature.

To determine the optimal coating speed for MISG layers, the real-time response characters of Bare/NISG/HAMISG coated sensors to HA vapor were investigated. The changes of

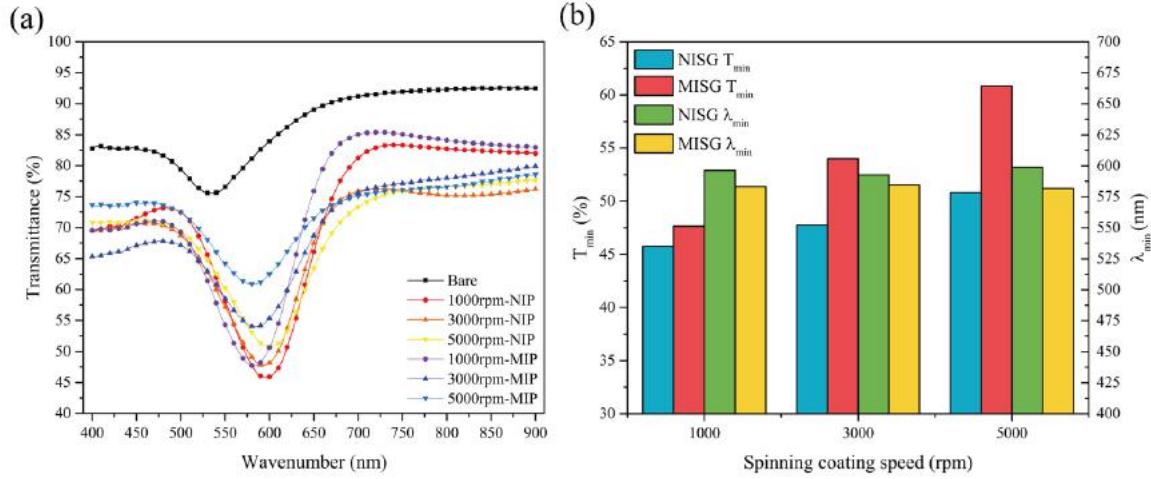


Fig. 4.3 Transmittance spectra (a) and optical features ( $\lambda_{min}$  and  $T_{min}$ ) (b) of NISG and HA-MISG coated AuNPs versus different spin coating speeds.

transmittance at plasmon peak ( $\lambda_{min}$ ) were detected and recorded. The  $\Delta T$  can be calculated by Formula 4.3.

$$\Delta T = T - T_0 \quad (4.3)$$

Where  $T_0$  indicated the transmittance in air, and T indicated the transmittance in organic acid vapors. The real-time response to HA vapor was as shown in Figure 4.4. To investigate the changes of LSPR response before and after coated MISG/NISG layers, bare Au nano-island was also considered in this study. The RIS of surface plasmon (SP) extinction bands to dielectric properties of the surrounding medium was depended on the particle size and the distance between particles [132]. The surface morphology of bare Au nano-island was studied as shown in Figure 4.5a. It indicated that all AuNPs were formed as arrays, which would induce a stranger RIS for LSPR [191].

Figure 4.4 showed that the response of bare sample was larger, and its response time was faster than NISG or MISG coated samples. Compared with MISG/NISG coated samples, more HA molecules could be absorbed in the RI sensing volume (SV) of AuNPs, and it would induce a stranger response for the vapor. We can also find that no responses were observed on samples coated NISG. It indicated that the adsorption capacity of pure titanate sol-gel matrix was weak, which agree with the results reported by Matsuguchi et al. [122]. The SEM image for NISG coated sample (spin coating speed 3000 rpm) is shown in Figure 4.5b. It suggested that the surface of NISG was full of cracks. Compared with MISG coated samples (Figure 4.5c), NISG coated sample showed a relative smooth surface and its surface area was smaller. This surface morphology would induce its poor gas responses. Besides, gas

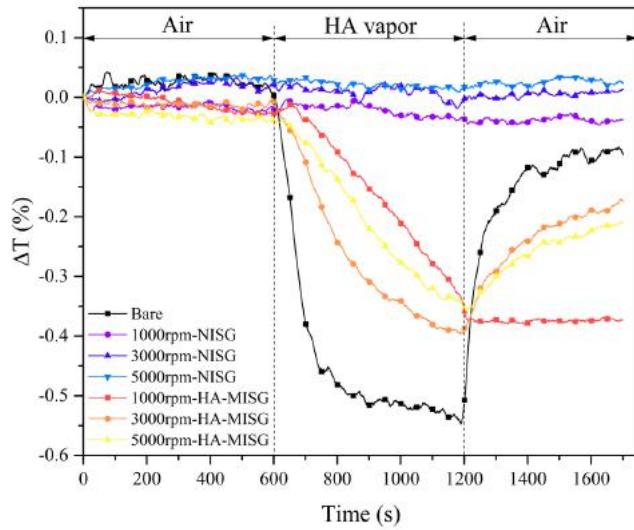


Fig. 4.4 Real-time response of HA-MISG and NISG with different coating speeds to HA vapor. Gas responses were obtained by keeping the switch on to HA vapor flow for 600 s and then to air flow for 600 s.

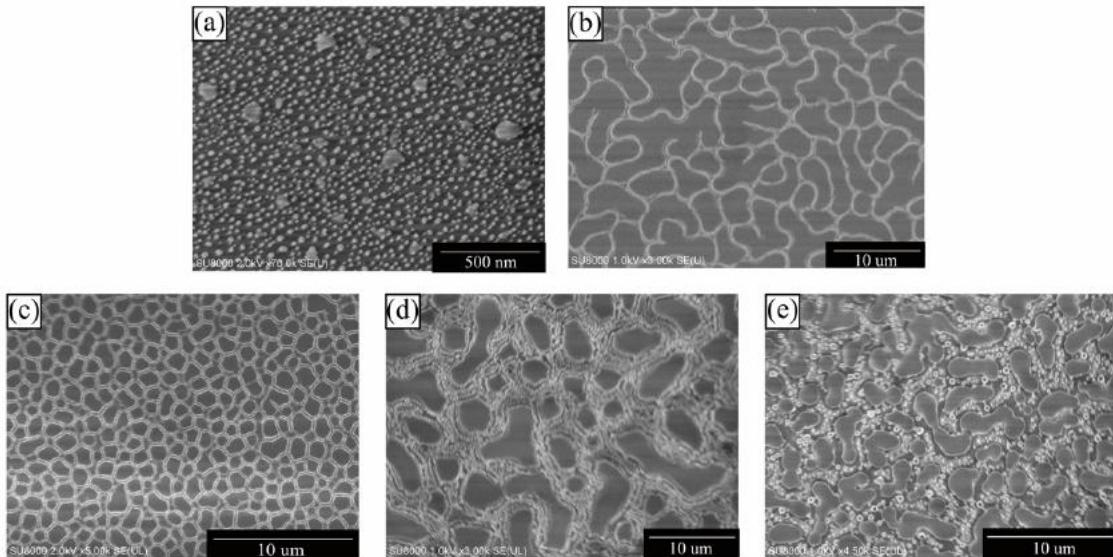


Fig. 4.5 SEM images of bare Au nano-island (a) and coated with NISG (b), HA-MISG (c), HPA-MISG (d) and OA-MISG (e). All the MISG/NISG films were fabricated by spin coating speed at 3000 rpm.

molecules would be obstructed by the pure sol-gel layer to be in the SV of AuNPs. Compared with the bare sample, the responses of MISG coated samples were smaller. It could be

explained that only the molecules absorbed by the nano-scale cavities of MISG in SV could be sensed. By the effect of MISG layer, a longer response and recovery time was observed in in-situ responses. Besides, too thick layer (spin coating speed: 1000 rpm) would induce a longer recovery time. Just like we discussed in former work, excessive thin MIP layer is difficult to realize a good selectivity for target molecules, while thick layer would induce a long recovery time [115]. Considered the response and recovery time simultaneously, the relative optimal spin coating speed was selected as 3000 rpm in this study.

To determine the absorption characteristics of HA-MISG films to HA vapor, the in-situ responses of HA-MISG coated LSPR sensor (spin coating speed: 3000 rpm) to PA, HA and OA vapors were investigated. The normalized response  $K$  can be defined by Formula 4.4.

$$K_j = \Delta T/C_j \quad (4.4)$$

Where  $j$  is on behalf of 3 types of organic acid vapors: PA, HA and OA.  $C_j$  is the concentration of organic acids. Here, the concentrations of PA, HA and OA were 40.93, 21.05 and 11.23 ppm, respectively. Figure 4.6 shows that the corresponding response signal of HA (0.01844) is the larger than that for PA (0.00523) or OA (0.00781). A faster response speed was also observed to HA vapor. Besides, the recovery time for HA is longer than other vapors. It indicated that more target gas molecules were absorbed in the SV of AuNPs, which would be contributed by the selectivity of the MISG layer.

### 4.3.2 Sensor array response for organic acid odors

By spin coating 3 types of MISG reaction solutions at 3000 rpm on Au nano-island layers, a MISG-LSPR sensor array was constructed. The sensor array was consisted of 4 channels: bare, HA-MISG, HPA-MISG and OA-MISG (Figure 4.1b). The transmission spectra for 3 types of MISG coated samples were shown in Figure 4.7. It suggested that by spin coating different type of MISGs, their  $\lambda_{\min}$  and  $T_{\min}$  were different. SEM images for different types of MISG coated samples are shown in Figure 4.5. It illustrated that AuNPs were covered by MISG films. We can also find that the surfaces of MISG coated samples were full of cracks, and the degrees of tearing and roughnesses were different. Therefore, the surface areas of MISG coated sample are larger than NISG coated samples', which would induce the different response intensities. By adding template molecules, the polymerization of sol-gel can be effected, which would be explained these diverse morphologies. Because the size of cavities generated by template molecules in MISG films were too small, it is different to observe from SEM images.

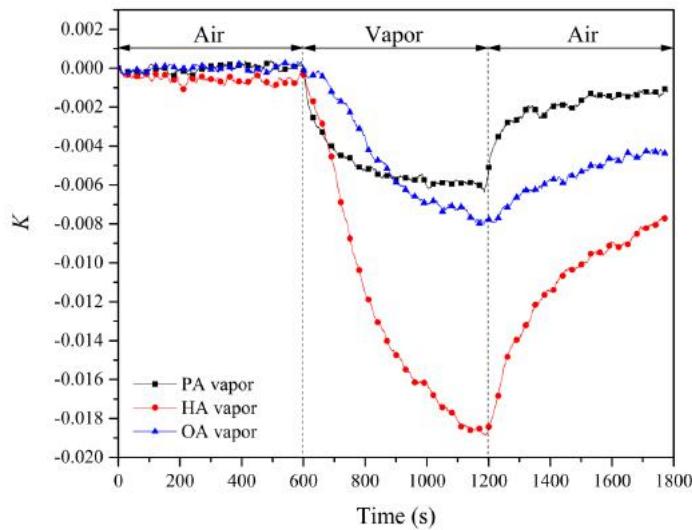


Fig. 4.6 Real-time responses of HAMISG-LSPR sensor to three fatty acid vapors (PA/HA/OA). Gas responses were obtained by keeping the switch on to organic acid vapor flow for 600 s and then to air flow for 600 s. The concentrations of PA, HA and OA were 40.93, 21.05 and 11.23 ppm. The  $K_{min}$  for PA, HA and OA were -0.00523, -0.01844 and -0.00781, respectively.

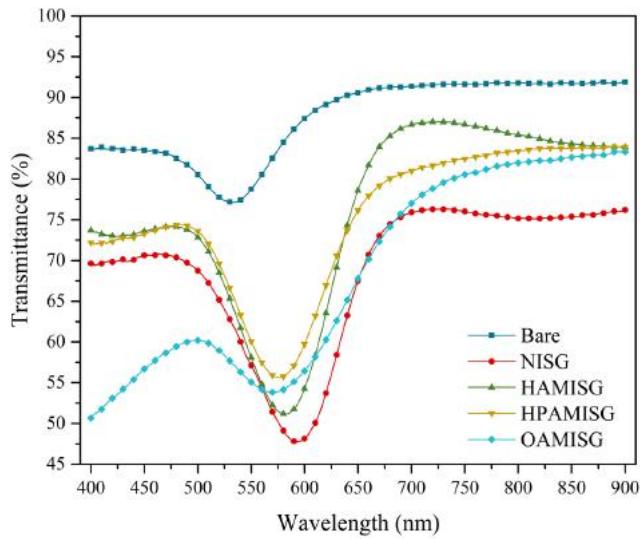


Fig. 4.7 Transmittance spectra of bare, NISG and three types of MISGs (HA-MISG/HPA-MISG/OA-MISG) coated samples.

For each measurement, sensors were exposed in dry air for 300 s firstly. Then, the target vapor exposure time 60 s and dry air was passed for next 300 s for recovery. By changing the flow rates (0.4 L/min, 0.5 L/min and 0.6 L/min), 3 concentrations of a vapor would be obtained. A typical response of HA-MISG coated LSPR sensor to 4 types of organic acid vapor (flow rate: 0.5 L/min) was shown in Figure 4.8. In this study, 9 samples (3 concentrations  $\times$  3 repetitions) from 7 types of vapors (PA, HA, HPA, OA, PA+HA, PA+OA, HA+OA), total 63 samples were considered. Hence, a response ( $\Delta T$ ) matrix  $M_{63 \times 4}$  for the sensor platform could be obtained for subsequent research.

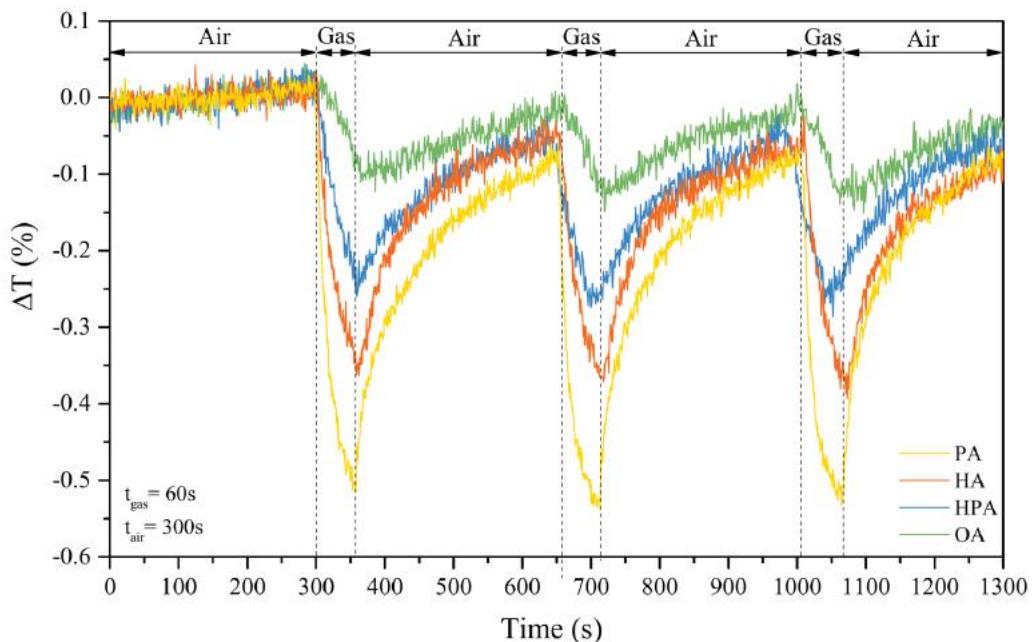


Fig. 4.8 Typical response of HA-MISG coated LSPR sensor to fatty acid vapors. Gas responses were obtained by keeping the switch on to organic acid vapor flow for 60 s and then to air flow for 300 s.

### 4.3.3 Discrimination of organic acid vapors

Before discriminating, the matrix was pre-processed by auto-scaling to reduce the large variations in response data for different channels. To visualize the cluster trends of vapor samples, PCA was performed on the normalized response matrix. PCA is a conventional unsupervised linear method for information concentrating, noise removing and data visualization

[165]. By this method, principal components (PC) could be constructed by linear combination of the original variables [159]. Based on these uncorrelated PCs, samples could be mapped in a low dimensional space for discriminating [160, 192]. The PC score plot of total 63 samples is shown in Figure 4.9. Because PCA is an unsupervised method, the samples were clustered together only based on the similarities and differences in their PC scores. In PC1-PC2 space (Figure 4.9a), each of 4 single odors occupied a separate region, and all binary mixtures were lying in new clusters. PC1 contained mainly information (88.78%) of original response matrix. Besides, for PC1, 4 single vapors were sorted by concentration in descending order. It indicated that the concentration information for vapors would be contained in PC1. In the PC1-PC3 space (Figure 4.9b), most of samples were patterned in individual clusters. But an overlapped was observed in the samples from PA and PA+OA vapors. In addition, a well clustering result was observed in PC2-PC3 space (Figure 4.9c) excepting a superposition between HA ( $C_6$ ) and HPA ( $C_7$ ) vapors. We can also find that for single vapor, the sort by PC2 is similar with that by molecular size. It might be contributed by the size effect of the imprinted template molecules. Besides, we can also find that the VOCs mixture samples were clustered on the centerline between two source VOCs in PC1-PC2 and PC1-PC3 spaces. However, in PC2-PC3 space, we did not find the similar result. It indicated that we could find a balance role in PCA spaces partly. The reason would be explained by the different concentrations for VOCs mixtures or pure organic acids.

To investigate the pattern recognition ability of MISG-LSPR multichannel sensor platform, LDA was applied in this study. Different from PCA, LDA is a supervised classifier by finding a discriminant function (DF), which is a linear combination of the original variables (features of the sensor responses) that tries to maximize the variance between groups and minimize the variance within groups [193, 194]. More detail information about LDA can be found else here [195–198]. Here, the optimal transformation in LDA was achieved by minimizing the intragroup distances and maximizing the intergroup distances simultaneously, thus the best group discrimination could be obtained [199]. Similar to PCA, 2 discrimination functions were obtained by a linear combination of the 4 variables in the sensor array. Consequently, all samples could be plotted in a LDA space as shown in Figure 4.10. It demonstrates the clear clustering of 7 distinct groups, which correspond to 4 single and 3 binary mixtures of organic acid vapors, with no overlap being observed. Taking into account a small number of samples in this study, full leave-one-out cross-validation (LOOCV) method was applied to validate LDA models in this study [200, 201]. The LOOCV of LDA scores revealed a classification accuracy of 100%. It suggested that the multichannel LSPR-MISG sensor platform developed in this study could be applied on the pattern recognition of single or binary mixture of organic acid.

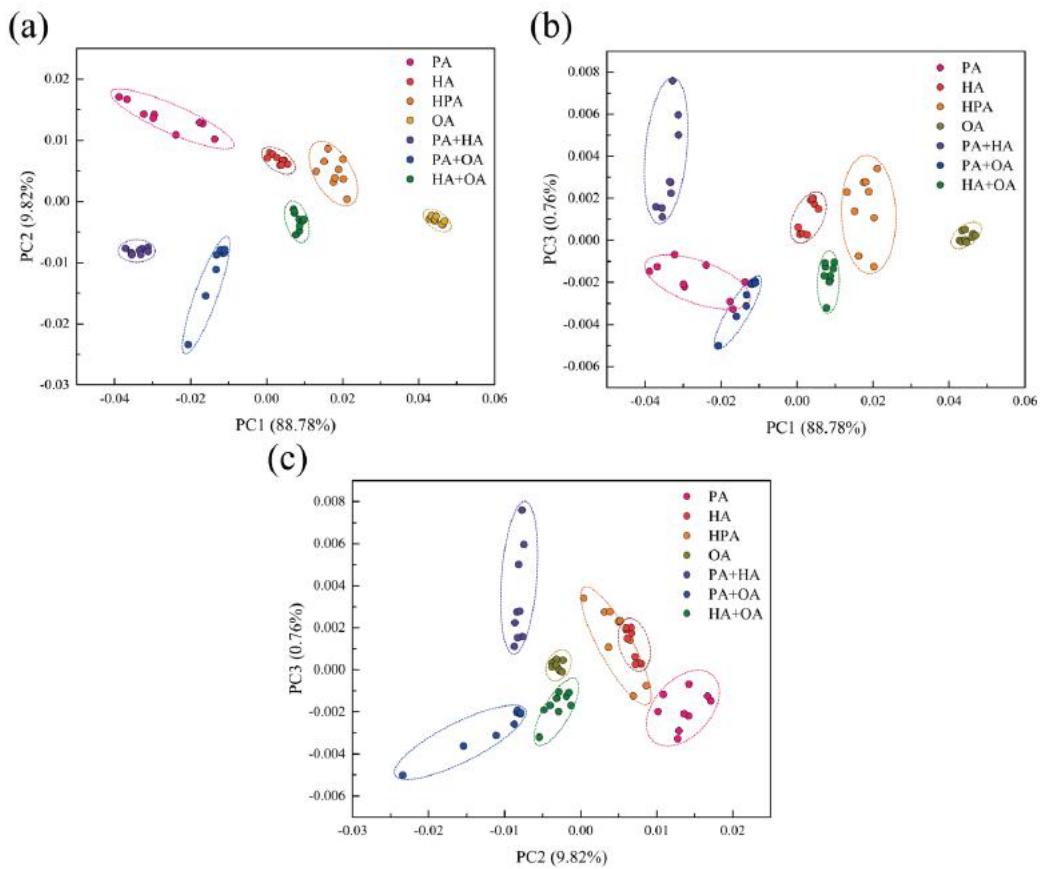


Fig. 4.9 PCA score plots of the multichannel responses for 63 samples from PA, HA, HPA, OA, binary mixture of PA+HA, PA+OA and HA+OA.

#### 4.3.4 Discussion

In this chapter, MISGs were designed for fatty acid vapor detection. Molecules with similar structure to imprinted molecules (with carboxyl group, different carbon chain length) were selected for MISGs selectivity evaluation. Results indicated that a selectivity shown to the target molecule. In addition, PCA results indicated that concentration information was shown in PC1 and molecular size information was shown in PC2. It indicated that molecular information (size, functional group and carbon chain length) can be captured by MISGs layer. Therefore, MISGs contained with different type of imprinted molecules would be developed for odor sensing system.

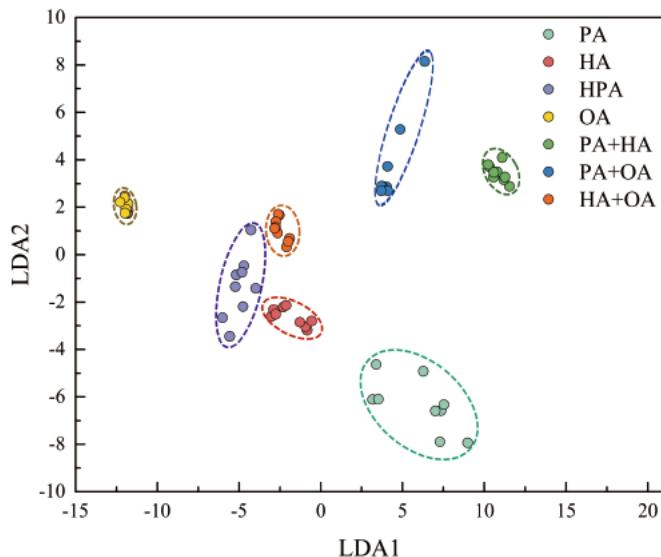


Fig. 4.10 LDA score plots of the first 2 discriminant factors (LDA1 and LDA2) achieved from transmittance change responses data of 63 samples from PA, HA, HPA, OA, binary mixture of PA+HA, PA+OA and HA+OA. Oval outlines indicate group of organic acid samples at 99% confidence level.

## 4.4 Conclusion

In summary, a MISG coated Au nano-island film was developed for determination of organic acids vapors selectively. The MISG reaction solution was spin coated on the Au nano-island layer. The results demonstrated that the adsorption capacity of pure TiO<sub>2</sub> sol-gel matrix was weak. In-situ response of HA-MISG was verified to be fast, selective and reversible. Eventually, by changing the template molecules in MISG reaction solutions, a 4 channels MISG-LSPR multichannel sensor array was constructed for the determination of 4 organic acids vapors (PA, HA, HPA and OA) in single and their binary mixtures. PCA and LDA were employed for pattern recognition of the response matrix. We also find a partly balance role between mixture VOCs and their source VOCs in PCA spaces. A 100% classification rate was achieved by leave-one-out cross-validation technique for the LDA model. It indicated that a sensor array combined MISG with LSPR could be an effective method for organic acid odor pattern recognition. This research offers some useful technologies for developing sensor system for organic acid from human body odor.



# **Chapter 5**

## **Development of MISG based LSPR sensor for detection of volatile cis-jasmone**

### **5.1 Introduction**

Plants have evolved a variety of sophisticated mechanisms to withstand stresses imposed by salt, cold, heat, herbivore attack, or pathogen infection [202]. A critical mechanism for self-protection, as well as a communication signal between plants, is the production of complex mixtures of plant volatile organic compounds (PVOCs) [203]. PVOCs can not only attract pollinators or seed-dispersing animals, but also repel potential herbivores [204]. Recently, much effort has been focused on clarifying the metabolic pathways and functions of these PVOCs [205, 206]. Sobhy et al. suggested that (E)-2-hexenal, methyl salicylate, cis-jasmone (CJ), and methyl benzoate were the main compounds in the PVOCs [207]. In particular, jasmone released by flowers or leaves can be used as a chemical cue for herbivorous insect infestation [208]. In addition, Birkett et al. reported that the biosynthesis of CJ was associated with stress-induced jasmonic acid or octadecanoid pathways [186, 209, 210]. Bruce et al. reported that CJ treatment of crop plants, such as soybeans and potatoes, directly defended against aphids, and also initiated PVOC release that repelled natural enemies [211]. Therefore, CJ can be regarded as a vital biomarker for plant pest and disease monitoring in agriculture [212]. Currently, PVOCs has been analyzed with gas chromatography/mass [36]. However, this is not suitable for real-time PVOC monitoring because it is time consuming, costly, and not portable [213, 214]. Therefore, alternative sensing strategies need to be con-

sidered for real-time monitoring of CJ vapors with high sensitivity, selectivity, and response speeds for agricultural applications [215].

The aim of this chapter was to develop a molecularly imprinted sol-gel (MISG) based localized surface plasmon resonance (LSPR) for detection of CJ. As illustrated in Figure 5.1, Au nano-islands were fabricated by vacuum sputtering and annealing. Then, a three dimensional "imprint" of CJ molecules within the MISG was created around the Au nano-islands. An optical sensor was developed for CJ detection by monitoring changes in the RI by absorbance spectra. The functional monomer was a vital element of the MISGs, and was investigated to enhance imprinting effects for sensitivity. Interactions between the functional group of functional monomers and imprinted molecules are important for cavities generation in MISG layer. In addition, the effect of the ratio of matrix material to functional monomer on the CJ response was examined. The feasibility of the MISG-LSPR sensor for CJ vapor was evaluated.

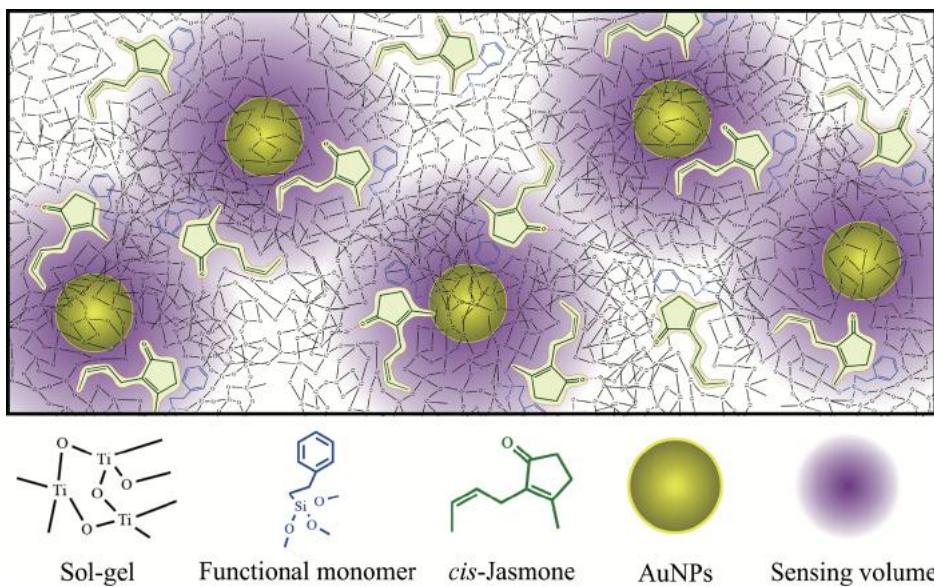


Fig. 5.1 Schematic of MISG-coated Au nano-islands with functional monomers for selective CJ vapor detection.

## 5.2 Materials and Methods

### 5.2.1 Materials and instrumentations

Tetrabutoxy titanium (TBOT), iso-propanol, CJ, limonene,  $\gamma$ -terpene, titanium tetrachloride ( $TiCl_4$ ), acetone, and ethanol were purchased from Wako Pure Chemical Industries, Japan.

3-aminopropyl tri-ethoxysilane (APTES) was purchased from Shin-Etsu Chemical, Japan.  $\alpha$ -pinene, trimethoxyphenylsilane (TMP), triethoxy phenylsilane (TEP), trimethoxy(2-phenylethyl)silane (TM2P) were purchased from Sigma-Aldrich Co. LLC, USA. Benzyltriethoxysilane (BTE) was purchased from Tokyo Chemical Industry Co., Japan. All reagents were used as received. Fourier transform-infrared spectroscopy (FT-IR, Nicolet iS 5, Thermo Fisher Scientific, USA) was used to analyze MISG-AuNP films before and after CJ absorption. Scanning electron microscopy (SEM, SU8000, Hitachi) and atomic force microscopy (AFM, Dimension Ion, Bruker IXS) were used to image morphologies of AuNP-coated NISGs/MISGs.

### 5.2.2 Preparation of MISG reaction solution

Preparation of the MISG reaction solution was reported previously [144]. Briefly, it was prepared by dissolving 0.441 mmol (150  $\mu$ L) TBOT precursor, 0.304 mmol (50  $\mu$ L) CJ templates, and 50  $\mu$ L of functional monomers (TMP, TEP, TM2P, or BTE) in 2 mL of isopropanol while stirring. The concentrations of TMP, TEP, TM2P, or BTE in the MISG solutions were 0.252 mmol, 0.208 mmol, 0.221 mmol, and 0.197 mmol, respectively (see Figure 5.2 for structures). Then, 0.132 mmol (25  $\mu$ L) of  $TiCl_4$  was added to initiate the reaction. Afterward, the MISG reaction solution was pre-hydrolyzed in a 60 °C water bath for 1 h while stirring. Finally, the solution was stirred at room temperature (25 °C) for 8 h to complete the MISG reaction.

### 5.2.3 MISG-coated LSPR sensor preparation

A 12-mm 9-mm glass substrate was ultrasonically cleaned in ultrapure water, acetone, ethanol, and then exposed to an argon plasma (PDC-001, Harrick plasma, USA). It was then immersed in a 1:10 (v:v) APTES ethanol solution for 8 h. After being cleaned with ethanol and dried with flowing nitrogen, the sample was put into a quick coater (SC-701 HMCII, Sanyu electron, Japan) for a 3-nm deposition of AuNP. Thermal annealing was then performed in a muffle furnace (SSTS-13K, ISUZU, Seisakusho, Japan) at 500 °C for 2 h, followed by cooling to room temperature. The sample was then subjected to sputtering and annealing again under the same conditions to form a high-sensitivity LSPR substrate. Then, 20  $\mu$ L of the MISG reaction solution was spin-coated on the Au nano-islands at 3000 rpm for 1 min. In the last step, the sample was kept at 130 °C for 1 h to finish the MISG fabrication and to evaporate the templates. All samples were stored under vacuum to remove volatile organic compounds in the MISGs.

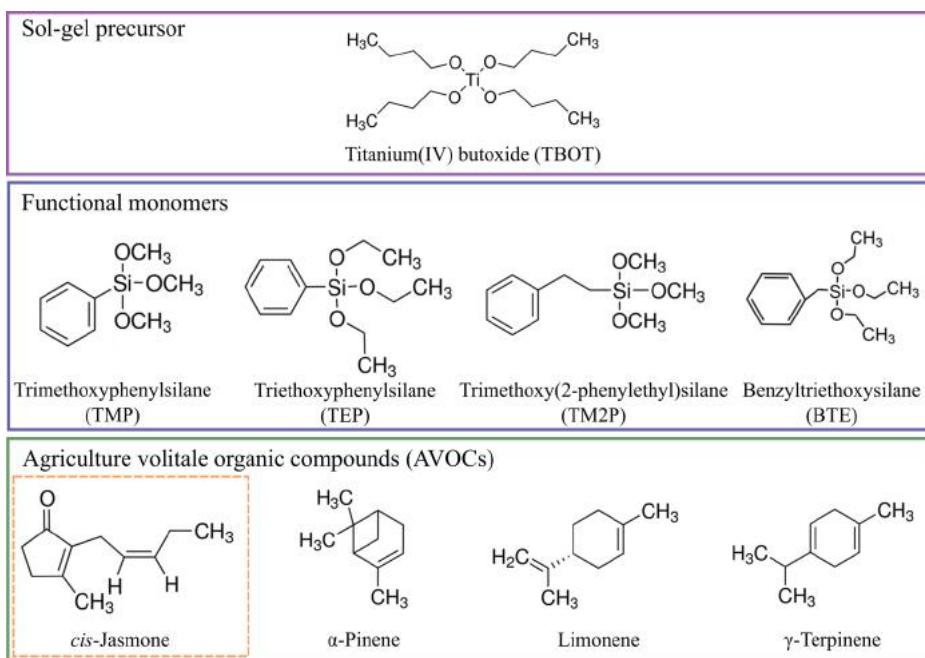


Fig. 5.2 Chemical structures of the matrix precursor, four functional monomers, and four types of PVOCs.

## 5.3 Results and discussion

### 5.3.1 Functional monomer selection

Strong noncovalent interactions between the functional monomers and the templates were critical for forming three-dimensional binding cavities during polymerization [216]. The functional monomer was a critical for MISGs preparation. CJ is very volatile, with a vapor pressure of 0.029 mm Hg at 25 °C. Therefore, it could be evaporated from the NISGs/MISGs layer with clean flowing air. Given the chemical structure of CJ (Figure 5.2), functional monomers with aromatic rings would be appropriate for cavity generation in the MISGs via electron, Van der Waals, and hydrogen-bond interactions [217, 218]. Therefore, four function monomers (TMP, TEP, TM2P and BTE) were considered (Figure 5.2).

Au nano-islands coated with various MISGs were optically characterized. Figure 5.3 plots differences in their UV-vis spectra. Relative to the bare sample, plasmon peaks in the NISG/MISG-coated samples shifted to the red because of the MISGs. In addition, relative to the MISG-modified sample, the surface plasmon (SP) peaks of the samples coated by MISG with functional monomers shifted to the blue. Their surface morphologies were imaged with SEM (Figure 5.4) and AFM (Figure 5.5, Table 5.1). SEM of the bare sample revealed that Au nano-islands were uniformly deposited on the substrate, which would induce

a stronger LSPR effect between Au nano-islands. The islands were covered by MISG/NISG films, which produced the diverse absorbance spectra. The surfaces of MISG-coated samples varied with functional monomers, which indicated that the sol-gel process was affected differently by the functional monomers (TMP, TEP, TM2P, and BTE).

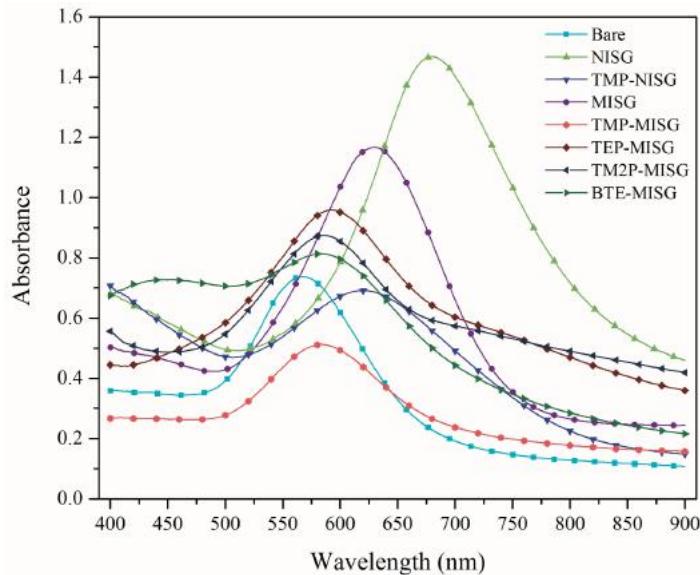


Fig. 5.3 Absorption spectra for the bare substrate, a NISGs coat, and MISGs coatings with different functional monomers.

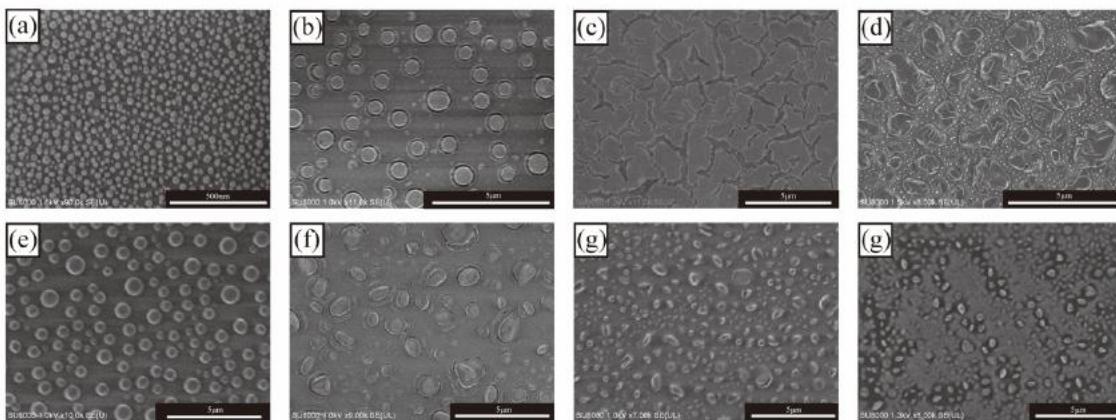


Fig. 5.4 SEM images of bare (a), NISG (b), TMP-NISG (c), MISG (d), TMP-MISG (e), TEP-MISG (f), TM2P-MISG (g) and BTE-MISG (h) coated samples.

Attenuated total reflection FT-IR spectra were plotted in Figure 5.6. The broadened band at  $1400\text{-}1480\text{ cm}^{-1}$  was attributed to stretching of the benzenoid ring. There were benzenoid ring peaks in the functional monomers contained in the NISGs/MISGs; thus, the

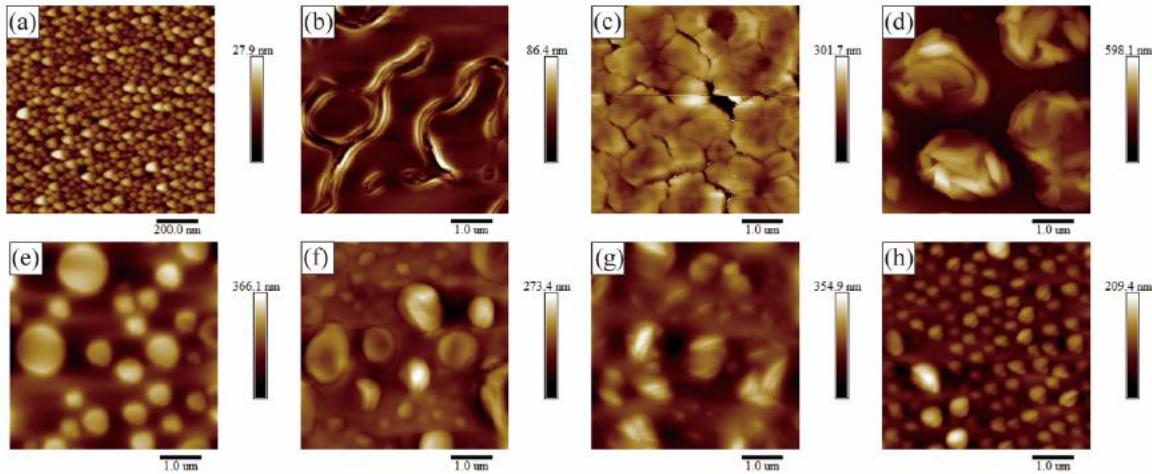


Fig. 5.5 AFM images of bare (a), NISG (b), TMP-NISG (c), MISG (d), TMP-MISG (e), TEP-MISG (f), TM2P-MISG (g) and BTE-MISG (h) coated samples.

Table 5.1 The roughness of Bare, NISGs and MISGs with different functional monomers coated samples.

Roughness	Bare	NISG	TMP-NISG	MISG	TMP-MISG	TEP-MISG	TM2P-MISG	BTE-MISG
$R_q$ (nm)	3.89	11.70	36.00	102.00	57.80	32.30	48.90	27.00
$R_a$ (nm)	3.08	8.59	25.60	86.60	46.90	23.10	37.30	20.60

monomers were polymerized in the sol-gel films. The strong band at 1620-1750  $\text{cm}^{-1}$  was attributed to C=O stretching of CJ, as was observed in the MISG-TMP- and MISG-TM2P-coated samples after gas absorption. Hence, CJ molecules diffused to the MISGs cavities (Figure 5.2). In addition, lower-energy peaks appeared in the MISG- and MISG-BTE-coated samples. In contrast, there were no C=O absorption peaks in the NISG-modified samples, which indicated less gas absorption.

In-situ responses of NISG-, TMP-NISG-, MISG-, TMP-MISG-, TEP-MISG-, TM2P-MISG-, and BTE-MISG-coated samples to CJ vapor ( $10.56 \pm 0.96$  ppm) were measured by absorption changes  $\Delta A$ , given by Formula 5.1.

$$\Delta A = A_{\text{gas}} - A_{\text{air}} \quad (5.1)$$

where  $A_{\text{air}}$  was the absorption in air, and  $A_{\text{gas}}$  was the absorption in the PVOC vapor. Both the CJ vapor generation and the sensing measurements were performed at room temperature.

Figure 5.7a illustrated that no responses to CJ vapor were observed for NISG-modified LSPR sensors, which indicated that CJ absorption by the pure sol-gel matrix was poor, as

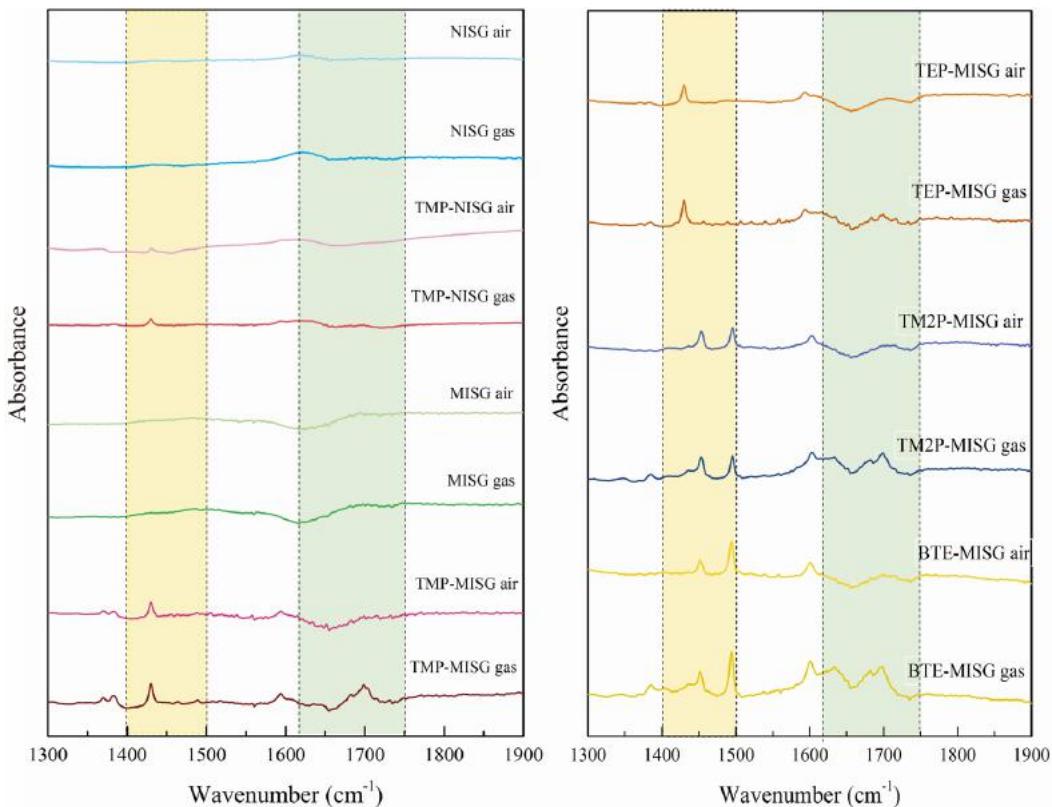


Fig. 5.6 FT-IR spectra of NISG-, TMP-NISG-, MISG-, TMP-MISG-, TEP-MISG-, TM2P-MISG-, and BTE-MISG-coated samples before and after CJ vapor absorption.

reported previously [219]. The responses to CJ vapor for NISG/MISG-coated samples are summarized in Figure 5.7b. MISGs without functional monomers exhibited lower responses than did the TMP-MISGs and TM2P-MISGs, indicating that the functional monomers improved the MISG responses and response speeds to target molecules. The TEP and BTE functional monomers had lower responses, indicating that functional monomers associated with -OCH<sub>2</sub>CH<sub>3</sub> (TEP and BTE) were less effective in enhancing CJ imprinting than those containing -OCH<sub>3</sub> (TMP and TM2P). In addition, NISG-coated samples with the functional monomer TMP were studied. As shown in Figure 5.7a, the affinity of NISGs to CJ vapor increased by adding TMP because of the TMP-NISG matrix effect. The response of TMP-MISGs was 2.25 times that of the TMP-NISGs. The larger sensitivity of the TMP-MISGs was attributed to both the matrix effect and imprinting. Therefore, TMP appeared to be the optimal functional monomer for the CJ-MISGs.

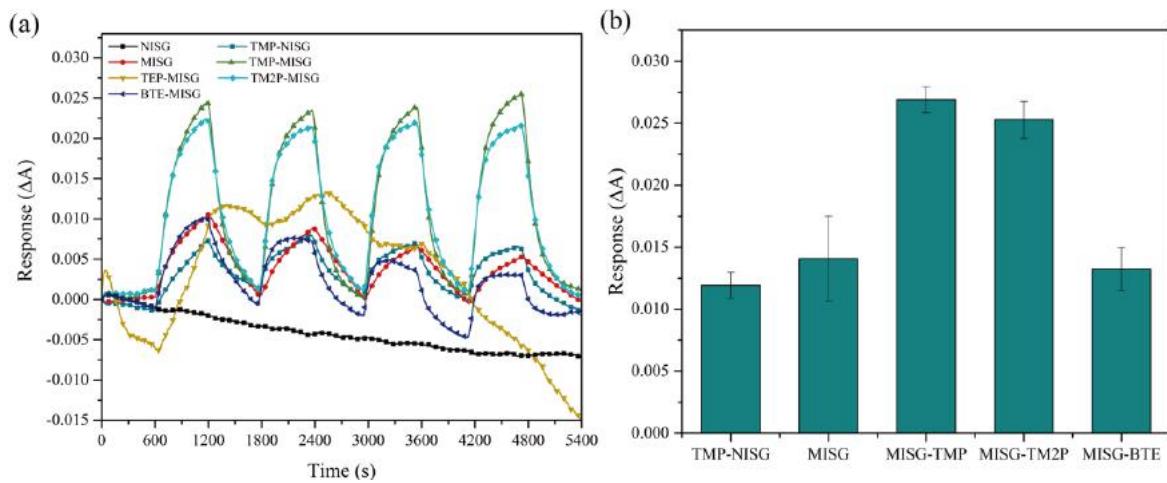


Fig. 5.7 Real-time responses for NISG-, MISG-, TMP-MISG-, TEP-MISG-, TM2P-MISG-, and BTE-MISG-coated samples (a) and their quantitative responses (b). Responses were obtained for a 600-s CJ vapor flow ( $10.56 \pm 0.96$  ppm), followed by a 600-s air flow.

### 5.3.2 Optimization of the TBOT/TMP ratio

Imprinting with MIP materials was affected by the ratio of the matrix to the functional monomers [219]. Specifically, the ratios of matrix material (TBOT) to functional monomer (TMP) were adjusted to be 75/125, 100/100, 125/75, 150/50, and 175/25 ( $\mu\text{L}$ , v/v) to optimize the LSPR sensor for CJ detection. UV-vis absorption spectra vs. the ratios were plotted in Figure 5.8, where  $A_{max}$  increased and its wavelength was blue-shifted as the amount of TMP decreased. SEM and AFM images for these MISGs (Figure 5.9, 5.10, Table 5.2) revealed that the size of the  $\text{TiO}_2$  sol-gel was affected by the ratios. In particular, larger  $\text{TiO}_2$  discs were formed by a higher TBOT/TMP ratio.

In-situ responses of LSPR sensors coated with MISGs having the various TBOT/TMP ratios were plotted in Figure 5.11. The results indicated that sensitivity was affected, and that Au nano-islands coated by MISG-TMP with the ratio TBOT/TMP=150/50 had the highest CJ sensitivity.

Table 5.2 The roughness of MISGs coated samples with different ratios of TBOT to TMP (v:v).

Roughness	The ratio of TBOT to TMP (v:v)				
	75:125	100:100	125:75	150:50	175:25
$R_q$ (nm)	92.1	98.4	106.0	57.8	96.3
$R_a$ (nm)	67.2	74.5	80.4	46.9	76.6

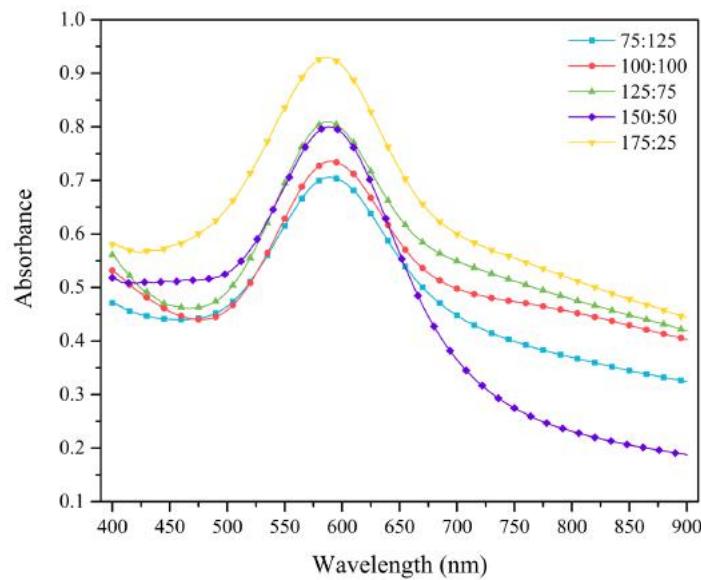


Fig. 5.8 Absorption spectra for TMP-MISG-coated samples at different TBOT/TMP (v/v).

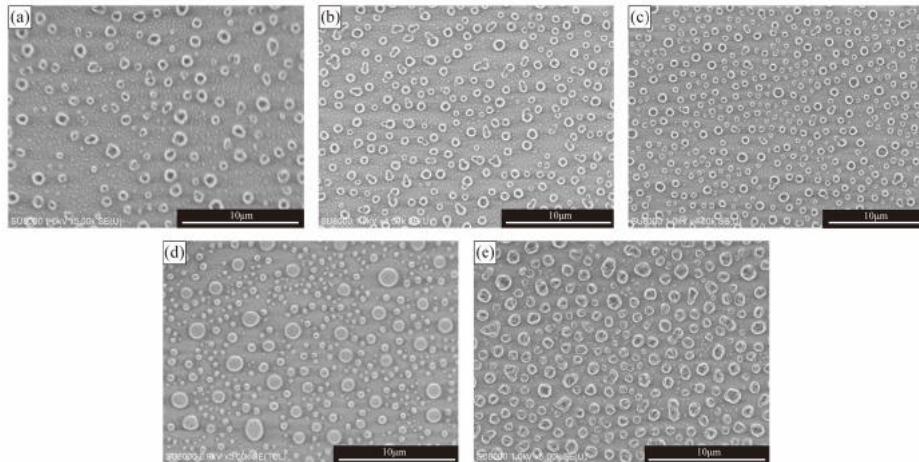


Fig. 5.9 SEM images of samples coated with MISG at TBOT/TMP=75/125 (a), 100/100 (b), 125/75 (c), 150/50 (d), 175/25 (e) ( $\mu\text{L}$ , v:v).

### 5.3.3 CJ detection with a MISG-coated Au nano-island sensor

Both the shape and correct orientation of the functional groups enabled selective re-binding of the imprinted target by subsequent removal of the template binding sites during sol-gel processing. However, other PVOCS, specially terpene molecules, were present in the ambient environment and could be captured in MISG cavities. To evaluate the selectivity of the MISG-coated sensors, the three primary terpenes  $\alpha$ -pinene, limonene, and  $\gamma$ -terpinene (Figure 5.2) were tested as interference PVOCS. All the responses were normalized to these concentrations (Formula 5.2)

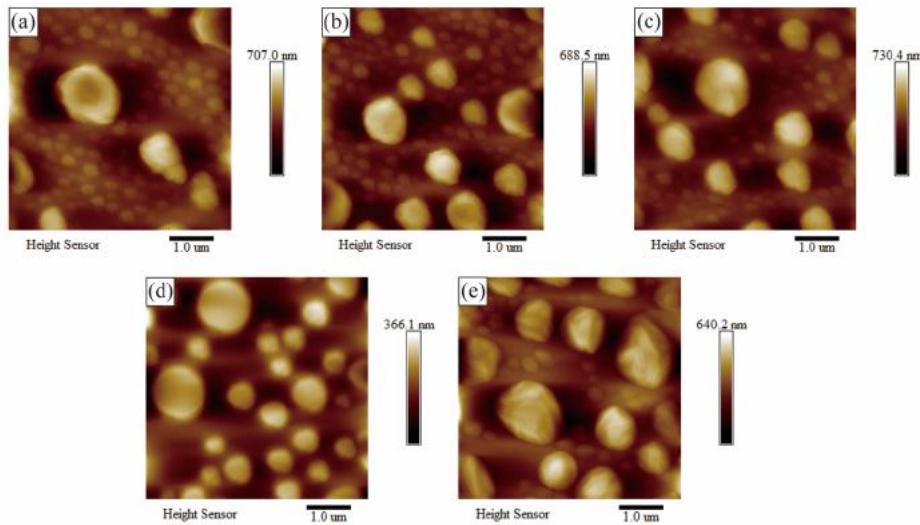


Fig. 5.10 SEM images of samples coated with MISG at TBOT/TMP=75/125 (a), 100/100 (b), 125/75 (c), 150/50 (d), 175/25 (e) ( $\mu\text{L}$ , v:v).

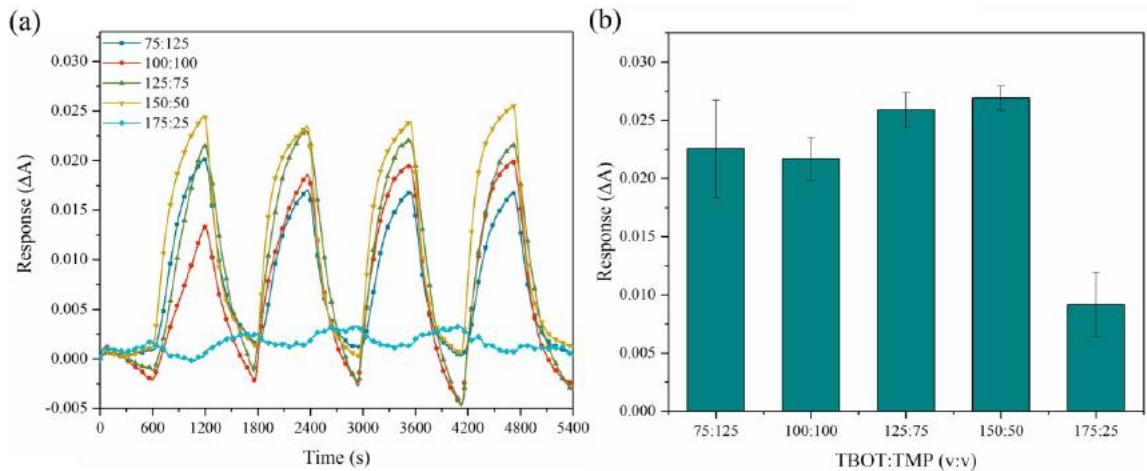


Fig. 5.11 Real-time responses of samples coated with TMP-MISG at TBOT/TMP=75/125, 100/100, 125/75, 150/50, 175/25 (v/v) (a) and their response summary (b). Responses were obtained for a CJ vapor flow for 600 s ( $10.56 \pm 0.96$  ppm), followed by an air flow for 600 s.

$$R_{normalized} = R / \lg(C_{test}) \quad (5.2)$$

where  $R$  was the original LSPR sensor response.  $C_{test}$  was the testing concentration of each PVOCS; specifically,  $10.56 \pm 0.96$  ppm,  $187.72 \pm 33.64$  ppm,  $971.20 \pm 58.89$  ppm, and  $750.05 \pm 36.25$  ppm for CJ,  $\alpha$ -pinene, limonene and  $\gamma$ -terpinene vapors, respectively. The ratios of CJ to interference PVOCS ( $\alpha$ -pinene, limonene and  $\gamma$ -terpinene) were calculated as 1:18, 1:92, 1:71, respectively.

The normalized responses of the TMP-MISG-coated LSPR sensors to these interference PVOCS were plotted in Figure 5.12. The CJ response of the molecularly imprinted sensor was much higher than that for the interfering molecules. This selectivity of the nanocomposite MISG-LSPR electrode was attributed to cavities that matched the shape and size of the CJ molecule. It has been reported that concentrations in agricultural environments of CJ,  $\alpha$ -pinene, limonene, and  $\gamma$ -terpinene are about 85 ppm, 941 ppm, 171 ppm, and 371 ppm, respectively. Therefore, the ratios of CJ to  $\alpha$ -pinene, limonene, and  $\gamma$ -terpinene in agricultural environments can be calculated as 1:11, 1:2, and 1:4, respectively, which are larger than that in present study. It indicated that the sensor developed in present study would have enough selectivity in agricultural environments.

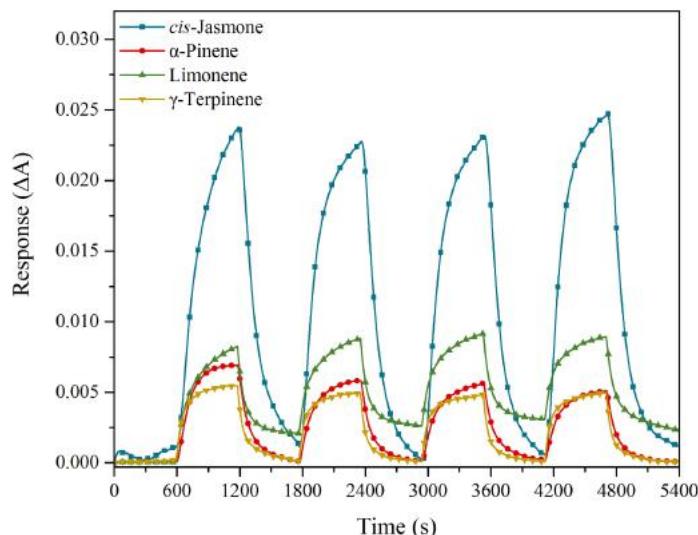


Fig. 5.12 Real-time responses of TMP-MISG-modified Au-islands to four types of PVOCS.

Quantitative CJ detection was performed by monitoring changes in the UV-vis absorption spectra for different CJ concentrations. A baseline signal was first collected in clean air. The LSPR sensors were then exposed to different CJ concentrations and the absorbances were obtained, as shown in Figure 5.13. A linear calibration curve was fitted with Formula 5.3.

$$y = 0.0132 + 0.00233x, R^2 = 0.9689; n = 3 \quad (5.3)$$

The limit of detection LOD=3.494 ppm was calculated as the CJ concentration that resulted in a signal that was at least three times the baseline noise ( $3\delta/m$ ), where  $m=0.00233$  was the slope of the calibration curve and  $\delta$  was the standard deviation. Previous studies reported that CJ concentrations in shoots, leaves, and flowers were 0-85, 0-125 and 0-114 ppm, respectively [220–222]. Therefore, the sensor here would be sensitive enough for agricultural CJ detection.

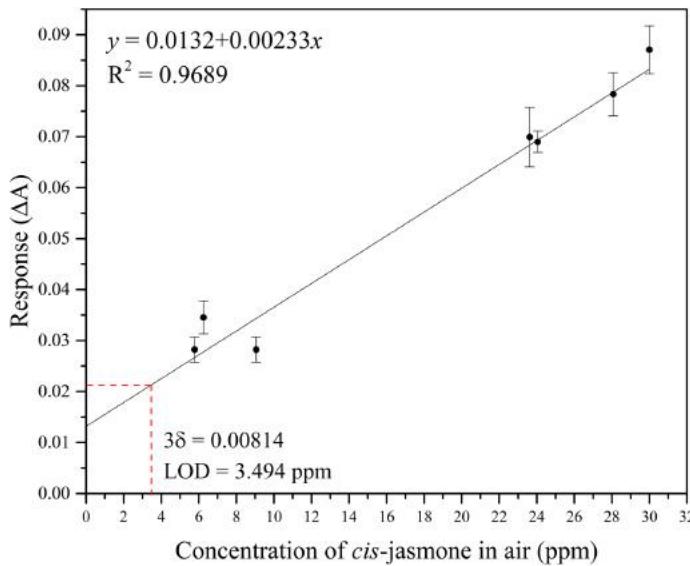


Fig. 5.13 Linear response vs. CJ concentration in air. The limit of CJ detection (LOD) was 3.5 ppm.

### 5.3.4 Discussion

In this part, the sensitivities of MISGs contained 4 types of functional monomers were detected and compared. Just as former discussed, the absorption of MISGs were consisted by matrix effect and imprinted effect. The functional group in functional monomers played an important role in enhancing the selectivity/affinity to molecules with structure similarly for MISGs. The interaction between VOC molecules and MISGs layer was contributed by electron, Van der Waals and  $\pi$ - $\pi$  effects. It indicated that functional monomers contained different types functional monomers (such as amino group, carboxyl group, hydroxyl or benzene ring etc.) would be attempted for developing olfaction inspired odor sensing system.

## 5.4 Conclusion

LSPR sensors based on MISG-modified Au nano-islands was demonstrated for CJ vapor detection. The absorption of interference PVOCS ( $\alpha$ -pinene, limonene and  $\gamma$ -terpinene) by MISG<sub>CJ</sub> were tested for CJ selectivity evaluation. The results demonstrated that sensor covered by MISGs with TMP had the strongest responses. In addition, the matrix to functional monomer ratio was optimized for better responsively. Under optimal conditions, the volume ratio TBOT/TMP=150/50 resulted in a 3.494-ppm LOD for CJ vapor. This was attributed to a porous imprinted composite film with CJ-selective binding sites. Real-time responses

of the sensors displayed good selectivity, broad linearity, and repeatability. In conclusion, they are expected to provide sensitive PVOC detection for agricultural applications.



# **Chapter 6**

## **LSPR sensor array coated AuNPs@MISGs for PVOCs recognition**

### **6.1 Introduction**

Chapter 5 revealed that a molecularly imprinted sol–gel (MISG)-coated LSPR sensor was effective for cis-jasmone (CJ) vapor detection. Although the selectivity of the sensor was enhanced by the MISG coating, its sensitivity was decreased. Recently, molecularly imprinted polymers (MIPs) have been combined with gold nanoparticles (AuNPs) to amplify the LSPR causing the hot spot effect [223]. Based on the LSPR coupling between AuNPs in an MIP and a thin gold film, refractive index (RI) changes can be enhanced for use in optical sensing [224]. Therefore, AuNPs embedded in an MISG (denoted as AuNPs@MISG) should be effective at enhancing the signal intensity of a sensor while maintaining high selectivity. In addition, considering the complex nature of the agricultural environment, a multi-channel sensor array combined with a pattern recognition method should be developed for plant volatile organic compounds (PVOC) identification.

The primary goal of this chapter is to develop an AuNPs@MISG LSPR sensor platform to detect and identify PVOCs. As illustrated in Figure 7.1, the sensor consists of an LSPR sensing layer coated with an MISG layer. The LSPR sensing layer is fabricated by vacuum sputtering and annealing. An AuNPs@MISG layer is formed around the AuNPs in the sensing layer by spin coating. Sensor responses are captured by monitoring changes in the RI by absorbance spectra. AuNPs doped in the MISG are expected to increase the signal intensity through the hot spot effect. Critical parameters of the AuNPs@MISG (AuNP size, amount, and spin coating speed) are tuned to optimize sensitivity. By spin coating MISG solutions with diverse template molecules on AuNPs, a multi-channel optical sensor platform

for PVOC identification is constructed. The identification capability of the sensor platform is tested using four PVOCs (CJ,  $\alpha$ -pinene, limonene, and  $\gamma$ -terpinene) and their binary mixtures. Principal component analysis (PCA) and linear discriminant analysis (LDA) are used to visualize the cluster trends of vapor samples in low dimensions. To assess the potential of the sensor platform, three common supervised approaches, LDA, k-nearest neighbor (KNN), and naïve Bayes classifier (NBC), are used to establish PVOC identification models (Figure 6.2). The objective of this study is to develop a new sensing strategy for PVOC detection in agricultural applications.

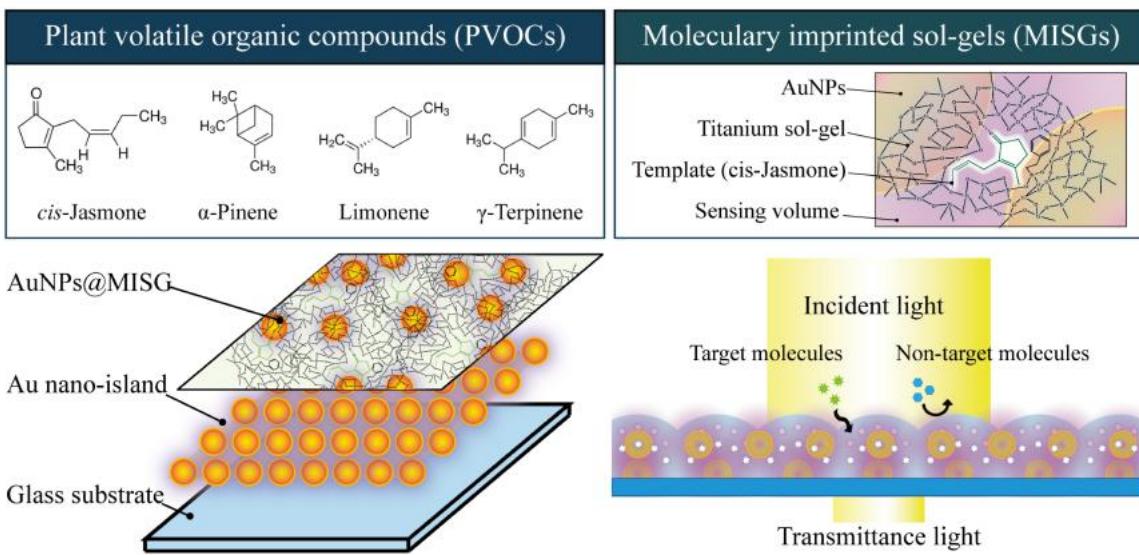


Fig. 6.1 Schematic of AuNPs@MISG-coated Au nano-islands for selective PVOC detection.

## 6.2 Experimental

### 6.2.1 Materials, chemicals and instrumentations

Titanium tetrabutoxide (TBOT), iso-propanol, CJ, limonene,  $\gamma$ -terpiene, titanium tetrachloride ( $TiCl_4$ ), acetone, and ethanol were purchased from Wako Pure Chemical Industries Co., Ltd. (Osaka, Japan). (3-Aminopropyl)triethoxysilane (APTES) was purchased from Shin-Etsu Chemical Co., Ltd (Tokyo, Japan). AuNP suspensions (NP diameter: 10, 20, 30, and 40 nm),  $\alpha$ -pinene, and trimethoxyphenylsilane (TMP) were purchased from Sigma-Aldrich (St. Louis, MO, USA). All reagents were used as received. Scanning electron microscopy (SEM; SU8000, Hitachi, Japan) was used to image sensor morphology.

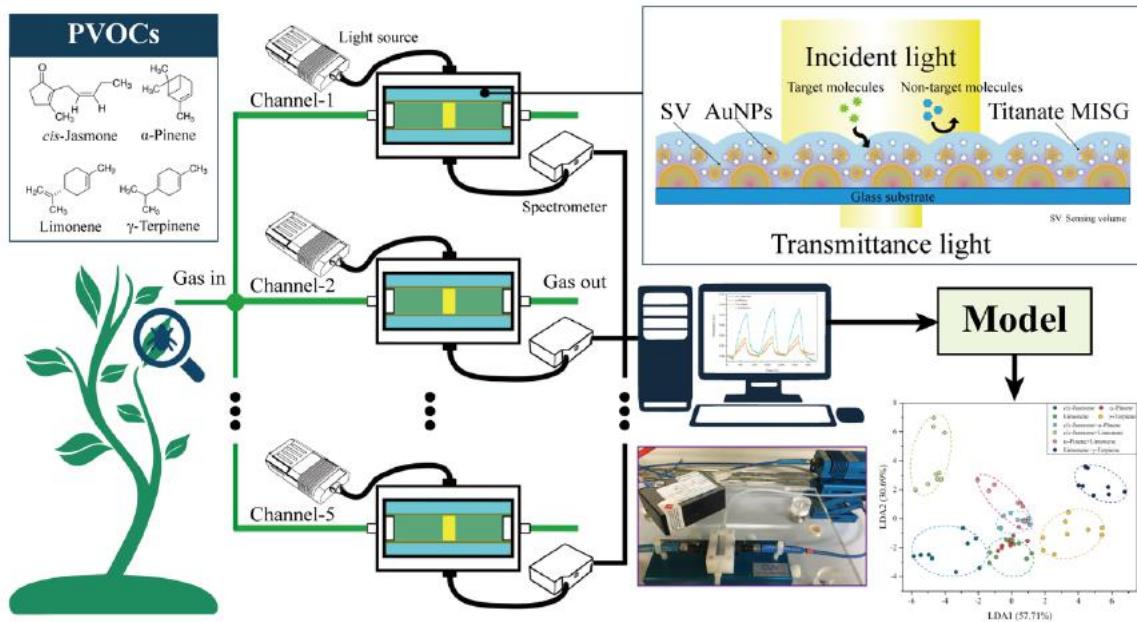


Fig. 6.2 Schematic of AuNPs@MISG-coated Au nano-islands sensor array for selective PVOC detection.

### 6.2.2 Preparation of AuNPs@MISGs reaction solution

A typical experiment was described here. First, TBOT (0.441 mmol, 150  $\mu$ L) was dissolved in isopropanol (2 mL) as a precursor. Then, CJ,  $\alpha$ -pinene, limonene, or  $\gamma$ -terpinene (50  $\mu$ L) as a template material and TMPS (0.252 mmol, 50  $\mu$ L) as a functional monomer were added with stirring.  $TiCl_4$  (0.132 mmol, 25  $\mu$ L) was added to initiate the reaction and then the MISG reaction solution was pre-hydrolyzed in a water bath at 60 °C for 1 h with stirring. The mixture was vigorously stirred for 8 h at room temperature (20 °C) to complete the MISG reaction. Finally, AuNP suspension (50  $\mu$ L) was added to the hydrolyzed MISG solution while stirring. Before spin coating, the reaction solution was stirred at room temperature for 8 h.

### 6.2.3 AuNPs@MISGs coated LSPR sensor fabrication

Briefly, a glass substrate (12×9 mm) was ultrasonically cleaned in ultrapure water, acetone, and ethanol and then immersed in a 1:10 (v:v) APTES/ethanol solution for 8 h. After being cleaned with ethanol and dried with flowing nitrogen, a 3-nm-thick layer of AuNPs was deposited using a quick coater (SC-701 HMCII, Sanyu Electron, Japan). The substrate was heated in a muffle furnace (SSTS-13K, ISUZU, Seisakusho, Japan) at 500 °C for 2 h and then cooled to room temperature. The sample was sputtered and annealed again under the

same conditions to form a high-sensitivity LSPR substrate. The MISG reaction solution (20  $\mu\text{L}$ ) was then spin-coated on the AuNP layer for 1 min. Finally, the sample was annealed at 130 °C for 1 h to complete MISG fabrication and evaporate the template molecules. All samples were stored under vacuum to remove VOCs from the MISG layer.

## 6.3 Results and discussion

### 6.3.1 Effect of Au NP size on LSPR

Because the size of AuNPs is a critical factor affecting their LSPR signals, MISG-coated AuNPs with diameters of 10, 20, 30, and 40 nm were considered. The typical features of the MISG-coated samples were analyzed using UV-vis spectroscopy and SEM, as shown in Figure 6.3 and 6.4, respectively. Relative to that of the MISG-coated sample, the LSPR peaks of the AuNPs@MISG-coated samples were blue-shifted. In addition, the surfaces of the AuNPs@MISG-coated samples varied with the AuNP size, indicating that the sol-gel process was affected by the diameter of the AuNPs. The responses of the MISG- and AuNPs@MISG-coated samples to CJ vapor were measured by the change of absorption  $\Delta A$ , which was calculated using Formula n 5.1.

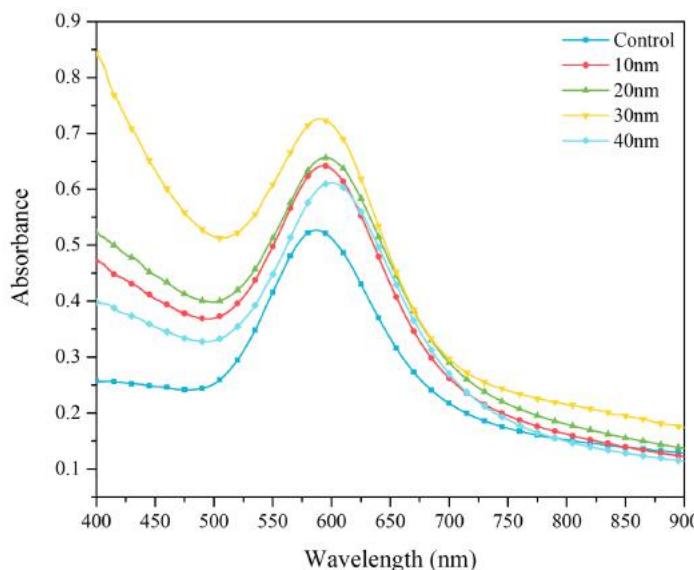


Fig. 6.3 Absorption spectra of MISG-coated samples with different size of AuNPs.

The sensitivity to CJ vapor of various MISG-coated samples are summarized in Figure 6.5. The MISG without AuNPs offered lower sensitivity than that of the AuNPs@MISG-coated samples, revealing that the AuNPs improved the response of the MISG to target

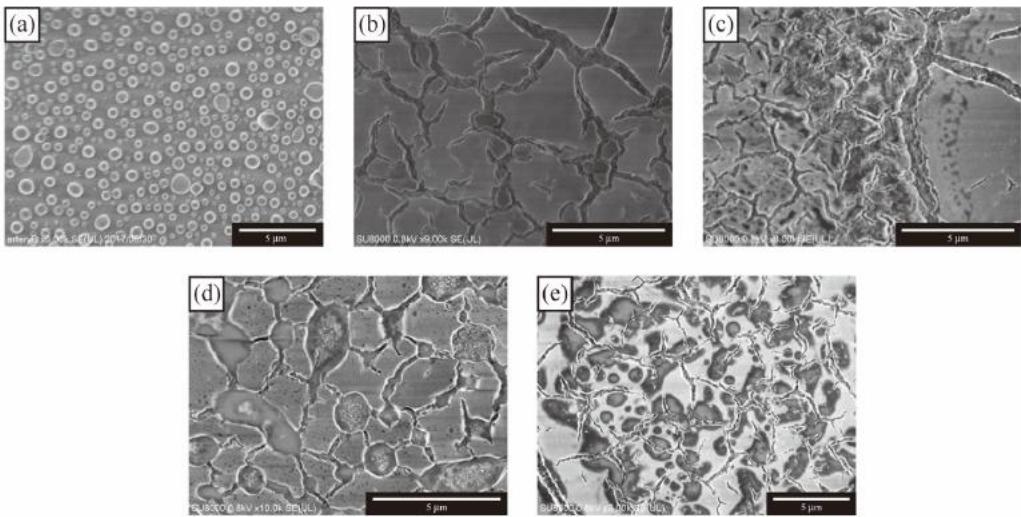


Fig. 6.4 SEM images of control (a) and MISGs with 10-nm (b), 20-nm (c), 30-nm (d), and 40-nm (e) AuNPs.

molecules. The response of AuNPs@MISG-coated sensor with 30-nm AuNPs was 6.33 times that of the one without NPs. An SEM image of an AuNP layer is illustrated in Figure 6.6a. The particle size distribution histogram of the spherical AuNPs was analyzed by ImageJ (Figure 6.7). The analysis indicated that the diameter of the AuNPs on the substrate was  $34.13 \pm 9.41$  nm, which is close to that of the AuNPs in the MISG (30 nm). The high sensitivity of the sensor was therefore caused by hot-spot coupling between the AuNPs on the substrate surface and those in the MISG [225]. Therefore, 30 nm is the optimal size for the AuNPs in the MISG.

### 6.3.2 Optimization of the amount of AuNPs

To obtain the best performance for PVOC detection, the effect of the amount of AuNPs on sensing behavior was investigated. UV-vis spectra and SEM images of MISG samples with different amounts of AuNPs are presented in Figure 6.8 and 6.9, respectively. The responses of the AuNPs@MISG-coated samples to CJ vapor were measured and are depicted in Figure 6.10. Evidently, the sensitivity of the sensors increased with the AuNP concentration initially and then decreased. The results revealed that the sensor coated with the MISG containing 20 μL of 30-nm AuNPs had the highest sensitivity of those investigated.

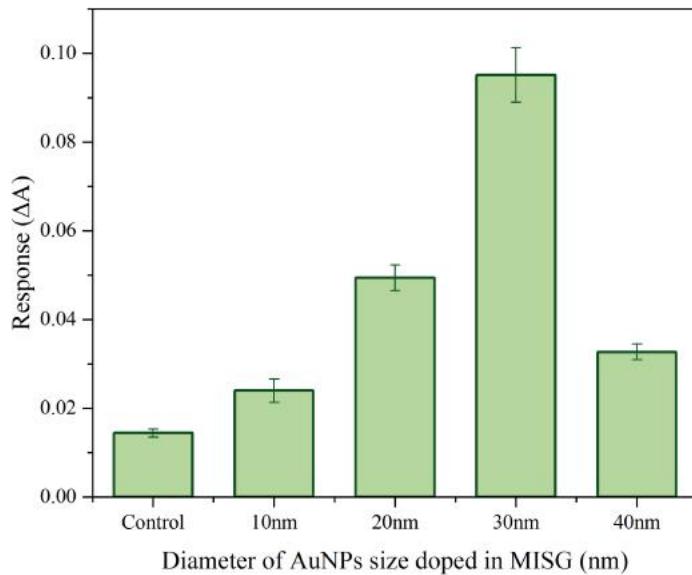


Fig. 6.5 Response effected by AuNPs size in MISGs.

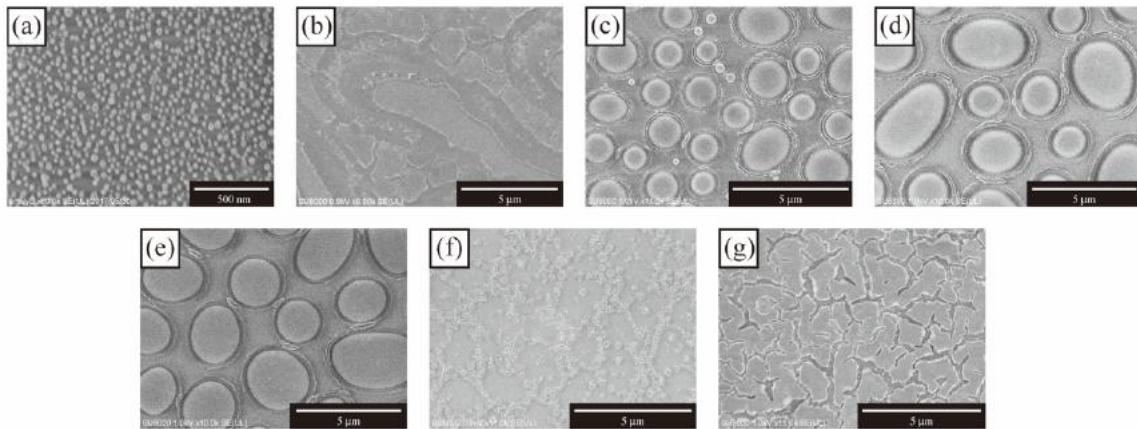


Fig. 6.6 SEM images of bare (a), MISG<sub>cis</sub>-jasmonate (b), MISG<sub>α</sub>-pinene (c), MISG<sub>limonene</sub> (d), and MISG<sub>γ</sub>-terpiene (e), NISG (f), and AuNPs doped NISG-coated (g) sample.

### 6.3.3 Optimization of spin coating speed

The thickness of the sensing film influences the sensitivity of LSPR sensors [132]. Here, spin coating speeds of 2000, 3000, 4000, and 5000 rpm were selected to optimize the thickness of the AuNPs@MISG coatings. The sensitivities to CJ vapor of samples coated with the optimal AuNPs@MISG solution at different coating speeds are illustrated in Figure 6.11. A thinner MISG layer exhibited lower selectivity for target molecules. However, a layer that was too thick would have a long recovery time [144]. Considering the observed sensor responses, the optimal spin coating speed was selected as 3000 rpm in the present study.

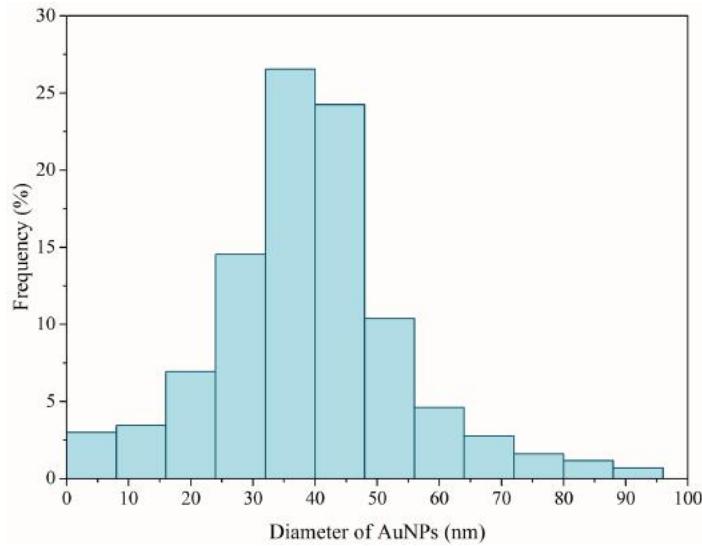


Fig. 6.7 Particle size distribution histogram of spherical AuNPs determined from bare sample (Figure 6.6 a). SEM image was analyzed by ImageJ. The diameter of AuNPs was  $34.13 \pm 9.41$  nm.

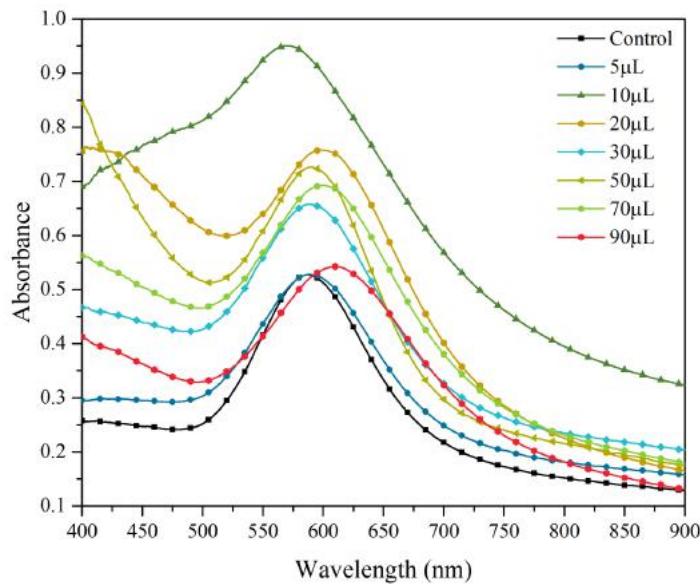


Fig. 6.8 Absorption spectra of MISG-coated samples with different amount of 30-nm AuNPs.

### 6.3.4 CJ detection with the optimized AuNPs@MISG-coated LSPR sensor

To evaluate the interference immunity of the AuNPs@MISG-coated sensor, it was exposed to the three primary PVOCs:  $\alpha$ -pinene, limonene, and  $\gamma$ -terpinene. All the responses were normalized to the PVOC concentrations as chapter 5. The normalized in situ response of the

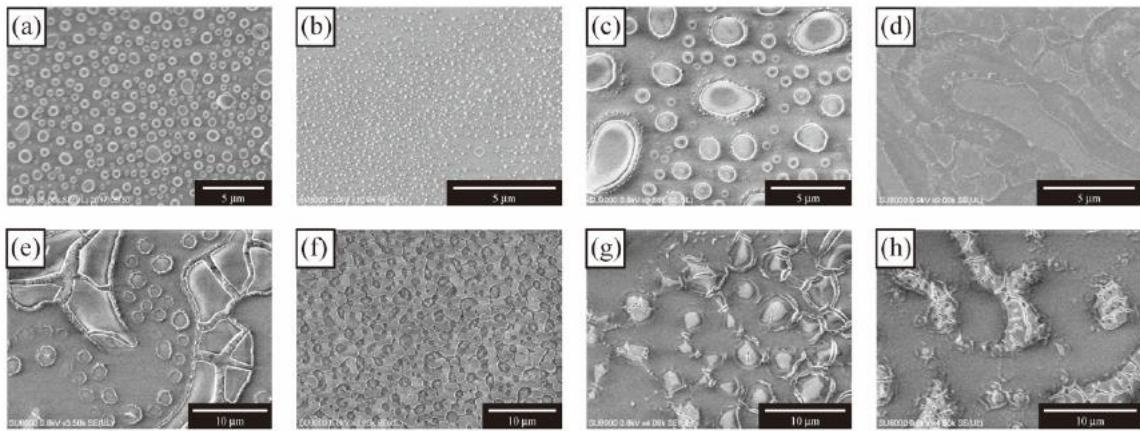


Fig. 6.9 SEM images of control (a), and MISGs with 30 nm AuNPs 5- $\mu\text{L}$  (b), 10- $\mu\text{L}$  (c), 20- $\mu\text{L}$  (d), 30- $\mu\text{L}$  (e), 50- $\mu\text{L}$  (f), 70- $\mu\text{L}$  (g), and 90- $\mu\text{L}$  (h).

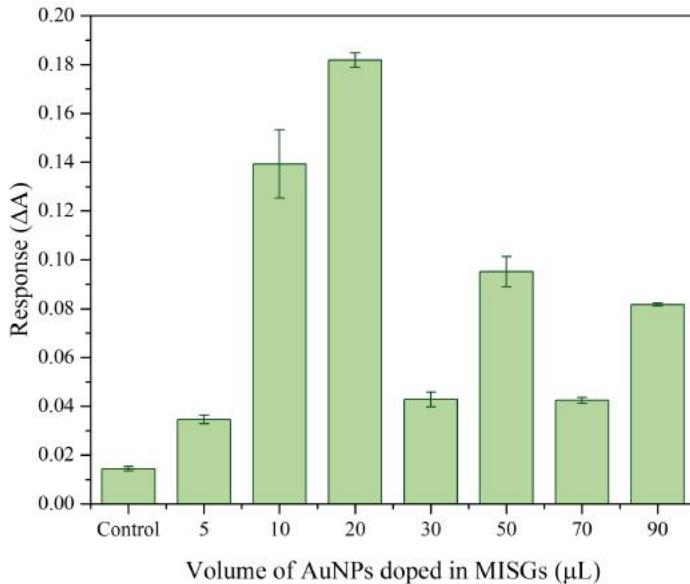


Fig. 6.10 Response effected by Au nanoparticles (30 nm) amount in MISGs.

optimized AuNPs@MISG-coated LSPR sensor to these interferents is shown in Figure 6.12. The response to CJ was much higher than that to the interfering PVOCs. This indicates that the developed sensor has sufficient interference immunity for use in agricultural applications. A linear train was fitted with Formula 6.1.

$$y = 0.03758 + 0.00278x, R^2 = 0.9402; n = 3 \quad (6.1)$$

The limited of detection LOD=3.07 ppm was calculated as the CJ concentration that resulted in a signal that was at least three times the baseline noise ( $3\sigma/m$ ).

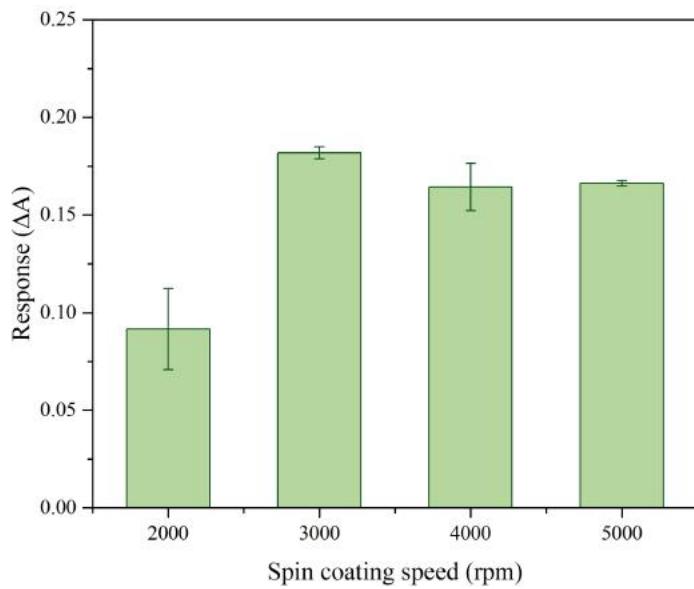


Fig. 6.11 Response effected by spin coating speed.

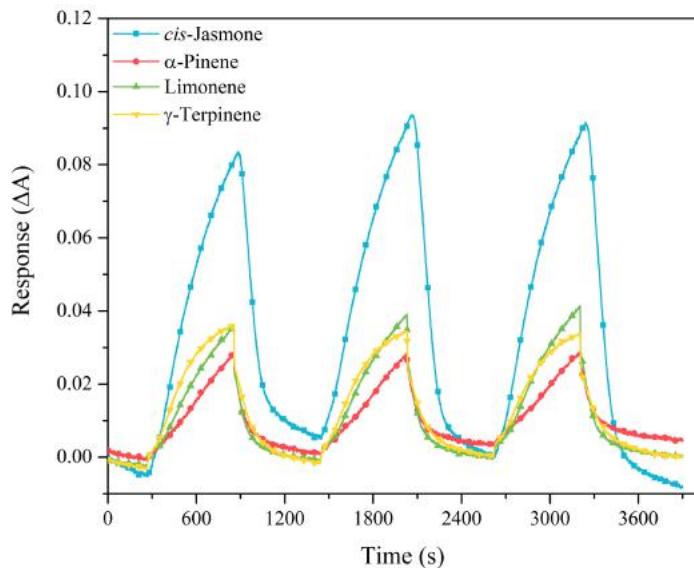


Fig. 6.12 Real-time responses of AuNPs@MISG<sub>CJ</sub>-modified Au-islands to four types of PVOCs. Responses were normalized by concentration.

### 6.3.5 AuNPs@MISGs-coated LSPR sensor array

Although the sensitivity of the sensor was enhanced because of the hot spot effect, the response intensity of the sensor was affected by the high concentrations of interfering PVOCs. To address this problem, an AuNPs@MISG LSPR sensor array was constructed for identification and detection of PVOCs. Four types of AuNPs@MISG reaction solutions were

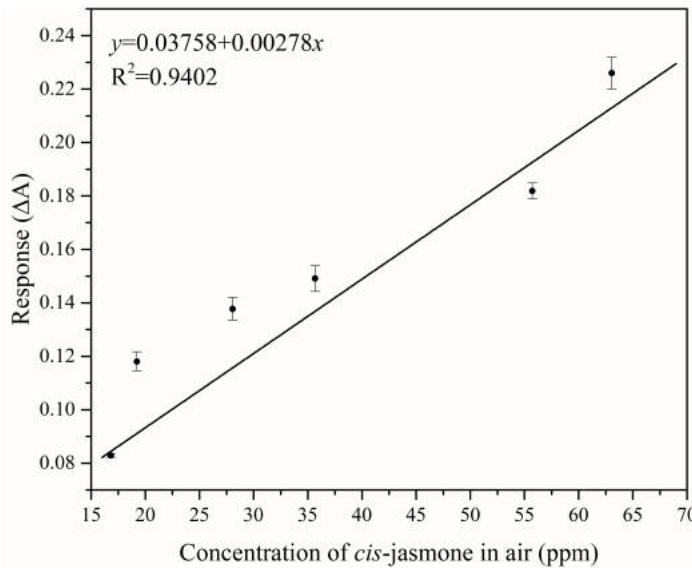


Fig. 6.13 Linear responses vs. CJ concentration in air. The limit of detection (LOD) was 3.07 ppm.

prepared using the optimal preparation conditions. By spin coating these solutions at 3000 rpm on the AuNP-coated surface, a sensor array was developed for pattern recognition of PVOCs (Figure 6.6). The sensor array included five channels: bare, MISG<sub>CJ</sub>, MISG <sub>$\alpha$ -pinene</sub>, MISG<sub>limonene</sub>, and MISG <sub>$\gamma$ -terpinene</sub>.

The response characteristics of the array to four single PVOCs (CJ, $\alpha$ -pinene, limonene, and  $\gamma$ -terpinene) and four binary mixtures (CJ+ $\alpha$ -pinene, CJ+limonene,  $\alpha$ -pinene+limonene, and limonene+ $\gamma$ -terpinene) were investigated. For each type of vapor, three vapor flow rates (0.3, 0.5, and 0.7 L/min) were measured at each concentration, and the measurements were repeated three times. Therefore, a data set containing 72 samples (8 PVOCs×3 flow rates×3 repeats) was collected for subsequent research. Both PVOC vapor generation and sensing measurements were performed at room temperature.

Correlation analysis was carried out initially to evaluate the relationship of each channel. The Pearson correlation coefficient was calculated using Formula 6.2 [142].

$$Cor(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (6.2)$$

where,  $\vec{x}$  and  $\vec{y}$  indicated response vectors for 2 channels.  $\bar{x}$  and  $\bar{y}$  are the mean values of vector  $\vec{x}$  and  $\vec{y}$ , respectively.  $N$  is 72, the dimension of vector  $\vec{x}$  or  $\vec{y}$ .

Figure 6.14 reveals there was a low correlation (under 0.4) between the bare and MISG-coated channels, indicating that the bare channel has different information to others. Because the bare LSPR sensor has low selectivity for the target molecules, the response of the bare channel would be related to the concentration of the PVOCs. In contrast, the responses of the MISG-coated channels contained contributions from the MIP and matrix effects, which increase selectivity for target molecules.

	CH1	CH2	CH3	CH4	CH5
CH1	1	0.06	-0.05	-0.17	-0.34
CH2	0.06	1	0.53	0.31	0.06
CH3	-0.05	0.53	1	0.59	0.1
CH4	-0.17	0.31	0.59	1	0.51
CH5	-0.34	0.06	0.1	0.51	1

Fig. 6.14 Correlation matrix for sensor array. CH1, CH2, CH3, CH4, and CH5 are bare, MISG<sub>CJ</sub>-, MISG <sub>$\alpha$ -Pinene</sub>-, MISG<sub>Limonene</sub>-, and MISG <sub>$\gamma$ -Terpinene</sub>-coated sample, respectively.

To visualize the clustering trends of vapor samples in low dimensions, PCA was performed on the normalized response matrix ( $M_{72 \times 5}$ ). The first three principal components (PCs), which captured 84.33% of the cumulative variance proportion of the response data, are plotted in Figure 6.15. In PCA space, only limonene (PC1-PC2 space, Figure 6.15a) and  $\gamma$ -terpinene (PC2-PC3 space, Figure 6.15c) samples formed differentiable clusters. It was difficult to distinguish clearly between the groups for other categories, such as the binary mixtures.

The response matrix was also analyzed by LDA. Unlike PCA, LDA is a supervised approach that aims to achieve an optimal transformation by minimizing the intragroup distances and maximizing the intergroup distances simultaneously [226–228]. Figure 6.16 shows 72 samples from eight clusters plotted in LDA spaces. In the LDA1-LDA2 space (Figure 6.16a), CJ,  $\gamma$ -terpinene, CJ+limonene, and limonene+ $\gamma$ -terpinene can be separated eas-

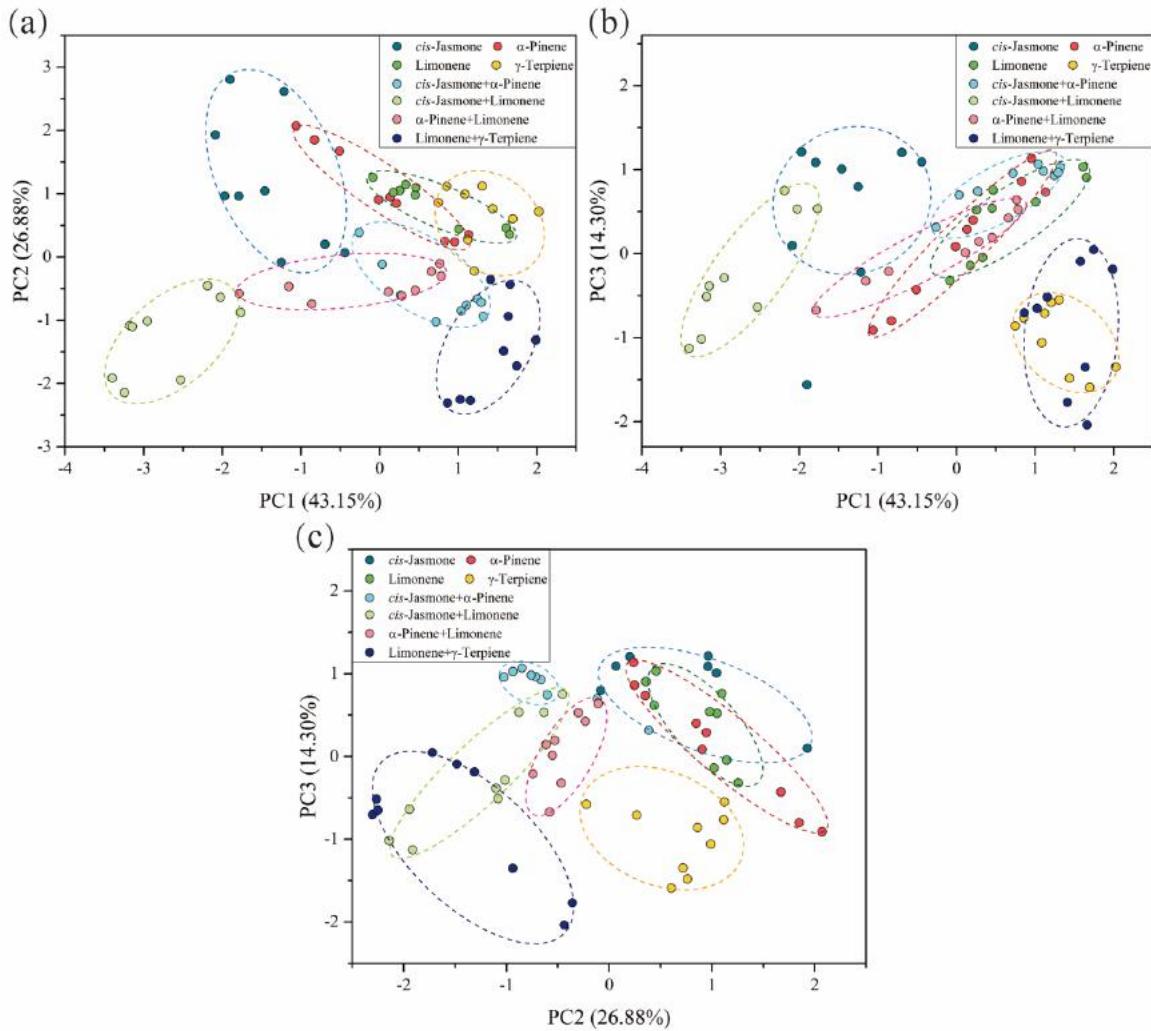


Fig. 6.15 PCA score plots of the responses of 72 samples from four PVOCs (CJ,  $\alpha$ -pinene, limonene, and  $\gamma$ -terpiene) and their binary mixtures (CJ+ $\alpha$ -pinene, CJ+limonene,  $\alpha$ -pinene+limonene, and limonene+ $\gamma$ -terpiene).

ily. Considering LDA1-LDA2 and LDA2-LDA3 (Figure 6.16b) simultaneously,  $\alpha$ -pinene and limonene occupied separate regions. We also found overlap between CJ+ $\alpha$ -pinene and  $\alpha$ -pinene+limonene in all LDA spaces. This means that it is hard to separate the PVOCs in low-dimension spaces. Therefore, pattern recognition approaches need to be explored and evaluated.

### 6.3.6 Identification model calibration

To investigate the ability of the AuNPs@MISG-coated LSPR sensor array to discriminate various types of PVOCs, three commonly used classification frameworks (LDA, KNN, and

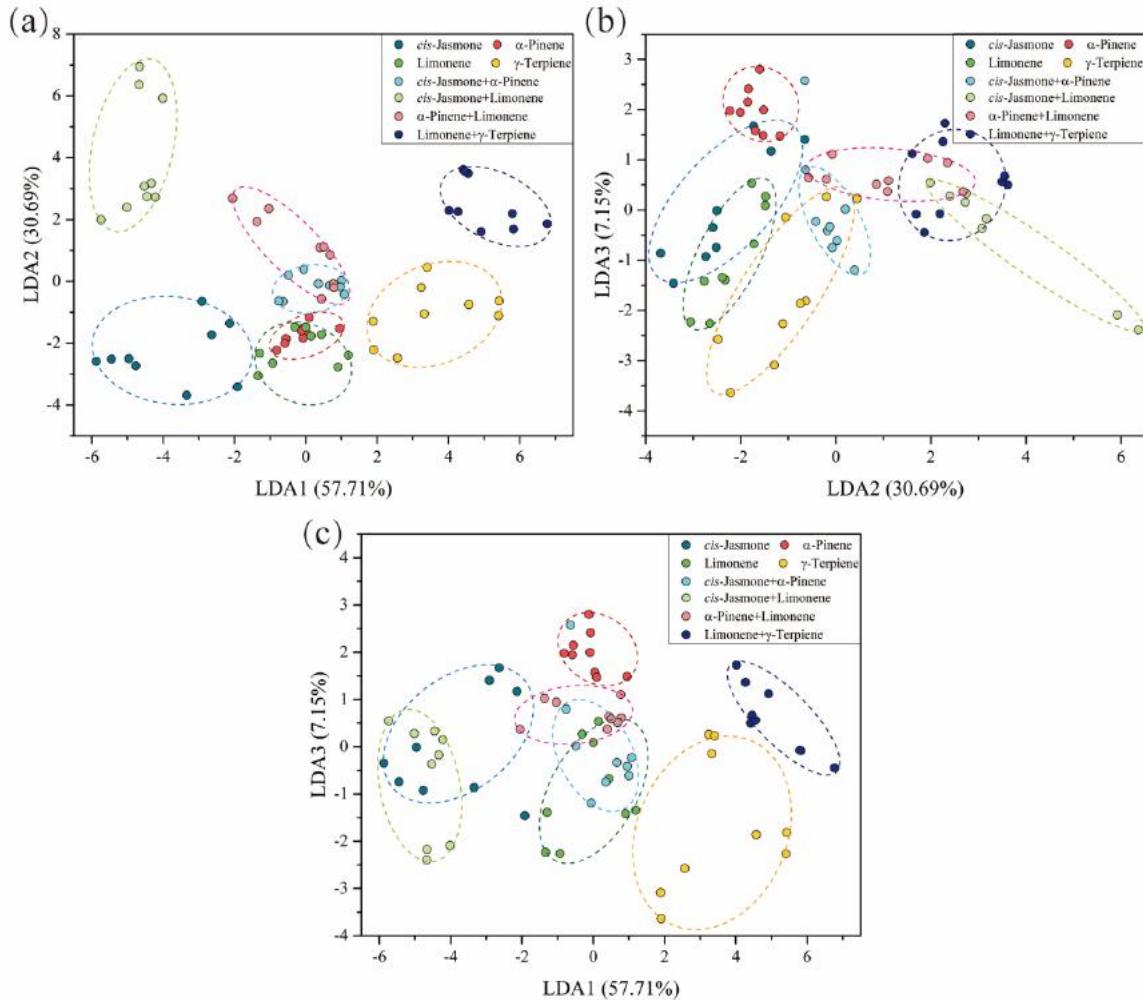


Fig. 6.16 LDA score plot of the first 3 discriminant factors achieved from the response matrix of 72 samples from four PVOCS (CJ,  $\alpha$ -pinene, limonene, and  $\gamma$ -terpinene) and their binary mixtures (CJ+ $\alpha$ -pinene, CJ+limonene,  $\alpha$ -pinene+limonene, and limonene+ $\gamma$ -terpinene).

NBC) were used. KNN has been widely applied as a supervised pattern recognition approach because of its robust nature and suitability for limited sample sets [229]. The main assumption of KNN is that the closer the samples, the more likely it is that they belong in the same category [230, 231]. Considering the sample size, three nearest neighbors were considered in the present study. Similarly, NBC is a supervised statistical model established by calculating the probability that a given sample belongs to a certain class [232]. Because of its simple structure and ease of implementation, NBC has been widely used [233]. More details on KNN and NBC can be found elsewhere [234–237].

In this study, 72 samples were divided into train and test sets by the random selection method [238]. Samples from the train set were used as training models, and samples from

the test set were used to evaluate the established models. The ratio of samples in the train and test sets was 7:3. Because the sample partition was random, the performance of the model was unstable. Therefore, sample partition was repeated 100 times to overcome this problem. The identification performance of a classification framework was evaluated by accuracy using Formula 6.3 [165].

$$\text{Accuracy} = \frac{TPs + TNs}{TPs + TNs + FPs + FNs} \times 100\% \quad (6.3)$$

where, the true positive (TP) is the event that a positive sample is classified as a positive example, the true negative (TN) is the event that a negative sample is classified as a negative example, the false positive (FP) is the event that a negative sample is classified as a positive example, the false negative (FN) is the event that a positive sample is classified as a negative example. According to the average accuracies, the optimal model was chosen.

The PVOC identification accuracies for the train and test sets of the three models are listed in Table 6.1. Other standard performance measures (sensitivity, specificity, precision, recall, and F1 score) are summarized in Table C.1 and C.2. For each PVOC, the LDA model showed the highest average accuracy in identifying CJ ( $99.40 \pm 3.20\%$ ),  $\gamma$ -terpinene ( $99.56 \pm 1.02\%$ ), and  $\alpha$ -pinene+limonene ( $97.49 \pm 3.40\%$ ). The KNN model readily identified  $\alpha$ -pinene ( $89.83 \pm 8.87\%$ ) and limonene ( $95.57 \pm 4.49\%$ ). The NBC model exhibited the highest identification accuracy for mixtures, including CJ+ $\alpha$ -pinene ( $95.30 \pm 6.34\%$ ), CJ+limonene (100%), and limonene+ $\gamma$ -terpinene (100%). This indicates that the NBC model was more suitable to deal with complicated samples than the other models. For the train set, the accuracy of the NBC model reached  $97.02 \pm 2.79\%$ , which was higher than that of the LDA model ( $86.66 \pm 2.49\%$ ) and KNN model ( $95.58 \pm 8.06\%$ ). For the test set, KNN had the highest accuracy ( $97.02 \pm 2.79\%$ ), followed by LDA ( $94.72 \pm 8.52\%$ ) and NBC ( $94.39 \pm 9.15\%$ ), indicating that the generalization ability of the KNN model is higher than that of the others. Considering total average accuracy, KNN had the highest accuracy ( $96.30 \pm 6.03\%$ ), followed by LDA ( $95.69 \pm 6.28\%$ ) and then NBC ( $95.71 \pm 6.77\%$ ). Overall, we found that KNN was the optimal model to identify PVOCs based on accuracies for the train and test sets. We also found that the lowest accuracy achieved by the models was higher than 95%. This indicated that enough molecular information was captured by the AuNPs@MISG LSPR sensor array to allow PVOC identification.

## 6.4 Conclusion

An LSPR sensor coated with an MISG containing AuNPs to amplify the sensing signal was developed for PVOC detection. The optimal size and amount of AuNPs doped in the

Table 6.1 PVOCS identification accuracies of LDA, KNN, and NBC models based on the sensor array response matrix

PVOCS	Modeling approach	Train set (%)		Test set (%)		Total average (%)	
		Mean	SD	Mean	SD	Mean	SD
cis-Jasmone	LDA	99.93	0.63	98.87	4.48	99.4	3.2
	KNN	97.54	6.63	96.98	3.53	97.26	5.31
	NBC	96.98	3.53	91.79	11.3	94.39	8.37
$\alpha$ -Pinene	LDA	89.98	4.2	86.85	13.13	88.42	9.75
	KNN	91.1	11.69	88.55	4.55	89.83	8.87
	NBC	88.55	4.55	82.74	15.86	85.65	11.67
Limonene	LDA	98.78	1.4	93.55	12.81	96.17	9.11
	KNN	97.31	6.29	99.83	0.9	98.57	4.49
	NBC	99.83	0.9	94.54	9.6	97.19	6.82
$\gamma$ -Terpiene	LDA	99.87	0.34	99.24	1.4	99.56	1.02
	KNN	97.6	6.13	99.94	0.63	98.77	4.36
	NBC	99.94	0.63	98.21	4.58	99.08	3.27
cis-Jasmone+ $\alpha$ -Pinene	LDA	88.13	4.32	86.86	11.59	87.5	8.75
	KNN	92.98	11.17	96.48	4.29	94.73	8.46
	NBC	96.48	4.29	94.11	7.87	95.3	6.34
cis-Jasmone+Limonene	LDA	99.55	1.6	97.74	5.65	98.65	4.15
	KNN	98.94	3.79	100	0	99.47	2.68
	NBC	100	0	100	0	100	0
$\alpha$ -Pinene+Limonene	LDA	97.62	2.2	97.35	4.28	97.49	3.4
	KNN	89.91	10.69	94.4	3.11	92.16	7.87
	NBC	94.4	3.11	93.71	10.77	94.06	7.93
Limonene+ $\gamma$ -Terpiene	LDA	99.38	1.88	97.33	6.14	98.36	4.54
	KNN	99.27	2.99	100	0	99.64	2.11
	NBC	100	0	100	0	100	0
Summary	LDA	96.66	2.49	94.72	8.52	95.69	6.28
	KNN	95.58	8.06	97.02	2.79	96.3	6.03
	NBC	97.02	2.79	94.39	9.15	95.71	6.77

MISG were 30 nm and 20  $\mu$ L, respectively. Under optimal conditions, the sensitivity of the AuNPs@MISG-coated sensor was 12.33 times higher than that of the equivalent without AuNPs, which was caused by hot spot enhancement. The real-time responses of the sensor displayed good interference immunity and repeatability. A five-channel AuNPs@MISG LSPR sensor array was designed to detect and identify four PVOCS alone and in binary mixtures. Correlation analysis, PCA, and LDA were used to process the response matrix. The results indicated that it was difficult to distinguish clearly between the groups for the binary mixtures. Three supervised modeling approaches (LDA, KNN, and NBC) were used to establish PVOCS pattern recognition models. KNN displayed high accuracy (96.03%), identifying PVOCS quickly and efficiently. This study demonstrated that an AuNPs@MISG-coated LSPR sensor array combined with a pattern recognition approach can be used for PVOCS detection and identification, which may become a useful technology for agricultural applications.



# **Chapter 7**

## **Conclusion and prospect**

### **7.1 Conclusion**

#### **7.1.1 Structure–odor relationship**

Progress in the molecular biology of olfaction has revealed a close relationship between the structural features of odorants and the response patterns they elicit in the olfactory bulb. Molecular feature-related response patterns, termed odor maps (OMs), may represent information related to basic odor quality. Thus, studying the relationship between OMs and the molecular features of odorants is helpful for better understanding the relationships between odorant structure and odor.

Firstly, we explored the correlation between OMs and the molecular parameters (MPs) of odorants by taking OMs from rat olfactory bulbs and extracting feature profiles of the corresponding odorant molecules. Coefficient maps for molecular parameters turned out to be clustered in seven groups, and the parameters in each group had a similar effect on the images of olfactory responses. In addition, we tested how well different models could identify functional groups when the classification models were established based on olfactory information or molecular parameters. The results showed that although classifiers with molecular parameters as feature quantity were weaker than those with odor map as feature quantity, a comparative model could be established if it was based on enough molecular features. This research would be applied in developing biology olfaction based odor recognition sensors.

Secondly, we present a proof-of-concept model by which odor information can be obtained by machine-learning-based prediction from MPs of odorant molecules. The results showed that Boruta (Confirmed only) SVM has good potential in predicting odor perceptions rapidly and precisely. It demonstrated that MPs associated with machine-learning models can be adopted for odor perceptual senses identification.

### 7.1.2 Molecular imprinted material coated optical odor sensor

To develop olfaction inspired odor sensor, we presented a possibility to use localized surface plasmon resonance (LSPR) of Au nanoparticles (AuNPs) to recognize odorants. Molecularly imprinted sol-gels (MISGs) as employed as a sensitive layer for enhancing the selectivity of LSPR sensors.

At the beginning, MISGs coated LSPR sensor was prepared for determination of organic acids vapors selectively. Molecules with similar structure to imprinted molecules (with carboxyl group, different carbon chain length) were selected for MISGs selectivity evaluation. We found that the adsorption capacity of pure TiO<sub>2</sub> sol-gel matrix was weak. In-situ response of LSPR sensor was verified to be fast, selective and reversible. Eventually, a multichannel sensor array was constructed for the determination of 4 organic acids vapors in single and their binary mixtures. Based on principal component analysis (PCA) results, molecular structure information (such as molecular size and carbon chain length) can be detected and captured by MISG layer. A 100% classification rate was achieved by leave-one-out cross-validation technique for the linear discriminant analysis (LDA) model, which indicated that a sensor array combined MISG with LSPR could be an effective method for organic acid odor pattern recognition.

Next, a sensitive and selective nanocomposite-imprinted, LSPR sensor for cis-jasmone vapor was fabricated. The functional monomer and the ratio of matrix materials to functional monomers in the MISGs were investigated and optimized. Result indicated that MISGs contained the functional monomer trimethoxyphenylsilane at a 3:1(v:v) ratio exhibited a higher sensitivity and selectivity than other films. The optical sensor would have advantages of low cost, selectivity, sensitivity, and repeatability. Furthermore, AuNPs were doped in MISGs for enhancing response intensity by hot spots generation. Size and amount of the AuNPs were investigated and optimized. It indicated that MISGs contained the 30 nm AuNPs at 20  $\mu$ L exhibited a higher sensitivity than the other films. Finally, an optical multi-channel sensor platform was developed to detect PVOCS in single and binary mixtures. The result indicated that k-nearest neighbor model had good potential in identifying PVOCS quickly and efficiently. Thus, the sensor platform is expected to be a potential tool for PVOCS monitoring in agriculture applications. In additionally, the result indicated that by changing template molecules and functional monomers, the absorption ability of MISGs can be controlled and designed. The interaction between VOC molecules and MISGs layer was contributed by electron, Van der Waals and  $\pi$ - $\pi$  effects. It indicated that functional monomers contained different types functional monomers (such as amino group, carboxyl group, hydroxyl or benzene ring etc.) would be attempted for developing olfaction inspired odor sensing system.

Therefore, MISG was expected to be a talent molecularly recognized materials for developing olfaction inspired sensing system.

## 7.2 Prospect

### 7.2.1 Basic odor clusters and molecular parameters

The odor clustering in olfactory bulb is not well understood due to the limited amount of reported odorant molecules. To discover all the primary clusters may be a massive undertaking for clarify the mechanism of olfactory perception. The more data that can be acquired, the higher the model accuracy will be. In this study, only 178 odorants with non-intersecting labels were considered. Therefore, the model is applicable for limited types of chemicals. The current model cannot predict the functional groups for molecules with high molecular weight and complicated structures. In the future, more odor-response images of neuronal activity in the OB will be investigated to determine the hidden patterns in t-SNE space. Additionally, a larger variety of molecular parameters will be considered so that the possibility of describing an odorant by molecular information closer to the mammal olfaction.

### 7.2.2 Possibility of machine-learning-based GC-O

Chapter 3 reported a proof-of-concept study aimed at testing the feasibility of ML-based GC-O. In this study, 10 odor descriptors (OD) were tested due to their relatively larger sample size and higher occurrence frequency in the selected database. Although 10 descriptors are obviously not enough for a practical application, the results of this study demonstrate the possibility that a machine learning approach can be used to obtain sensory information of GC effluents. Additionally, more models for ODs would be expected if sufficient samples can be acquired. About 3000 odorants with odor types were reported in existing databases include Flavornet, GoodScents and SuperScent et al.

As future work, in order to extend the prediction models to more ODs, more odorants with OD information would be collected and a summarized odor database would be established. In fact, the number of prediction models actually needed in a GC-O system may be not as many as expected. A recent study suggested that the dimensionality of odor percepts may be around 20 or less although our nose has 400 olfactory receptors [239]. This may mean 20 or less descriptors are enough for their application in GC-O. As shown in Table 3.3, multi ODs can be predicted for one compound. These ODs can be used as reference for panelists to obtain a relatively credible odor evaluation, which can enhance their work efficiency and accuracy. In future, an enough reliable model system would be established

rely on adequate samples to instead of panelists for odor type evaluation. In conclusion, the prediction model combined with descriptor indexing may be a good candidate to replace the human panelist in GC-O.

### 7.2.3 Olfaction inspired optical odor sensing system

Although a higher identification sensing platform was developed, information captured by the sensor array is still not enough to deal with complicated natural environment. For further improvement in sensitivity of LSPR sensors, AuNPs heterostructure should be considered and attempted by hot spot effect. In addition, by selecting template molecules and functional monomers in sol-gel, MISGs affined to odorants with characteristic structures or functional groups would be designed. Furthermore, we believe that by creating more channels for sensor array (coated by different polymer or sol-gel), complicated mixture vapors can be recognized in the future. As illustrated in Figure 7.1, based on inkjet technology, more types of MISGs can be doped on Au nano-islands, respectively. By a hyperspectral camera, a time series image data can be captured and analyzed. Finally, OMs can be employed as a supervised data for creating artificial OMs based on the responses of multi-MISG sensing platform. Based on these technologies, an olfaction inspired optical odor sensor would be developed.

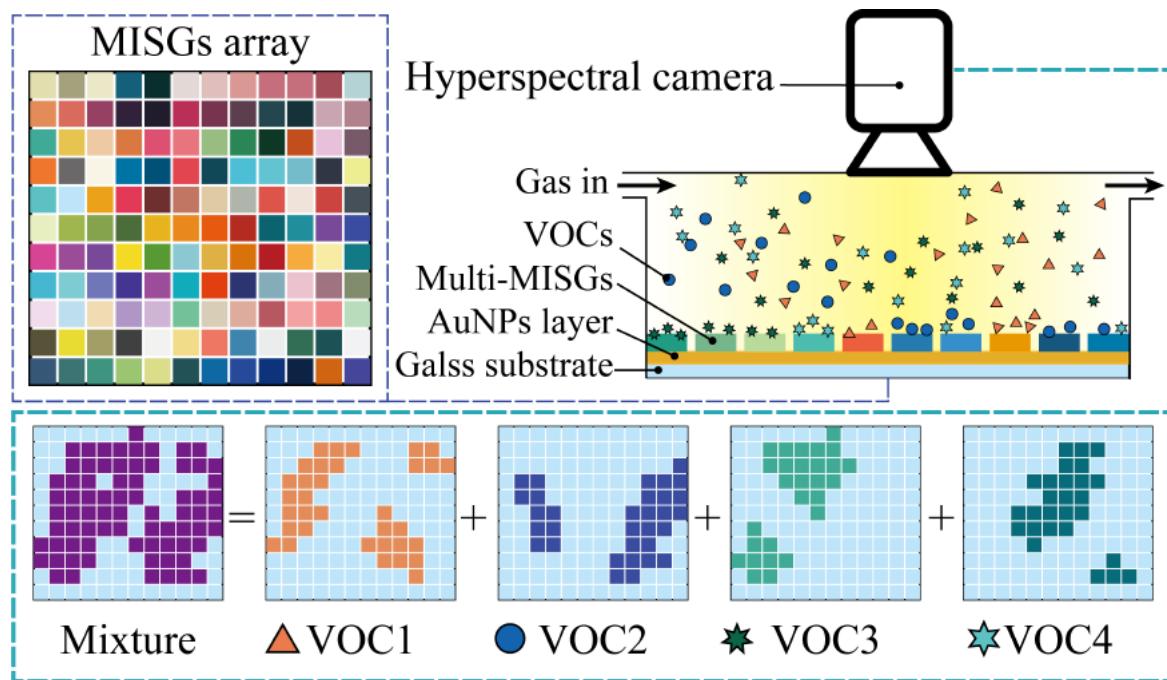


Fig. 7.1 Schematic of hyperspectral camera based multi-MISG-LSPR optical sensing platform.

# References

- [1] C. A. Levitan, J. Ren, A. T. Woods, S. Boesveldt, J. S. Chan, K. J. McKenzie, M. Dodson, J. A. Levin, C. X. R. Leong and J. J. F. v. d. Bosch, *PLoS One*, 2014, **9**, e101651.
- [2] M. Meister and T. Bonhoeffer, *The Journal of Neuroscience*, 2001, **21**, 1351–1360.
- [3] D. Saha, C. Li, S. Peterson, W. Padovano, N. Katta and B. Raman, *Nat. Commun.*, 2015, **6**, 6953.
- [4] C. Bushdid, M. O. Magnasco, L. B. Vosshall and A. Keller, *Science*, 2014, **343**, 1370–1372.
- [5] S. X. Luo, R. Axel and L. F. Abbott, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 10713–10718.
- [6] L. B. Buck, *Annu. Rev. Neurosci.*, 1996, **19**, 517–544.
- [7] R. Pellegrino, P. G. Crandall and H. S. Seo, *Sci. Rep.*, 2016, **6**, year.
- [8] R. Vassar, J. Ngai and R. Axel, *Cell*, 1993, **74**, 309–318.
- [9] M. Stopfer, S. Bhagavan, B. H. Smith and G. Laurent, *Nature*, 1997, **390**, 70–74.
- [10] J. A. Gottfried, J. O’Doherty and R. J. Dolan, *Science*, 2003, **301**, 1104–1107.
- [11] B. D. Rubin and L. C. Katz, *Neuron*, 1999, **23**, 499–511.
- [12] M. Wachowiak and L. B. Cohen, *Neuron*, 2001, **32**, 723–735.
- [13] M. Brattoli, G. De Gennaro, V. De Pinto, A. Demarinis Loiotile, S. Lovascio and M. Penza, *Sensors*, 2011, **11**, 5290–5322.
- [14] E. R. Soucy, D. F. Albeanu, A. L. Fantana, V. N. Murthy and M. Meister, *Nat. Neurosci.*, 2009, **12**, 210–220.
- [15] P. Mombaerts, F. Wang, C. Dulac, S. K. Chao, A. Nemes, M. Mendelsohn, J. Edmondson and R. Axel, *Cell*, 1996, **87**, 675–686.
- [16] N. Uchida, Y. K. Takahashi, M. Tanifuji and K. Mori, *Nat. Neurosci.*, 2000, **3**, 1035–1043.
- [17] K. Snitz, A. Yablonka, T. Weiss, I. Frumin, R. M. Khan and N. Sobel, *PLoS Comput. Biol.*, 2013, **9**, e1003184.

- [18] B. Malnic, J. Hirono, T. Sato and L. B. Buck, *Cell*, 1999, **96**, 713–723.
- [19] B. A. Johnson, H. Farahbod and M. Leon, *The Journal of Comparative Neurology*, 2005, **483**, 205–216.
- [20] B. A. Johnson, H. Farahbod, S. Saber and M. Leon, *The Journal of Comparative Neurology*, 2005, **483**, 192–204.
- [21] M. Imahashi and K. Hayashi, *Sensors and Materials*, 2014, **26**, 171–180.
- [22] B. A. Johnson, Z. Xu, S. S. Ali and M. Leon, *The Journal of Comparative Neurology*, 2009, **514**, 658–673.
- [23] K. Mori, H. Nagao and Y. Yoshihara, *Science*, 1999, **286**, 711–715.
- [24] L. Secundo, K. Snitz, K. Weissler, L. Pinchover, Y. Shoenfeld, R. Loewenthal, N. Agmon-Levin, I. Frumin, D. Bar-Zvi, S. Shushan and N. Sobel, *Proc. Natl. Acad. Sci. U.S.A.*, 2015, **112**, 8750–8755.
- [25] B. Raman and R. Gutierrez-Osuna, in *Relating Sensor Responses of Odorants to Their Organoleptic Properties by Means of a Biologically-Inspired Model of Receptor Neuron Convergence onto Olfactory Bulb*, ed. A. Gutiérrez and S. Marco, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 93–108.
- [26] A. Arzi and N. Sobel, *Trends in Cognitive Sciences*, 2011, **15**, 537–545.
- [27] R. Haddad, R. Khan, Y. K. Takahashi, K. Mori, D. Harel and N. Sobel, *Nat. Methods*, 2008, **5**, 425–9.
- [28] R. Kumar, R. Kaur, B. Auffarth and A. P. Bhondekar, *PLoS One*, 2015, **10**, e0141263.
- [29] A. Keller, R. C. Gerkin, Y. F. Guan, A. Dhurandhar, G. Turu, B. Szalai, J. D. Mainland, Y. Ihara, C. W. Yu, R. Wolfinger, C. Vens, L. Schietgat, K. De Grave, R. Norel, G. Stolovitzky, G. A. Cecchi, L. B. Vosshall, P. Meyer and D. O. Prediction, *Science*, 2017, **355**, 820–826.
- [30] Y. Nozaki and T. Nakamoto, *PLoS One*, 2016, **11**, e0157030.
- [31] B. Auffarth, *Neurosci. Biobehav. Rev.*, 2013, **37**, 1667–1679.
- [32] K. Kaepller and F. Mueller, *Chem. Senses*, 2013, **38**, 189–209.
- [33] M. Falasconi, A. Gutierrez-Galvez, M. Leon, B. A. Johnson and S. Marco, *Chem. Senses*, 2012, **37**, 639–653.
- [34] S. K. Jha and K. Hayashi, *Sens. Actuators, B*, 2015, **206**, 471–487.
- [35] H. Ishida, Y. Wada and H. Matsukura, *IEEE Sensors J.*, 2012, **12**, 3163–3173.
- [36] M. Delgado-Rodriguez, M. Ruiz-Montoya, I. Giraldez, R. Lopez, E. Madejon and M. J. Diaz, *Atmos. Environ.*, 2012, **51**, 278–285.

- [37] D. Liu, M. Y. Liu, G. H. Liu, S. C. Zhang, Y. Y. Wu and X. R. Zhang, *Anal. Chem.*, 2010, **82**, 66–68.
- [38] J. Bartsch, E. Uhde and T. Salthammer, *Anal. Chim. Acta*, 2016, **904**, 98–106.
- [39] M. Brattoli, E. Cisternino, P. R. Dambruoso, G. de Gennaro, P. Giungato, A. Mazzone, J. Palmisani and M. Tutino, *Sensors*, 2013, **13**, 16759–16800.
- [40] B. D. Zellner, P. Dugo, G. Dugo and L. Mondello, *J. Chromatogr. A*, 2008, **1186**, 123–143.
- [41] A. Casilli, E. Decorzant, A. Jaquier and E. Delort, *J. Chromatogr. A*, 2014, **1373**, 169–178.
- [42] E. S. Erten and K. R. Cadwallader, *Food Chem.*, 2017, **217**, 244–253.
- [43] T. E. Acree, *Anal. Chem.*, 1997, **69**, A170–A175.
- [44] S. T. Chin, G. T. Eyres and P. J. Marriott, *Anal. Chem.*, 2012, **84**, 9154–9162.
- [45] C. M. Delahunty, G. Eyres and J. P. Dufour, *J. Sep. Sci.*, 2006, **29**, 2107–2125.
- [46] H. Guth, *J. Agric. Food. Chem.*, 1997, **45**, 3022–3026.
- [47] P. Pollien, L. B. Fay, M. Baumgartner and A. Chaintreau, *Anal. Chem.*, 1999, **71**, 5391–5397.
- [48] S. M. van Ruth, *Biomol. Eng.*, 2001, **17**, 121–128.
- [49] F. Y. Yi, D. X. Chen, M. K. Wu, L. Han and H. L. Jiang, *Chempluschem*, 2016, **81**, 675–690.
- [50] L. X. Chen, X. Y. Wang, W. H. Lu, X. Q. Wu and J. H. Li, *Chem. Soc. Rev.*, 2016, **45**, 2137–2211.
- [51] G. Lal and D. C. Tiwari, *Sens. Actuators, B*, 2018, **262**, 64–69.
- [52] H. Bai and G. Q. Shi, *Sensors*, 2007, **7**, 267–307.
- [53] R. Paolesse, S. Nardis, D. Monti, M. Stefanelli and C. Di Natale, *Chem. Rev.*, 2017, **117**, 2517–2583.
- [54] Y. P. Chen, Y. L. Xianyu and X. Y. Jiang, *Acc. Chem. Res.*, 2017, **50**, 310–319.
- [55] G. Korotcenkov, *Materials Science and Engineering B-Solid State Materials for Advanced Technology*, 2007, **139**, 1–23.
- [56] C. M. Widdifield and R. W. Schurko, *Concepts in Magnetic Resonance Part A*, 2009, **34A**, 91–123.
- [57] C. O. Chui, F. Ito and K. C. Saraswat, *IEEE Trans. Electron Devices*, 2006, **53**, 1501–1508.

- [58] A. Mirzaei, K. Janghorban, B. Hashemi and G. Neri, *J. Nanopart. Res.*, 2015, **17**, year.
- [59] M. T. Soo, K. Y. Cheong and A. F. M. Noor, *Sens. Actuators, B*, 2010, **151**, 39–55.
- [60] X. H. Liu, T. T. Ma, Y. S. Xu, L. Sun, L. L. Zheng, O. G. Schmidt and J. Zhang, *Sens. Actuators, B*, 2018, **264**, 92–99.
- [61] H. I. Chen, C. H. Chang, H. H. Lu, I. P. Liu, W. C. Chen, B. Y. Ke and W. C. Liu, *Sens. Actuators, B*, 2018, **262**, 852–859.
- [62] Y. W. Li, N. Luo, G. Sun, B. Zhang, L. Lin, H. H. Jin, Y. Wang, H. Bala, J. L. Cao and Z. Y. Zhang, *Ceram. Int.*, 2018, **44**, 6836–6842.
- [63] F. Rahimi and A. I. Zad, *Journal of Physics D-Applied Physics*, 2007, **40**, 7201–7209.
- [64] S. Sharma and M. Madou, *Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences*, 2012, **370**, 2448–2473.
- [65] A. Kaushik, R. Kumar, S. K. Arya, M. Nair, B. D. Malhotra and S. Bhansali, *Chem. Rev.*, 2015, **115**, 4571–4606.
- [66] M. Setka, J. Drbohlavova and J. Hubalek, *Sensors*, 2017, **17**, year.
- [67] J. H. Xu, Y. Z. Wang and S. S. Hu, *Microchim. Acta*, 2017, **184**, 1–44.
- [68] C. O. Baker, X. W. Huang, W. Nelson and R. B. Kaner, *Chem. Soc. Rev.*, 2017, **46**, 1510–1525.
- [69] T. F. Otero and S. Beaumont, *Sens. Actuators, B*, 2018, **263**, 493–501.
- [70] P. G. Mineo, A. Abbadessa, A. Rescifina, A. Mazzaglia, A. Nicosia and A. A. Scamporrino, *Colloids and Surfaces a-Physicochemical and Engineering Aspects*, 2018, **546**, 40–47.
- [71] M. Eslamian, *Nano-Micro Letters*, 2017, **9**, year.
- [72] S. J. Kim and D. H. Shin, in *The Multisensory Effects of Atmospheric Cues on Online Shopping Satisfaction*, ed. F. F. H. Nah and C. H. Tan, 2016, vol. 9751, pp. 406–416.
- [73] H. Y. Jia, J. Wang, X. Y. Zhang and Y. P. Wang, *ACS Macro Lett.*, 2014, **3**, 86–90.
- [74] R. Khalil, S. Homaeigohar, D. Haussler and M. Elbahri, *Sci. Rep.*, 2016, **6**, year.
- [75] A. Ghoorchian and N. Alizadeh, *Sens. Actuators, B*, 2018, **255**, 826–835.
- [76] A. Dominguez-Aragon, J. A. Olmedo-Martinez and E. A. Zaragoza-Contreras, *Sens. Actuators, B*, 2018, **259**, 170–176.
- [77] M. M. Gong, Y. S. Li, Y. A. Guo, X. Lv and X. C. Dou, *Sens. Actuators, B*, 2018, **262**, 350–358.
- [78] M. R. Li, W. Wang, Z. Chen, Z. L. Song and X. L. Luo, *Sens. Actuators, B*, 2018, **260**, 778–785.

- [79] B. Mettaï, A. Mekki, F. Merdj, Z. B. D. Sayah, K. M. Soumia, Z. Safiddine, R. Mahmoud and M. M. Chehimi, *J. Polym. Res.*, 2018, **25**, year.
- [80] A. J. Slavin, *Journal of Vacuum Science & Technology A*, 2018, **36**, year.
- [81] D. Z. Zhang, D. Y. Wang, P. Li, X. Y. Zhou, X. Q. Zong and G. K. Dong, *Sens. Actuators, B*, 2018, **255**, 1869–1877.
- [82] C. J. Liu, L. Shang, H. T. Yoshioka, B. Chen and K. Hayashi, *Anal. Chim. Acta*, 2018, **1010**, 1–10.
- [83] A. G. Ayankojo, J. Reut, R. Boroznjak, A. Opik and V. Syritski, *Sens. Actuators, B*, 2018, **258**, 766–774.
- [84] X. L. Cha, F. F. Yu, Y. Fan, J. F. Chen, L. Y. Wang, Q. Xiang, Z. M. Duan and J. Q. Xu, *Sens. Actuators, B*, 2018, **263**, 436–444.
- [85] Y. Zhu, Z. X. Cheng, Q. Xiang, Y. H. Zhu and J. Q. Xu, *Sens. Actuators, B*, 2018, **256**, 888–895.
- [86] M. Peris and L. Escuder-Gilabert, *Anal. Chim. Acta*, 2009, **638**, 1–15.
- [87] L. Huang, J. W. Zhao, Q. S. Chen and Y. H. Zhang, *Food Chem.*, 2014, **145**, 228–236.
- [88] L. Zhang and D. Zhang, *IEEE Trans. Instrum. Meas.*, 2015, **64**, 1790–1801.
- [89] A. Loutfi, S. Coradeschi, G. K. Mani, P. Shankar and J. B. B. Rayappan, *J. Food Eng.*, 2015, **144**, 103–111.
- [90] A. D. Wilson and M. Baietto, *Sensors*, 2009, **9**, 5099–5148.
- [91] L. Dutta, C. Talukdar, A. Hazarika and M. Bhuyan, *IEEE Trans. Ind. Electron.*, 2018, **65**, 4983–4990.
- [92] Q. Q. Chen, J. X. Song, J. F. Bi, X. J. Meng and X. Y. Wu, *Food Res. Int.*, 2018, **105**, 605–615.
- [93] B. Chen, C. J. Liu, M. Ota and K. Hayashi, *IEEE Sensors J.*, 2013, **13**, 1307–1314.
- [94] H. Nanto, F. Yagi, H. Hasunuma, Y. Takei, S. Koyama, T. Oyabu and T. Mihara, *Sensors and Materials*, 2009, **21**, 201–208.
- [95] K. M. Mayer and J. H. Hafner, *Chem. Rev.*, 2011, **111**, 3828–3857.
- [96] J. B. Hu, Y. Yu, B. Jiao, S. Y. Ning, H. Dong, X. Hou, Z. J. Zhang and Z. X. Wu, *Org. Electron.*, 2016, **31**, 234–239.
- [97] M. J. Hou, Y. Huang, L. W. Ma and Z. J. Zhang, *Nanoscale Res. Lett.*, 2015, **10**, year.
- [98] D. S. Boyuk, L. W. Chou and M. A. Filler, *ACS Photonics*, 2016, **3**, 184–189.
- [99] C. C. Evans, C. Y. Liu and J. Suntivich, *ACS Photonics*, 2016, **3**, 1662–1669.

- [100] C. H. Fang, G. L. Zhao, Y. L. Xiao, J. Zhao, Z. J. Zhang and B. Y. Geng, *Sci. Rep.*, 2016, **6**, year.
- [101] S.-W. Kim, J.-S. Lee, S.-W. Lee, B.-H. Kang, J.-B. Kwon, O.-S. Kim, J.-S. Kim, E.-S. Kim, D.-H. Kwon and S.-W. Kang, *Sensors*, 2017, **17**, year.
- [102] B. Sepulveda, P. C. Angelome, L. M. Lechuga and L. M. Liz-Marzan, *Nano Today*, 2009, **4**, 244–251.
- [103] K. Lodewijks, J. Ryken, W. Van Roy, G. Borghs, L. Lagae and P. Van Dorpe, *Plasmonics*, 2013, **8**, 1379–1385.
- [104] M. S. Kirschner, C. M. Lethiec, X. M. Lin, G. C. Schatz, L. X. Chen and R. D. Schaller, *ACS Photonics*, 2016, **3**, 758–763.
- [105] K. Q. Lin, J. Yi, S. Hu, J. J. Sun, J. T. Zheng, X. Wang and B. Ren, *ACS Photonics*, 2016, **3**, 1248–1255.
- [106] J. R. L. Guerreiro, N. Teixeira, V. De Freitas, M. G. F. Sales and D. S. Sutherland, *Food Chem.*, 2017, **233**, 457–466.
- [107] S. Kamimura, S. Yamashita, S. Abe, T. Tsubota and T. Ohno, *Appl. Catal., B*, 2017, **211**, 11–17.
- [108] A. V. Skinner, S. Han and R. Balasubramanian, *Sens. Actuators, B*, 2017, **247**, 706–712.
- [109] A. Thakur, G. Y. Qiu, N. G. Siu-Pang, J. T. Guan, J. B. Yue, Y. J. Lee and C. M. L. Wu, *Biosensors & Bioelectronics*, 2017, **94**, 400–407.
- [110] Y. C. Wang, L. Lu and S. Gunasekaran, *Biosensors & Bioelectronics*, 2017, **92**, 109–116.
- [111] K. Shrivastava, N. Nirmalkar, S. S. Thakur, M. K. Deb, S. S. Shinde and R. Shankar, *Food Chem.*, 2018, **250**, 14–21.
- [112] T. R. Lin, M. Q. Zhang, F. H. Xu, X. Y. Wang, Z. F. Xu and L. Q. Guo, *Sens. Actuators, B*, 2018, **261**, 379–384.
- [113] X. M. Li, H. B. Feng, Y. Y. Wang, C. P. Zhou, W. Jiang, M. Zhong and J. H. Zhou, *Sens. Actuators, B*, 2018, **262**, 411–417.
- [114] M. T. Alula, L. Karamchand, N. R. Hendricks and J. M. Blackburn, *Anal. Chim. Acta*, 2018, **1007**, 40–49.
- [115] B. Chen, C. J. Liu, L. P. Ge and K. Hayashi, *Sens. Actuators, B*, 2016, **231**, 787–792.
- [116] J. M. Lim, N. S. Heo, S. Y. Oh, M. Y. Ryu, J. H. Seo, T. J. Park, Y. S. Huh and J. P. Park, *Biosensors & Bioelectronics*, 2018, **99**, 289–295.
- [117] A. Urrutia, J. Goicoechea, P. J. Rivero, A. Pildain and F. J. Arregui, *Sens. Actuators, B*, 2018, **255**, 2105–2112.

- [118] M. D. Allendorf, R. J. T. Houk, L. Andruszkiewicz, A. A. Talin, J. Pikarsky, A. Choudhury, K. A. Gall and P. J. Hesketh, *J. Am. Chem. Soc.*, 2008, **130**, 14404–+.
- [119] V. T. Tran, H. J. Zhou, S. Kim, J. Lee, J. Kim, F. M. Zou, J. Kim, J. Y. Park and J. Lee, *Sens. Actuators, B*, 2014, **203**, 817–823.
- [120] H. N. Iqbal and P. A. Lieberzeit, *Sens. Actuators, B*, 2013, **176**, 1090–1095.
- [121] K. Kotova, M. Hussain, G. Mustafa and P. A. Lieberzeit, *Sens. Actuators, B*, 2013, **189**, 199–202.
- [122] U. Latif, A. Rohrer, P. A. Lieberzeit and F. L. Dickert, *Anal. Bioanal. Chem.*, 2011, **400**, 2457–2462.
- [123] M. Matsuguchi and T. Uno, *Sens. Actuators, B*, 2006, **113**, 94–99.
- [124] Z. P. Yang and C. J. Zhang, *Sens. Actuators, B*, 2009, **142**, 210–215.
- [125] D. Z. Zhou, T. Y. Guo, Y. Yang and Z. P. Zhang, *Sens. Actuators, B*, 2011, **153**, 96–102.
- [126] R. Takemura, H. Sakata and H. Ishida, *Journal of Sensors*, 2016, **2016**, 11.
- [127] F. Yoshino and T. Nakamoto, *Sensors and Materials*, 2014, **26**, 137–147.
- [128] B. Babamiri, A. Salimi, R. Hallaj and M. Hasanzadeh, *Biosensors & Bioelectronics*, 2018, **107**, 272–279.
- [129] L. J. Zhang, G. H. Wang, C. Xiong, L. Zheng, J. B. He, Y. S. Ding, H. B. Lu, G. B. Zhang, K. Cho and L. Z. Qiu, *Biosensors & Bioelectronics*, 2018, **105**, 121–128.
- [130] W. L. Liu, H. F. Li, S. M. Yu, J. X. Zhang, W. H. Zheng, L. T. Niu and G. G. Li, *Biosensors & Bioelectronics*, 2018, **104**, 79–86.
- [131] J. Huang, Y. N. Wu, J. J. Cong, J. Luo and X. Y. Liu, *Sens. Actuators, B*, 2018, **259**, 1–9.
- [132] B. Chen, C. J. Liu and K. Hayashi, *IEEE Sensors J.*, 2014, **14**, 3458–3464.
- [133] M. Arabi, A. Ostovan, M. Ghaedi and M. K. Purkait, *Talanta*, 2016, **154**, 526–532.
- [134] Y. Zeng, Y. Zhou, L. Kong, T. Zhou and G. Shi, *Biosensors & Bioelectronics*, 2013, **45**, 25–33.
- [135] A. Mujahid, P. A. Lieberzeit and F. L. Dickert, *MATERIALS*, 2010, **3**, 2196–2217.
- [136] J. E. Lofgreen and G. A. Ozin, *CHEMICAL SOCIETY REVIEWS*, 2014, **43**, 911–933.
- [137] C. Hu, J. Deng, Y. Zhao, L. Xia, K. Huang, S. Ju and N. Xiao, *Food Chem.*, 2014, **158**, 366–373.
- [138] E. L. Holthoff and F. V. Bright, *Acc. Chem. Res.*, 2007, **40**, 756–767.

- [139] Y. J. Tong, H. D. Li, H. M. Guan, J. M. Zhao, S. Majeed, S. Anjum, F. Liang and G. B. Xu, *Biosensors & Bioelectronics*, 2013, **47**, 553–558.
- [140] A. Ostovan, M. Ghaedi and M. Arabi, *Talanta*, 2018, **179**, 760–768.
- [141] E. Bou-Maroun, J. Rossignol, B. De Fonseca, C. Lafarge, R. D. Gougeon, D. Stuerga and P. Cayot, *Sens. Actuators, B*, 2017, **244**, 24–30.
- [142] L. Shang, C. Liu, Y. Tomiura and K. Hayashi, *Sens. Actuators, B*, 2018, **255**, 508–518.
- [143] L. Shang, C. Liu, Y. Tomiura and K. Hayashi, *Analytical Chemistry*, 2017, **89**, 11999–12005.
- [144] L. Shang, C. J. Liu, M. Watanabe, B. Chen and K. Hayashi, *Sens. Actuators, B*, 2017, **249**, 14–21.
- [145] L. Shang, C. Liu, B. Chen and K. Hayashi, *Sens. Actuators, B*, 2018, **260**, 617–626.
- [146] W. Guo, J. Gu, D. Liu and L. Shang, *Comput. Electron. Agric.*, 2016, **123**, 297–303.
- [147] X. Zhu, L. Fang, J. Gu and W. Guo, *Food Anal. Methods*, 2016, **9**, 1789–1798.
- [148] L. Van der Maaten, *Journal of Machine Learning Research*, 2014, **15**, 3221–3245.
- [149] J. Cheng, H. J. Liu, F. Wang, H. S. Li and C. Zhu, *IEEE Trans. Image Process.*, 2015, **24**, year.
- [150] X. H. Li, W. Kong, W. M. Shi and Q. Shen, *Chemom. Intell. Lab. Syst.*, 2016, **155**, 145–150.
- [151] H. Kaneko and K. Funatsu, *Chemom. Intell. Lab. Syst.*, 2016, **153**, 75–81.
- [152] W. Gani and M. Limam, *J. Stat. Comput. Simul.*, 2016, **86**, 135–148.
- [153] G. B. da Costa, D. D. S. Fernandes, A. A. Gomes, V. E. de Almeida and G. Veras, *Food Chem.*, 2016, **196**, 539–543.
- [154] T. H. Sun, F. C. Tien, F. C. Tien and R. J. Kuo, *Journal of Intelligent Manufacturing*, 2016, **27**, 639–651.
- [155] T. Liu, C. S. Chen, X. Z. Shi and C. Y. Liu, *Laser Phys.*, 2016, **26**, year.
- [156] A. Bohnsack, K. Domaschke, M. Kaden, M. Lange and T. Villmann, *Neurocomputing*, 2016, **192**, 104–114.
- [157] J. Cho, R. Anandakathir, A. Kumar, J. Kumar and P. U. Kurup, *Sens. Actuators, B*, 2011, **160**, 1237–1243.
- [158] R. Kumar, A. P. Bhondekar, R. Kaur, S. Vig, A. Sharma and P. Kapur, *Sens. Actuators, B*, 2012, **171**, 1046–1053.
- [159] W. C. Guo, L. Shang, X. H. Zhu and S. O. Nelson, *Food Bioprocess Technol.*, 2015, **8**, 1126–1138.

- [160] L. Zhang, F. C. Tian, H. Nie, L. J. Dang, G. R. Li, Q. Ye and C. Kadri, *Sens. Actuators, B*, 2012, **174**, 114–125.
- [161] C. C. Chang and C. J. Lin, *ACM Trans. Intell. Syst. Technol.*, 2011, **2**, year.
- [162] G. B. Huang, H. M. Zhou, X. J. Ding and R. Zhang, *IEEE Transactions on Systems Man and Cybernetics Part B-Cybernetics*, 2012, **42**, 513–529.
- [163] G. B. Huang, D. H. Wang and Y. Lan, *International Journal of Machine Learning and Cybernetics*, 2011, **2**, 107–122.
- [164] L. Cornejo-Bueno, J. C. Nieto-Borge, P. Garcia-Diaz, G. Rodriguez and S. Salcedo-Sanz, *Renewable Energy*, 2016, **97**, 380–389.
- [165] L. Shang, W. C. Guo and S. O. Nelson, *Food Anal. Methods*, 2015, **8**, 1042–1052.
- [166] Y. K. Takahashi, M. Kurosaki, S. Hirono and K. Mori, *J. Neurophysiol.*, 2004, **92**, 2413–2427.
- [167] J. A. Gottfried, J. S. Winston and R. J. Dolan, *Neuron*, 2006, **49**, 467–479.
- [168] B. N. Johnson, J. D. Mainland and N. Sobel, *J. Neurophysiol.*, 2003, **90**, 1084–1094.
- [169] P. Comon, *Signal Process.*, 1994, **36**, 287–314.
- [170] L. Turgeman, J. H. May and R. Sciulli, *Expert Syst. Appl.*, 2017, **78**, 376–385.
- [171] S. Suresh, R. V. Babu and H. J. Kim, *Appl. Soft Comput.*, 2009, **9**, 541–552.
- [172] M. Xia, Y. C. Zhang, L. G. Weng and X. L. Ye, *Knowledge-Based Systems*, 2012, **36**, 253–259.
- [173] S. J. Lin, C. H. Chang and M. F. Hsu, *Knowledge-Based Systems*, 2013, **39**, 214–223.
- [174] F. Kermen, A. Chakirian, C. Sezille, P. Joussain, G. Le Goff, A. Ziessel, M. Chastrette, N. Mandairon, A. Didier, C. Rouby and M. Bensafi, *Sci. Rep.*, 2011, **1**, year.
- [175] D. Zakarya, D. Cherqaoui, M. Esseffar, D. Villemain and J. M. Cense, *J. Phys. Org. Chem.*, 1997, **10**, 612–622.
- [176] J. Y. Li, S. M. Fong, Y. S. Sung, K. G. Cho, R. Wong and K. K. L. Wong, *BioData Mining*, 2016, **9**, year.
- [177] N. H. Agjee, R. Ismail and O. Mutanga, *J. Appl. Remote Sens.*, 2016, **10**, year.
- [178] N. K. Poona, A. van Niekerk, R. L. Nadel and R. Ismail, *Appl. Spectrosc.*, 2016, **70**, 322–333.
- [179] J. C. Ren, *Knowledge-Based Systems*, 2012, **26**, 144–153.
- [180] G. Arvisenet, L. Billy, P. Poinot, E. Vigneau, D. Bertrand and C. Prost, *J. Agric. Food. Chem.*, 2008, **56**, 3245–3253.
- [181] S. T. Roweis and L. K. Saul, *Science*, 2000, **290**, 2323–2326.

- [182] M. Janus and A. W. Morawski, *Appl. Catal., B*, 2007, **75**, 118–123.
- [183] M. Gallagher, J. Wysocki, J. J. Leyden, A. I. Spielman, X. Sun and G. Preti, *Br. J. Dermatol.*, 2008, **159**, 780–791.
- [184] D. J. Penn, E. Oberzaucher, K. Grammer, G. Fischer, H. A. Soini, D. Wiesler, M. V. Novotny, S. J. Dixon, Y. Xu and R. G. Brereton, *Journal of the Royal Society Interface*, 2007, **4**, 331–340.
- [185] S. K. Pandey and K. H. Kim, *Trends in Analytical Chemistry*, 2011, **30**, 784–796.
- [186] M. Stitz, K. Gase, I. T. Baldwin and E. Gaquerel, *Plant Physiol.*, 2011, **157**, 341–354.
- [187] C. J. Liu, Y. Furusawa and K. Hayashi, *Sens. Actuators, B*, 2013, **183**, 117–123.
- [188] A. C. Little, B. C. Jones and R. P. Burriss, *Horm. Behav.*, 2007, **51**, 633–639.
- [189] R. J. Stevenson, *Chem. Senses*, 2010, **35**, 3–20.
- [190] S. K. Jha, C. J. Liu and K. Hayashi, *Sens. Actuators, B*, 2014, **204**, 74–87.
- [191] Y. Huang, X. Zhang, E. Ringe, M. J. Hou, L. W. Ma and Z. J. Zhang, *Sci. Rep.*, 2016, **6**, year.
- [192] L. Vera, L. Acena, J. Guasch, R. Boque, M. Mestres and O. Bustos, *Anal. Bioanal. Chem.*, 2011, **399**, 2073–2081.
- [193] A. Szczurek and M. Maciejewska, *Talanta*, 2004, **64**, 609–617.
- [194] H. Kong, S. C. Zhang, N. Na, D. Liu and X. R. Zhang, *Analyst*, 2009, **134**, 2441–2446.
- [195] R. Kurita, H. Tabei, Z. M. Liu, T. Horiuchi and O. Niwa, *Sens. Actuators, B*, 2000, **71**, 82–89.
- [196] J. Goschnick, I. Koronczi, M. Frietsch and I. Kiselev, *Sens. Actuators, B*, 2005, **106**, 182–186.
- [197] A. H. Gomez, J. Wang, G. X. Hu and A. G. Pereira, *Sens. Actuators, B*, 2006, **113**, 347–353.
- [198] S. Panigrahi, S. Balasubramanian, H. Gu, C. M. Logue and M. Marchello, *Sens. Actuators, B*, 2006, **119**, 2–14.
- [199] R. Banerjee, B. Tudu, L. Shaw, A. Jana, N. Bhattacharyya and R. Bandyopadhyay, *J. Food Eng.*, 2012, **110**, 356–363.
- [200] J. M. Gutierrez, Z. Haddi, A. Amari, B. Bouchikhi, A. Mimendia, X. Ceto and M. del Valle, *Sens. Actuators, B*, 2013, **177**, 989–996.
- [201] J. P. Santos, M. J. Fernandez, J. L. Fontech, J. Lozano, M. Aleixandre, M. Garcia, J. Gutierrez and M. C. Horrillo, *Sens. Actuators, B*, 2005, **107**, 291–295.
- [202] C. Di, C. Ning, L. Q. Huang and C. Z. Wang, *Insect Biochem. Mol. Biol.*, 2017, **84**, 48–62.

- [203] J. P. da Graca, T. E. Ueda, T. Janegitz, S. S. Vieira, M. C. Salvador, M. C. N. de Oliveira, S. M. Zingaretti, S. J. Powers, J. A. Pickett, M. A. Birkett and C. B. Hoffmann-Campo, *Phytochemistry*, 2016, **131**, 84–91.
- [204] E. L. Wason, A. A. Agrawal and M. D. Hunter, *J. Chem. Ecol.*, 2013, **39**, 1101–1111.
- [205] S. Tamogami, K. Noge, M. Abe, G. K. Agrawal and R. Rakwal, *Biotechnol. Lett.*, 2013, **35**, 1247–1252.
- [206] S. Tamogami, R. Rakwal and G. K. Agrawal, *Biochem. Biophys. Res. Commun.*, 2008, **376**, 723–727.
- [207] I. S. Sobhy, C. M. Woodcock, S. J. Powers, J. C. Caulfield, J. A. Pickett and M. A. Birkett, *J. Chem. Ecol.*, 2017, **43**, 39–52.
- [208] M. A. Birkett, C. A. M. Campbell, K. Chamberlain, E. Guerrieri, A. J. Hick, J. L. Martin, M. Matthes, J. A. Napier, J. Pettersson, J. A. Pickett, G. M. Poppy, E. M. Pow, B. J. Pye, L. E. Smart, G. H. Wadhams, L. J. Wadhams and C. M. Woodcock, *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, 9329–9334.
- [209] G. Buchi and H. Wuest, *J. Org. Chem.*, 1966, **31**, 977–978.
- [210] C. C. von Dahl and I. T. Baldwin, *Physiol. Plant.*, 2004, **120**, 474–481.
- [211] D. M. Mutyambai, T. J. A. Bruce, J. van den Berg, C. A. O. Midega, J. A. Pickett and Z. R. Khan, *PLoS One*, 2016, **11**, year.
- [212] Z. X. Han, M. M. Rana, G. F. Liu, M. J. Gao, D. X. Li, F. G. Wu, X. B. Li, X. C. Wan and S. Wei, *Food Chem.*, 2016, **212**, 739–748.
- [213] N. Alipour, R. L. Andersson, R. T. Olsson, U. W. Gedde and M. S. Hedenqvist, *Acs applied materials & interfaces*, 2016, **8**, 9946–9953.
- [214] A. Caron, N. Redon, F. Thevenet, B. Hanoune and P. Coddeville, *Build. Environ.*, 2016, **107**, 19–28.
- [215] J. H. Lee and H. G. Byun, *Anal. Bioanal. Chem.*, 2014, **406**, 3929–3929.
- [216] K. Wang, X. W. Guan, S. G. Chai, Q. C. Zou, X. H. Zhang and J. Z. Zhang, *Biosensors & Bioelectronics*, 2015, **64**, 94–101.
- [217] W. A. Braunecker and K. Matyjaszewski, *Prog. Polym. Sci.*, 2007, **32**, 93–146.
- [218] V. Percec, T. Guliashvili, J. S. Ladislaw, A. Wistrand, A. Stjerndahl, M. J. Sienkowska, M. J. Monteiro and S. Sahoo, *J. Am. Chem. Soc.*, 2006, **128**, 14156–14165.
- [219] C. J. Liu, B. Wyszynski, R. Yatabe, K. Hayashi and K. Toko, *Sensors*, 2017, **17**, year.
- [220] J. A. Duke, *Handbook of phytochemical constituent grass, herbs and other economic plants*, CRC press, 1992.
- [221] W. S. Schlotzhauer, S. D. Pair and R. J. Horvat, *J. Agric. Food. Chem.*, 1996, **44**, 206–209.

- [222] K. Tatsuka, S. Suekane, Y. Sakai and H. Sumitani, *J. Agric. Food. Chem.*, 1990, **38**, 2176–2180.
- [223] R. Ahmad, N. Griffete, A. Lamouri, N. Felidj, M. M. Chehimi and C. Mangeney, *Chem. Mater.*, 2015, **27**, 5464–5478.
- [224] J. J. Qiu and W. D. Wei, *J. Phys. Chem. C*, 2014, **118**, 20735–20749.
- [225] T. Shahar, T. Sicron and D. Mandler, *Nano Res.*, 2017, **10**, 1056–1063.
- [226] R. Mungkarndee, I. Techakriengkrai, G. Tumcharern and M. Sukwattanasinitt, *Food Chem.*, 2016, **197**, 198–204.
- [227] R. Selyanchyn, S. Wakamatsu, K. Hayashi and S.-W. Lee, *Sensors*, 2015, **15**, 18834–18850.
- [228] S. Iwamori, K. Yoshino, H. Matsumoto, K. Noda and I. Nishiyama, *Sens. Actuators, B*, 2012, **171**, 769–776.
- [229] U. R. Acharya, S. Bhat, J. E. W. Koh, S. V. Bhandary and H. Adeli, *Comput. Biol. Med.*, 2017, **88**, 72–83.
- [230] I. N. Pashchenko, K. V. Sokolovsky and P. Gavras, *Mon. Not. R. Astron. Soc.*, 2018, **475**, 2326–2343.
- [231] H. X. Luo, P. F. Jia, S. Q. Qiao and S. K. Duan, *Sens. Actuators, B*, 2018, **259**, 241–249.
- [232] C. R. Stephens, H. Huerta and A. Linares, *Machine Learning*, 2018, **107**, 397–441.
- [233] W. Q. Jiang, Y. F. Shen, Y. F. Ding, C. Y. Ye, Y. Zheng, P. Zhao, L. L. Liu, Z. Tong, L. F. Zhou, S. Sun, X. C. Zhang, L. S. Teng, M. P. Timko, L. J. Fan and W. J. Fang, *Int. J. Cancer*, 2018, **142**, 357–368.
- [234] J. Tian, C. Morillo, M. H. Azarian and M. Pecht, *IEEE Trans. Ind. Electron.*, 2016, **63**, 1793–1803.
- [235] C. Richter, E. King, E. Falvey and A. Franklyn-Miller, *J. Biomech.*, 2018, **66**, 1–9.
- [236] A. A. Mohammed, R. Minhas, Q. M. J. Wu and M. A. Sid-Ahmed, *Pattern Recognit.*, 2011, **44**, 2588–2597.
- [237] D. S. Guru, M. Suhil, L. N. Raju and N. V. Kumar, *Pattern Recognit. Lett.*, 2018, **103**, 23–31.
- [238] M. Staton, G. Ciciurkaite, J. Havens, M. Tillson, C. Leukefeld, M. Webster, C. Oser and B. Peteet, *J. Rural Health*, 2018, **34**, 31–41.
- [239] M. Meister, *Elife*, 2015, **4**, e07865.

# Appendix A

## Chapter 2 support information

Table A.1 The list of 178 odorants and their CAS numbers, chemical clusters and t-SNE results.

No.	Name	CAS	Cluster	Olfactory images		Molecular parameters	
				Dimension 1	Dimension 2	Dimension 1	Dimension 2
1	Ethyl propionate	105-37-3	1	0.308139	-1.636572	-4.013656	-0.164298
2	Ethyl butyrate	105-54-4	1	0.740831	-0.914779	-2.678833	-1.153896
3	Methyl caproate	106-70-7	1	4.116115	1.928249	-1.161941	-3.481872
4	Methyl heptanoate	106-73-0	1	4.400900	2.456325	0.004627	-5.879601
5	Ethyl caproate	123-66-0	1	4.103857	1.868916	-6.561180	-9.158452
6	Ethyl acetate	141-78-6	1	-0.597867	-3.402584	-7.224433	6.811835
7	Methyl isocaproate	2412-80-8	1	5.053745	0.341276	-1.703288	-2.271641
8	Ethyl valerate	539-82-2	1	3.422064	1.185465	-1.478769	-3.362447
9	Methyl valerate	624-24-8	1	-0.151081	-1.388754	-2.239833	-1.445676
10	Methyl acetate	79-20-9	1	5.553658	-5.863034	-8.132847	7.409454
11	Methyl 3-methylbutenoate	924-50-5	1	-4.678826	0.870475	-3.336869	0.920892
12	3,7-dimethyloctan-1-ol	106-21-8	2	5.330842	6.790664	2.224548	4.347022
13	Citronellol	106-22-9	2	4.944474	6.653876	-0.802323	-8.985935
14	Geraniol	106-24-1	2	7.101928	-6.798558	-1.023801	-8.971433
15	Nerol	106-25-2	2	2.154527	-10.148940	-1.050802	-8.961532
16	1-hexanol	111-27-3	2	5.655984	3.426531	-0.288197	11.736692
17	1-heptanol	111-70-6	2	4.400446	4.292415	-1.907948	4.784803
18	1-octanol	111-87-5	2	6.474554	4.037191	0.524609	3.650876
19	1-decanol	112-30-1	2	11.364222	5.181892	2.378490	5.246251
20	Isoamyl alcohol	123-51-3	2	8.087529	-2.713978	-3.477695	5.465726
21	9-decen-1-ol	13019-22-2	2	4.997338	6.954799	-0.222486	-8.998510
22	1-nonanol	143-08-8	2	8.240569	3.685157	1.867619	3.945528
23	Phytol	150-86-7	2	-3.118683	-9.460390	7.636538	-12.893249
24	Trans,trans-2,4-hexadien-1-ol	17102-64-6	2	-1.072378	-3.766085	-6.809980	4.291436
25	Lavandulol	1845-51-8	2	-2.980515	-6.876206	-1.590133	-8.613377
26	3-methyl-2-buten-1-ol	556-82-1	2	2.524267	-1.018963	-7.196845	3.229084
27	4-methyl-1-pentanol	626-89-1	2	2.593787	-0.834281	-5.518653	5.753223
28	1-propanol	71-23-8	2	-0.218590	-2.181966	-1.926540	13.824256
29	1-butanol	71-36-3	2	-4.237943	-1.373536	-2.093252	12.839695
30	1-pentanol	71-41-0	2	-7.353788	6.916110	-0.900430	13.160926

No.	Name	CAS	Cluster	Olfactory images		Molecular parameters	
				Dimension 1	Dimension 2	Dimension 1	Dimension 2
31	Nerolidol	7212-44-4	2	0.635197	0.328942	7.267100	-10.192150
32	Isobutanol	78-83-1	2	-3.885450	-0.123942	-2.107489	12.843531
33	5-hexen-1-ol	821-41-0	2	2.995946	3.823765	-7.856308	4.098797
34	Cis-2-hexen-1-ol	928-94-9	2	3.060304	3.245103	-7.607327	4.253326
35	Trans-2-hexen-1-ol	928-95-0	2	4.480986	4.548905	-7.609240	4.253963
36	Cis-3-hexen-1-ol	928-96-1	2	3.194082	3.661717	-7.692431	4.173042
37	Trans-3-hexen-1-ol	928-97-2	2	3.138133	3.770713	-7.640281	4.222150
38	Ethylbenzene	100-41-4	3	-7.059334	-0.479419	-5.856471	6.852228
39	Propylbenzene	103-65-1	3	-10.556522	-4.958207	-6.138080	7.193122
40	Butylbenzene	104-51-8	3	-10.644025	-5.121360	-6.419690	7.534016
41	P-xylene	106-42-3	3	-9.051429	-0.462666	-6.701300	7.874910
42	M-xylene	108-38-3	3	-8.837407	-0.940557	-6.982909	8.215804
43	Mesitylene	108-67-8	3	-9.713529	-0.834414	-7.264519	8.556698
44	Toluene	108-88-3	3	-9.040840	-0.107798	-7.546129	8.897592
45	1,2,3,4-tetramethylbenzene	488-23-3	3	-8.919945	-2.181804	-7.827738	9.238485
46	1,2,3-trimethylbenzene	526-73-8	3	-8.377890	-1.055092	-8.109348	9.579379
47	O-xylene	95-47-6	3	-8.444178	-0.927918	-8.390958	9.920273
48	1,2,4-trimethylbenzene	95-63-6	3	-8.728199	-1.406018	-8.672567	10.261167
49	Tert-butylbenzene	98-06-6	3	-5.398059	-3.568425	-8.954177	10.602061
50	Isopropylbenzene	98-82-8	3	-10.408720	-4.710147	-9.235786	10.942955
51	Methyl phenylacetate	101-41-7	4	-8.854176	5.905904	1.908405	-3.094372
52	Methyl cinnamate	103-26-4	4	-0.343790	7.793530	4.527574	-4.706611
53	Methyl salicylate	119-36-8	4	-8.854942	2.613182	-6.397805	-5.673913
54	Phenyl acetate	122-79-2	4	-8.973787	5.925920	2.129654	-0.571213
55	Methyl anthranilate	134-20-3	4	8.845433	-8.887504	-5.794615	-5.279344
56	Benzyl acetate	140-11-4	4	-8.786029	6.197205	2.464882	-3.134664
57	Phenethyl hexanoate	6290-37-5	4	9.810109	0.460857	5.916659	-11.069476
58	Phenyl propionate	637-27-4	4	-8.961146	5.818830	2.621772	-2.944608
59	Methyl o-toluate	89-71-4	4	-7.658230	1.023338	3.008999	-1.850014
60	Methyl benzoate	93-58-3	4	-5.681050	9.729186	2.696643	-1.480043
61	Ethyl heptanoate	106-30-9	5	6.712290	5.075457	1.215802	-7.245925
62	Ethyl caprylate	106-32-1	5	6.579837	5.248027	3.028166	-8.261959
63	Methyl octanoate	111-11-5	5	7.681458	5.308768	1.314122	-7.249644
64	Methyl 2-octynoate	111-12-6	5	5.066063	-2.530888	1.861780	-7.122384
65	Methyl trans-2-octenoate	7367-81-9	5	7.774600	5.364214	2.308403	-6.485896
66	Hexane	110-54-3	6	11.805715	-4.756393	0.316227	13.778552
67	Octane	111-65-9	6	7.655288	0.684015	5.084964	6.944221
68	Nonane	111-84-2	6	6.949939	8.091359	4.517381	4.969145
69	Undecane	1120-21-4	6	5.423283	-5.169142	4.588507	11.192247
70	Dodecane	112-40-3	6	5.181754	-10.544674	4.893858	11.158640
71	Decane	124-18-5	6	7.209423	8.545395	4.398122	11.209378
72	Heptane	142-82-5	6	-4.422161	3.006847	1.344280	13.614320
73	2,2,4-trimethylpentane	540-84-1	6	1.581120	-3.993377	5.465401	5.658543
74	Hexadecane	544-76-3	6	5.766831	8.530791	1.988600	-11.751704
75	2,3,4-trimethylpentane	565-75-3	6	-4.173204	-8.960370	5.465511	5.651912
76	2,3-dimethylhexane	584-94-1	6	1.371373	-2.842388	4.301026	6.322262
77	2,4-dimethylhexane	589-43-5	6	1.655887	-3.256643	4.231407	6.405684
78	4-methylheptane	589-53-7	6	2.359899	-2.906029	4.200278	6.433503
79	3-methylheptane	589-81-1	6	2.751822	-3.568203	4.237862	6.371823
80	2,2-dimethylhexane	590-73-8	6	1.917088	-3.766368	4.198862	6.195062
81	2,5-dimethylhexane	592-13-2	6	2.111306	-3.738331	5.432391	5.635077
82	2-methylheptane	592-27-8	6	2.777103	-3.349493	4.130500	6.106297

No.	Name	CAS	Cluster	Olfactory images		Molecular parameters	
				Dimension 1	Dimension 2	Dimension 1	Dimension 2
83	Tridecane	629-50-5	6	6.933434	9.894933	5.603600	11.087695
84	Tetradecane	629-59-4	6	6.828951	10.600637	5.931582	11.042735
85	Pentadecane	629-62-9	6	1.451709	2.957249	6.156611	11.018596
86	Benzaldehyde	100-52-7	7	-8.252113	1.638883	-1.408840	2.945818
87	Cyclamen aldehyde	103-95-7	7	-5.482128	5.280167	3.909895	-9.598995
88	P-tolualdehyde	104-87-0	7	-10.135870	0.966484	1.804289	0.492592
89	4-hydroxybenzaldehyde	123-08-0	7	-2.345065	11.974445	-8.184002	-2.710606
90	P-anisaldehyde	123-11-5	7	-11.251067	4.360659	1.265178	-1.179405
91	Cinnamaldehyde	14371-10-9	7	-4.095995	-3.163852	2.593487	0.798744
92	2,4-dimethylbenzaldehyde	15764-16-6	7	-9.841910	1.415331	2.593446	-0.125427
93	O-tolualdehyde	529-20-4	7	-9.999970	1.058446	1.564145	1.891187
94	M-anisaldehyde	591-31-1	7	-5.997152	2.212166	2.067555	-1.206375
95	M-tolualdehyde	620-23-5	7	-10.151842	1.419335	1.516890	1.883709
96	Salicylaldehyde	90-02-8	7	-9.354719	4.064882	-7.722842	-4.094855
97	1-octene	111-66-0	8	8.802660	0.414177	9.450730	3.787736
98	Trans-2-octene	13389-42-9	8	9.227975	0.504293	9.146700	3.735935
99	Trans-4-octene	14850-23-8	8	8.737789	0.253822	9.126331	3.769447
100	4-octyne	1942-45-6	8	9.421871	-0.460765	8.779719	4.666908
101	2-octyne	2809-67-8	8	9.733129	0.040741	8.782606	4.675715
102	1,7-octadiene	3710-30-3	8	10.101167	0.078784	9.069393	4.331640
103	Alpha-farnesene	502-61-4	8	6.489703	-3.689738	6.452899	-10.240498
104	2-methyl-2-pentene	625-27-4	8	4.459388	-0.096654	1.732936	7.557686
105	1-octyne	629-05-0	8	8.452683	-4.818006	8.996157	5.155472
106	Cis-4-octene	7642-15-1	8	9.826534	-0.512128	8.725622	3.787148
107	1,7-octadiyne	871-84-1	8	-2.285555	-8.029442	8.850776	5.957025
108	2-methyl-3-butene-2-ol	115-18-4	9	-10.529038	-2.484770	-7.530438	2.622327
109	2-octanol	123-96-6	9	5.464221	2.678425	0.534236	3.652331
110	2-heptanol	543-49-7	9	1.548013	0.987405	-1.944491	4.779940
111	4-heptanol	589-55-9	9	-2.289502	0.806572	1.438010	12.420286
112	3-heptanol	589-82-2	9	2.057118	0.456405	1.638529	12.510844
113	4-nonanol	5932-79-6	9	2.068266	-0.145569	2.558697	11.862305
114	2-pentanol	6032-29-7	9	1.430014	0.071962	-0.874578	13.146917
115	5-nonanol	623-93-8	9	8.611494	2.581700	2.145860	11.821026
116	3-nonanol	624-51-1	9	8.089071	2.789220	2.423837	12.220817
117	2-hexanol	626-93-7	9	1.515785	0.740559	-0.295948	11.725396
118	2-nonanol	628-99-9	9	7.933796	2.853832	2.613899	4.621985
119	Isopropanol	67-63-0	9	-2.884529	0.075958	-1.925531	13.824682
120	Tert-butanol	75-65-0	9	-3.260873	0.061624	-2.100547	12.841559
121	Linalool	78-70-6	9	-3.011682	-6.914612	-2.307149	-9.018756
122	Sec-butanol	78-92-2	9	-1.034204	0.310292	-2.107860	12.843922
123	2,6-dimethyl-5-heptenal	106-72-9	10	5.584676	0.390244	2.346862	-4.171114
124	3-methyl-2-butenal	107-86-8	10	9.432009	-8.236072	-4.230098	3.926343
125	Pentanal	110-62-3	10	0.629453	-6.296503	-3.418399	2.961632
126	Heptanal	111-71-7	10	3.423291	5.717902	-0.908830	-1.348438
127	Decanal	112-31-2	10	7.019346	9.244042	3.362765	-7.541365
128	Octanal	124-13-0	10	7.383443	6.407657	0.566433	-4.211043
129	Nonanal	124-19-6	10	7.562127	6.481057	1.883686	-5.949048
130	Trans-2,4-decadienal	25152-84-5	10	11.321676	5.201920	3.724621	-6.585705
131	Tiglic aldehyde	497-03-0	10	-3.460553	1.183222	-4.173035	3.922369
132	Trans-2,cis-6-nonadienal	557-48-2	10	9.165842	5.241045	2.886755	-4.793063
133	Hexanal	66-25-1	10	3.116937	5.718837	-2.316046	1.454263
134	Trans-2-tridecenal	7774-82-5	10	5.131192	-10.712749	5.898439	-9.826008

No.	Name	CAS	Cluster	Olfactory images		Molecular parameters	
				Dimension 1	Dimension 2	Dimension 1	Dimension 2
135	3-heptanone	106-35-4	11	-6.409458	2.311096	-1.590810	-0.454217
136	2-pentanone	107-87-9	11	5.348936	-8.072342	-4.640771	3.970572
137	5-methyl-2-hexanone	110-12-3	11	-7.002380	1.412871	-1.603469	0.572747
138	2-heptanone	110-43-0	11	5.897741	-1.107662	-1.606404	-0.432622
139	2-octanone	111-13-7	11	-3.629550	4.485588	-0.595361	-2.182909
140	4-heptanone	123-19-3	11	-6.396251	2.407712	-6.649379	-9.148906
141	4-methyl-3-penten-2-one	141-79-7	11	-3.826644	0.922202	-3.968557	3.683062
142	3-hexanone	589-38-8	11	-5.836923	2.549238	-2.931934	1.859659
143	2-hexanone	591-78-6	11	-2.306077	-2.982856	-2.909820	1.792638
144	Acetone	67-64-1	11	-0.350873	-9.070539	-8.374978	7.486774
145	Pinacolone	75-97-8	11	-11.909026	-0.456615	-3.236341	2.253358
146	2-butanone	78-93-3	11	-2.523891	-0.761153	-7.438159	6.972857
147	Tert-butylacetic acid	1070-83-3	12	3.044546	-0.999903	-8.590593	-0.594879
148	Butyric acid	107-92-6	12	-0.580538	-0.115720	-11.193853	5.081952
149	Valeric acid	109-52-4	12	-0.514718	4.383894	-7.998776	0.719526
150	Undecylenic acid	112-38-9	12	7.955585	10.057254	5.584067	-8.325031
151	Octanoic acid	124-07-2	12	0.045245	4.529572	-3.518018	-6.820454
152	Trans-2-hexenoic acid	13419-69-7	12	-5.963778	4.845471	-8.236337	-1.069232
153	Trans-2-pentenoic acid	13991-37-2	12	10.047475	-8.861898	-7.868631	0.989849
154	Caproic acid	142-62-1	12	0.497529	4.321718	-8.344157	-0.716757
155	2-methylcyclopropanecarboxylic acid	29555-02-0	12	5.485896	-5.120040	-9.312122	0.254366
156	Cyclopentanecarboxylic acid	3400-45-1	12	8.376621	-9.401661	-8.966192	-0.348175
157	Cyclobutanecarboxylic acid	3721-95-7	12	9.322282	-9.228657	-9.329647	0.231090
158	Isovaleric acid	503-74-2	12	9.765827	-8.728375	-9.156543	0.545336
159	2-octynoic acid	5663-96-7	12	5.077249	-2.579765	-3.546829	-6.870459
160	4-pentenoic acid	591-80-0	12	-7.908308	3.111154	-8.114101	1.102013
161	2-methylbutyric acid	600-07-7	12	9.184447	-8.970454	-9.239638	0.656164
162	Formic acid	64-18-6	12	0.345539	5.430430	-10.886136	5.689774
163	Acetic acid	64-19-7	12	-1.543464	5.440512	-11.006661	5.468649
164	4-methylvaleric acid	646-07-1	12	9.470038	-8.279746	-8.002273	0.058014
165	Trimethylacetic acid	75-98-9	12	-1.372275	-4.776557	-9.113965	1.514087
166	Propionic acid	79-09-4	12	-1.427947	5.316183	-11.115910	5.240089
167	Isobutyric acid	79-31-2	12	-2.312307	-2.924788	-9.263040	1.859283
168	2,5-hexanedione	110-13-4	13	-1.334655	-1.277514	-5.278023	0.358780
169	2,3-hexanedione	3848-24-6	13	-1.619389	-1.000069	-5.164887	0.537244
170	2,3-butanedione	431-03-8	13	-1.107181	-1.988397	-6.089848	2.077969
171	3,4-hexanedione	4437-51-8	13	-1.690373	-0.979479	-5.154532	0.619272
172	Acetoin	513-86-0	13	-0.480628	-0.657918	-9.037024	2.208270
173	Diethyl malonate	105-53-3	14	3.481725	-1.979833	-1.919724	-5.742164
174	Dimethyl malonate	108-59-8	14	4.528733	-1.451245	-6.020491	-2.013110
175	2-butoxyethanol acetate	112-07-2	14	9.527465	7.813824	-1.039225	-6.214310
176	Diethyl succinate	123-25-1	14	4.305611	-1.636995	-6.413418	-9.161686
177	Diethyl suberate	2050-23-9	14	4.009269	-1.891318	5.681720	-12.043036
178	Diethyl oxalate	95-92-1	14	3.633788	-2.025922	-5.282857	-2.938382

## Appendix B

### Chapter 3 support information

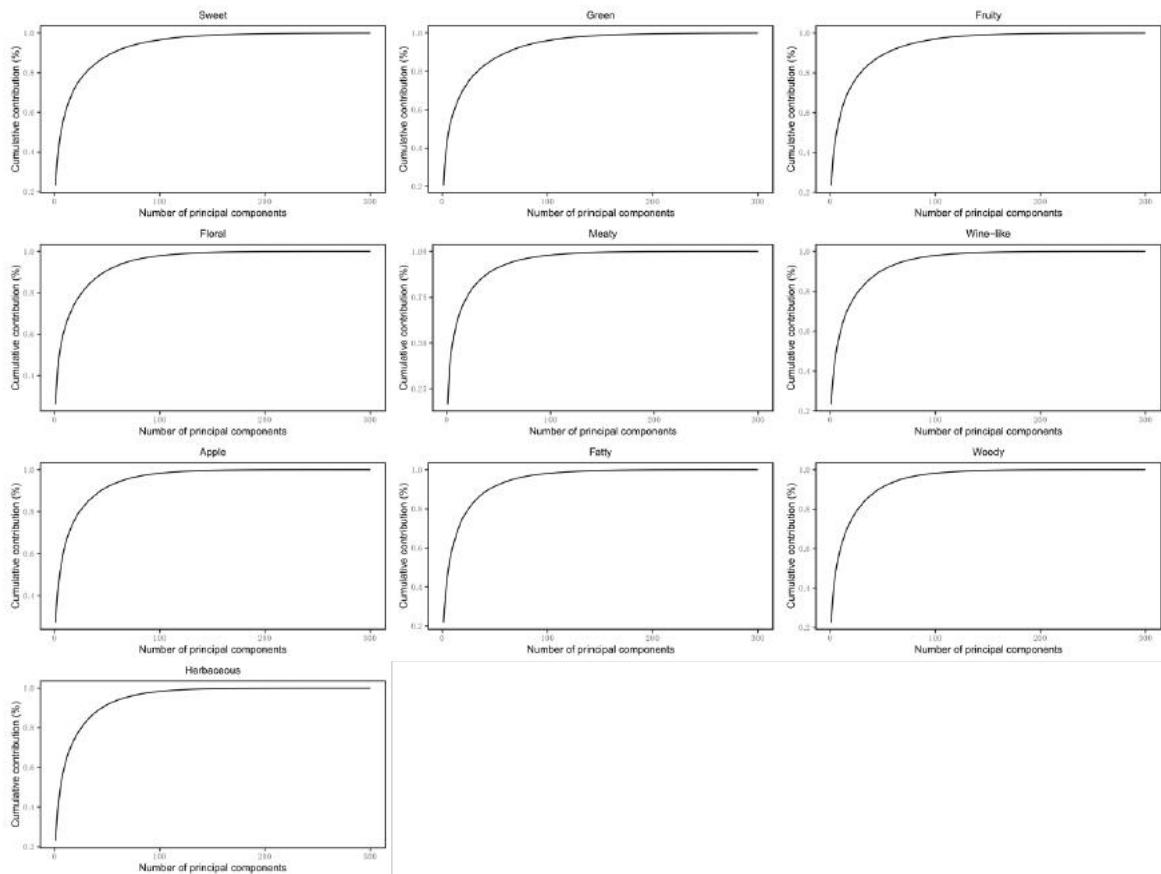


Fig. B.1 The accumulative contribution rates of the first 300 PCs for ten OPs.

Table B.1 Odor perceptions appeared in Sigma Aldrich database.

No.	Words	Num	No.	Words	Num	No.	Words	Num	No.	Words	Num
1	Sweet	198	41	Odorless	28	81	Mushroom	5	121	Vinegar	2
2	Green	192	42	Plum	26	82	Roasted	5	122	Animalic	1
3	Fruity	133	43	Orange	25	83	Blossom	4	123	Antiseptic	1
4	Floral	81	44	Balsam	24	84	Gardenia	4	124	Bacon	1
5	Meaty	80	45	Medicinal	24	85	Lavender	4	125	Baked	1
6	Wine-like	81	46	Pear	24	86	Leafy	4	126	Bitter	1
7	Apple	76	47	Strawberry	24	87	Mossy	4	127	Blackcurrant	1
8	Fatty	75	48	Anise	22	88	Phenolic	4	128	Brandy	1
9	Woody	74	49	Melon	22	89	Pungent	4	129	Bread	1
10	Herbaceous	72	50	Fishy	21	90	Tomato	4	130	Butterscotch	1
11	Sulfurous	71	51	Lemon	21	91	Tropical	4	131	Cantaloupe	1
12	Ethereal	65	52	Jasmine	18	92	Warm	4	132	Cashew	1
13	Nutty	63	53	Pepper	18	93	Whiskey	4	133	Cedar	1
14	Spicy	62	54	Raspberry	18	94	Acidic	3	134	Civet	1
15	Oily	58	55	Smoky	16	95	Alliaceous	3	135	Clover	1
16	Earthy	55	56	Camphoraceous	14	96	Carnation	3	136	Coriander	1
17	Pineapple	55	57	Cinnamon	14	97	Celery	3	137	Corn	1
18	Waxy	53	58	Hyacinth	13	98	Cheesy	3	138	Fleshy	1
19	Creamy	52	59	Sour	13	99	Cucumber	3	139	Fresh	1
20	Rose	52	60	Clove	12	100	Dairy	3	140	Garlic	1
21	Musty	48	61	Soapy	12	101	Fennel	3	141	Green bean	1
22	Cheese	47	62	Grapefruit	11	102	Leather	3	142	Hopoil	1
23	Chocolate	46	63	Hazelnut	11	103	Milk	3	143	Iris	1
24	Citrus	46	64	Tobacco	11	104	Pine	3	144	Lard	1
25	Coffee	46	65	Peanut	10	105	Plastic	3	145	Lily	1
26	Vanilla	46	66	Lime	9	106	Rum	3	146	Mango	1
27	Minty	44	67	Potato	9	107	Beer	2	147	Mesquite	1
28	Banana	43	68	Alcohol	8	108	Blueberry	2	148	Metallic	1
29	Berry	43	69	Violet	8	109	Cabbage	2	149	Papaya	1
30	Caramel	43	70	Balsamic	6	110	Caraway	2	150	Paper	1
31	Butter	40	71	Geranium	6	111	Clean	2	151	Peony	1
32	Vegetable	39	72	Hawthorne	6	112	Cranberry	2	152	Peppermint	1
33	Honey	36	73	Jam	6	113	Grassy	2	153	Quince	1
34	Grape	35	74	Maple	6	114	Licorice	2	154	Rhubarb	1
35	Apricot	34	75	Walnut	6	115	Mild	2	155	Rich	1
36	Cherry	34	76	Beef	5	116	Onion	2	156	Spearmint	1
37	Coconut	34	77	Burnt	5	117	Orchid	2	157	Sweet strawberry	1
38	Almond	33	78	Coumarin	5	118	Rancid	2	158	Tallow	1
39	Peach	30	79	Horseradish	5	119	Sage	2	159	Tart	1
40	Faint	29	80	Lilac	5	120	Seedy	2	160	Vanilla honey	1

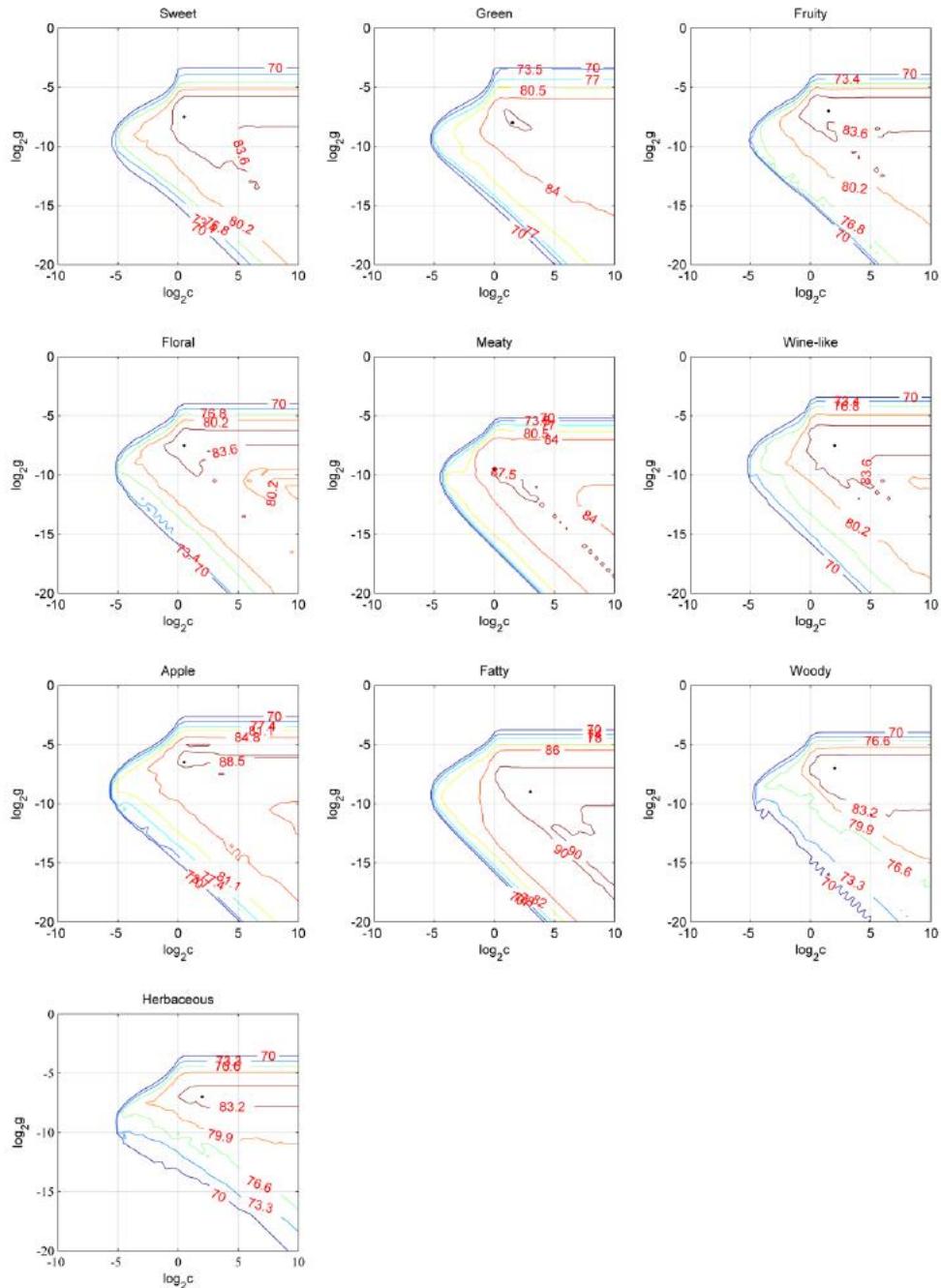


Fig. B.2 The grid search process of penalty factor ( $c$ ) and RBF parameter ( $g$ ) for SVM models by 5-fold cross validation under all parameters. The black circle represents the point at which the optimal  $c$  and  $g$  were chosen.

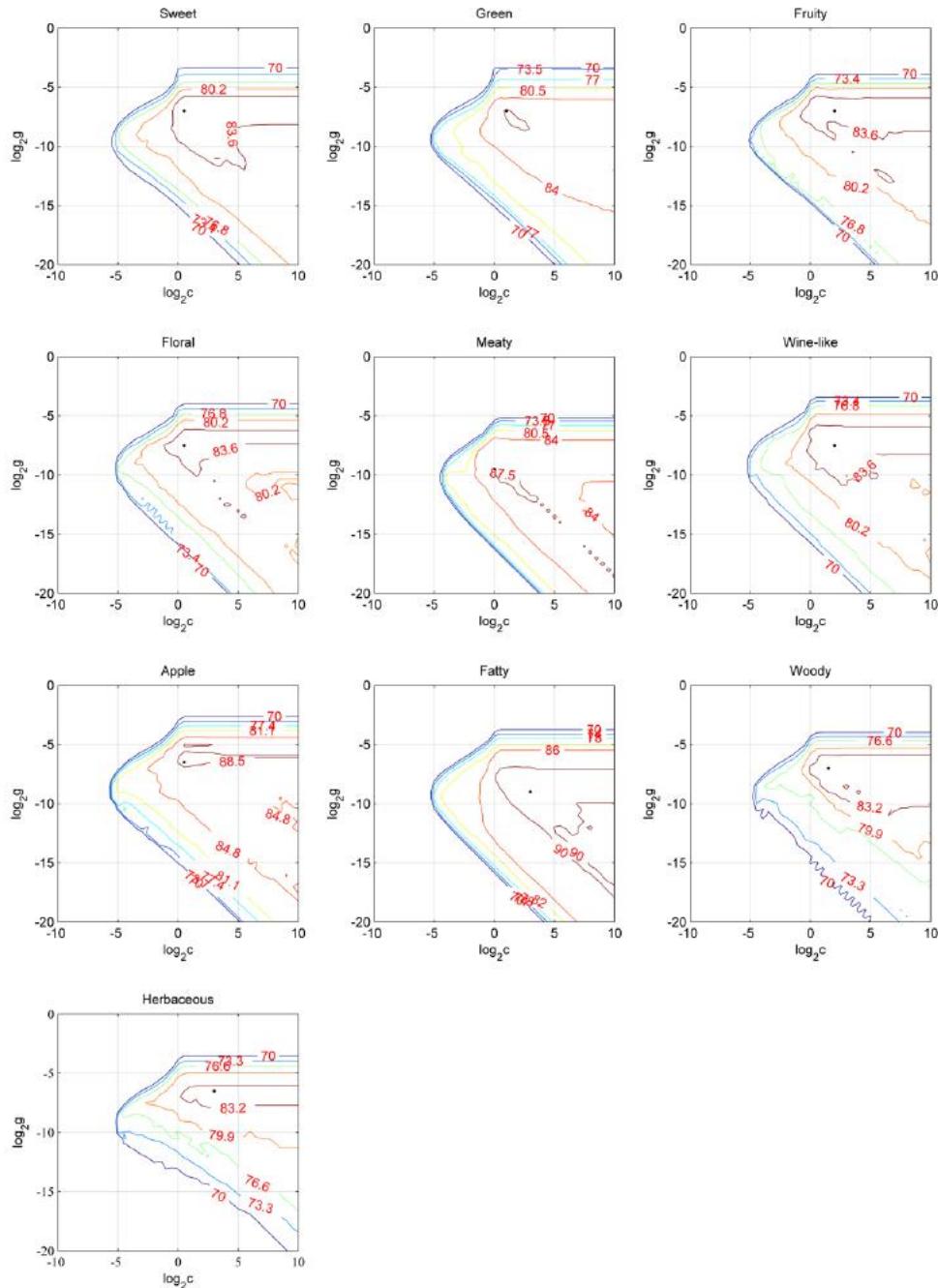


Fig. B.3 The grid search process of penalty factor ( $c$ ) and RBF parameter ( $g$ ) for SVM models by 5-fold cross validation under features obtained by PCA. The black circle represents the point at which the optimal  $c$  and  $g$  were chosen.

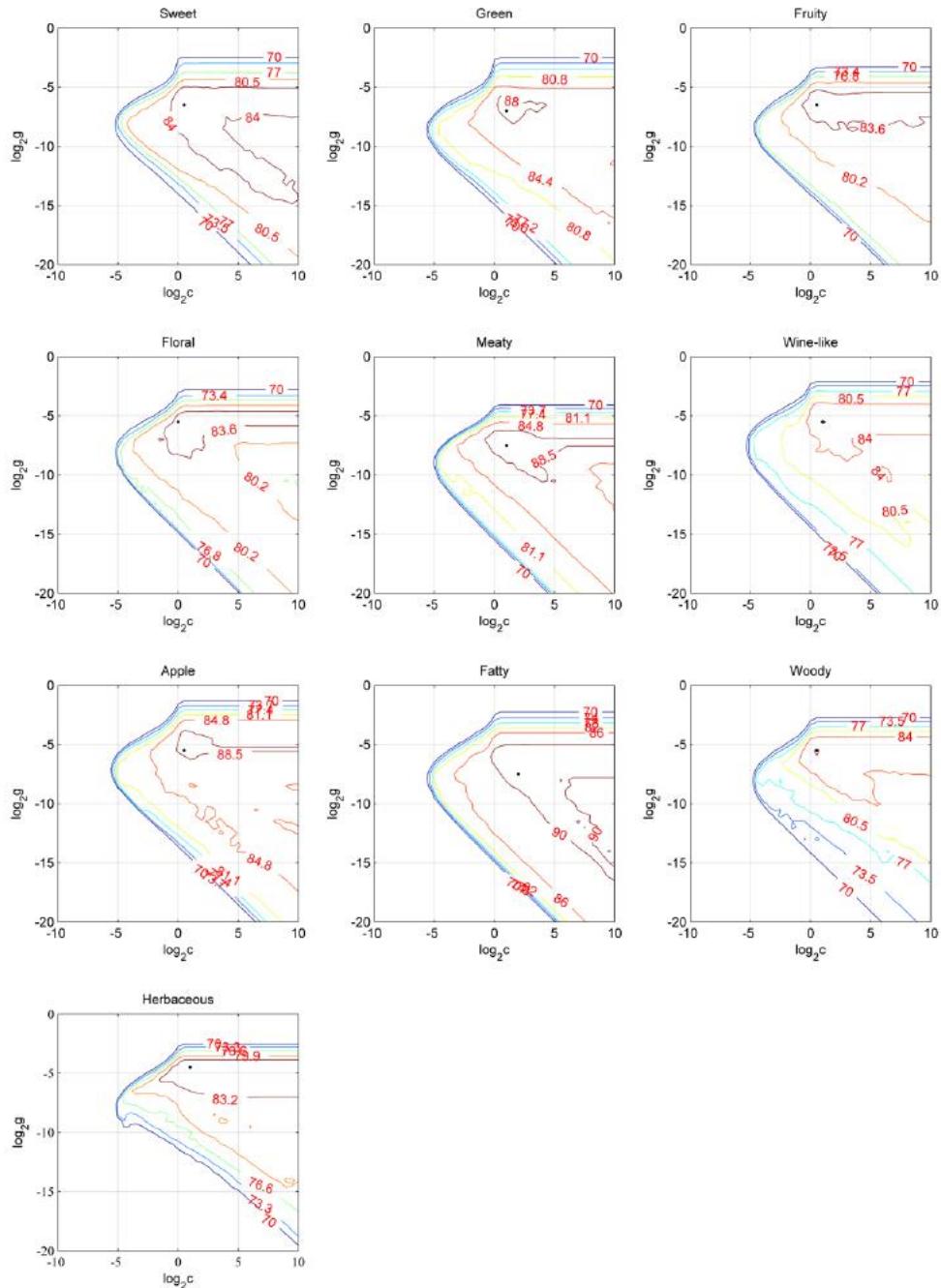


Fig. B.4 The grid search process of penalty factor ( $c$ ) and RBF parameter ( $g$ ) for SVM models by 5-fold cross validation under features obtained by BR-CT. The black circle represents the point at which the optimal  $c$  and  $g$  were chosen.

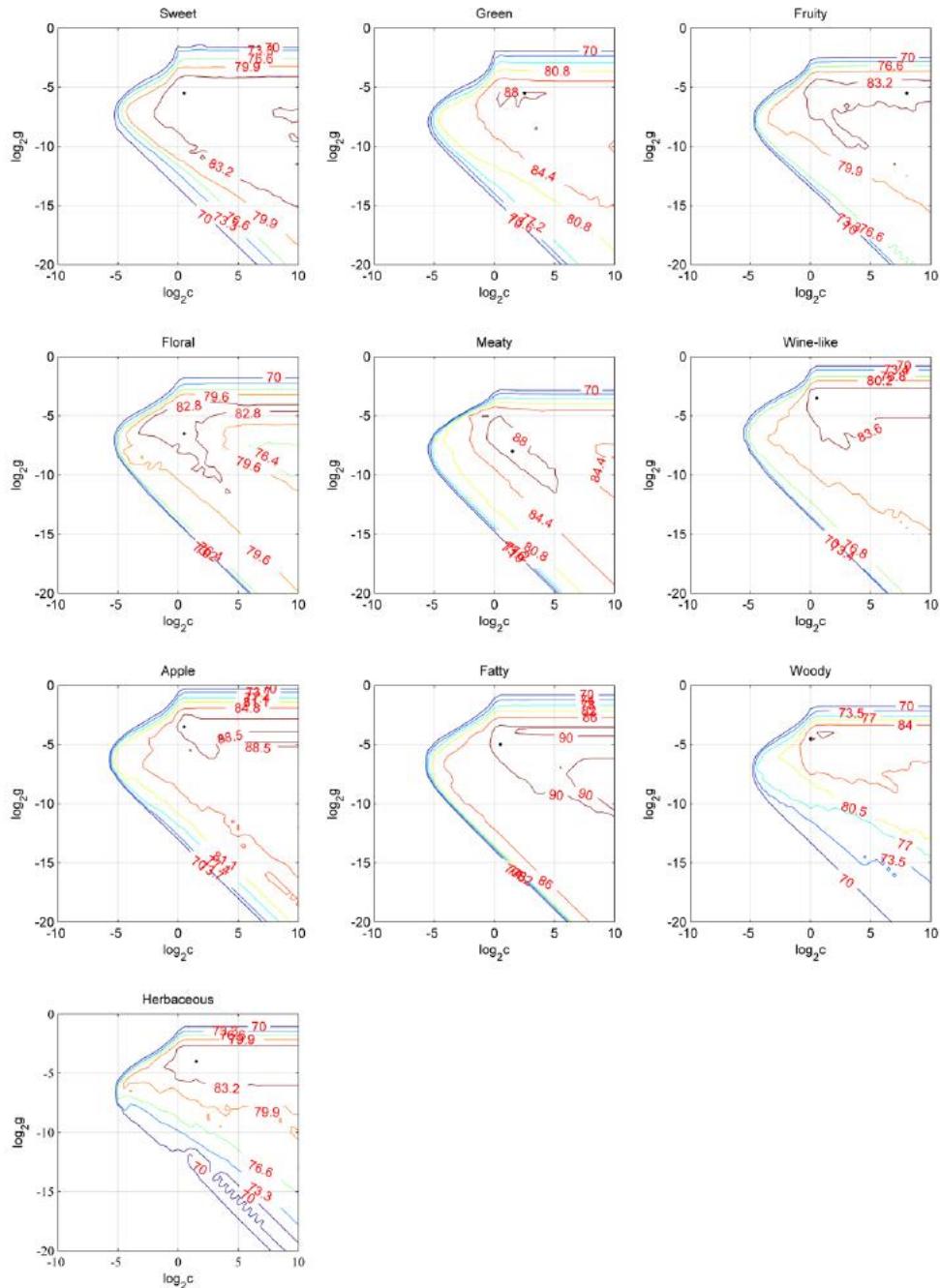


Fig. B.5 The grid search process of penalty factor ( $c$ ) and RBF parameter ( $g$ ) for SVM models by 5-fold cross validation under features obtained by BR-C. The black circle represents the point at which the optimal  $c$  and  $g$  were chosen.

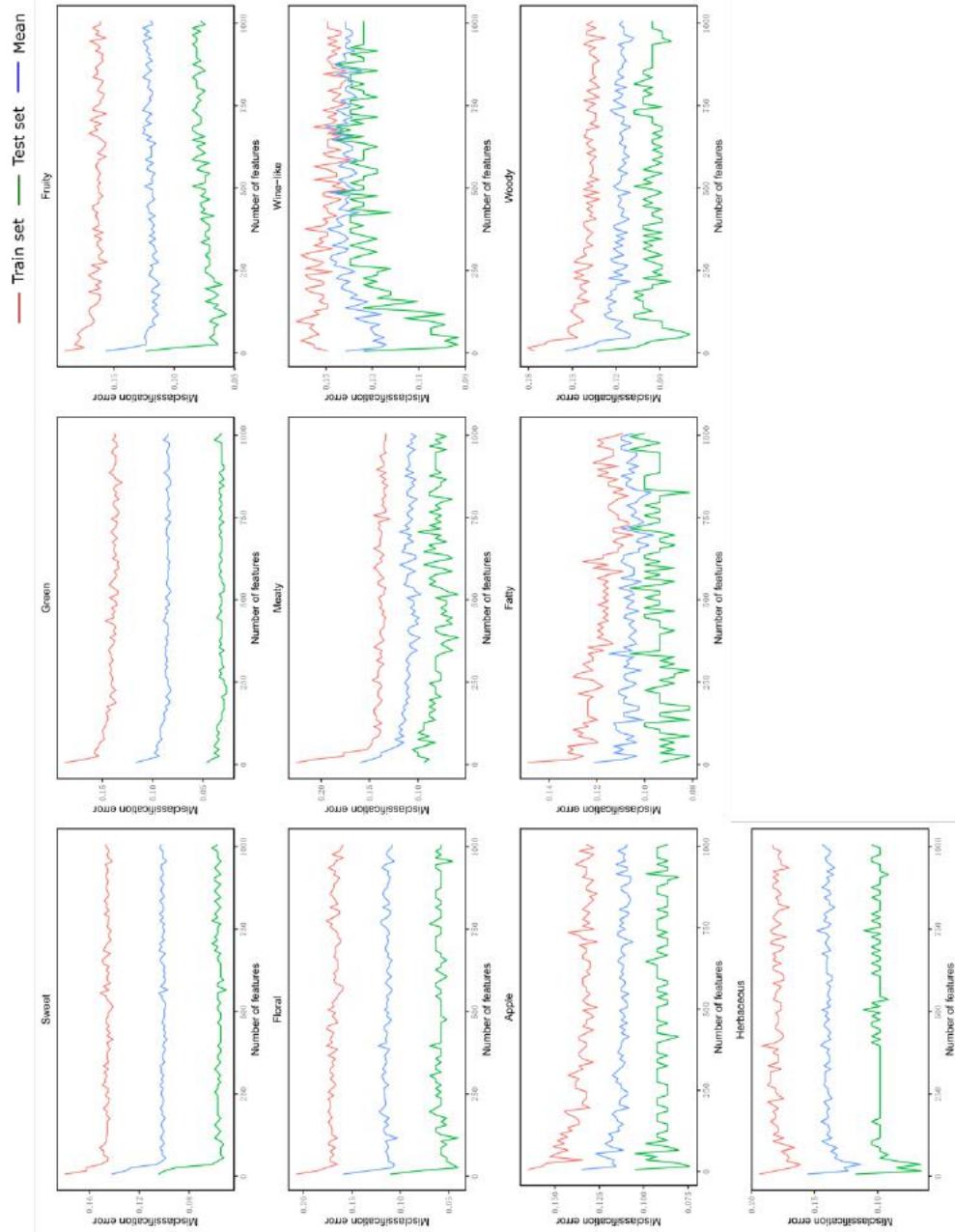


Fig. B.6 Impact of number of features on the misclassification error for RF models under all parameters.

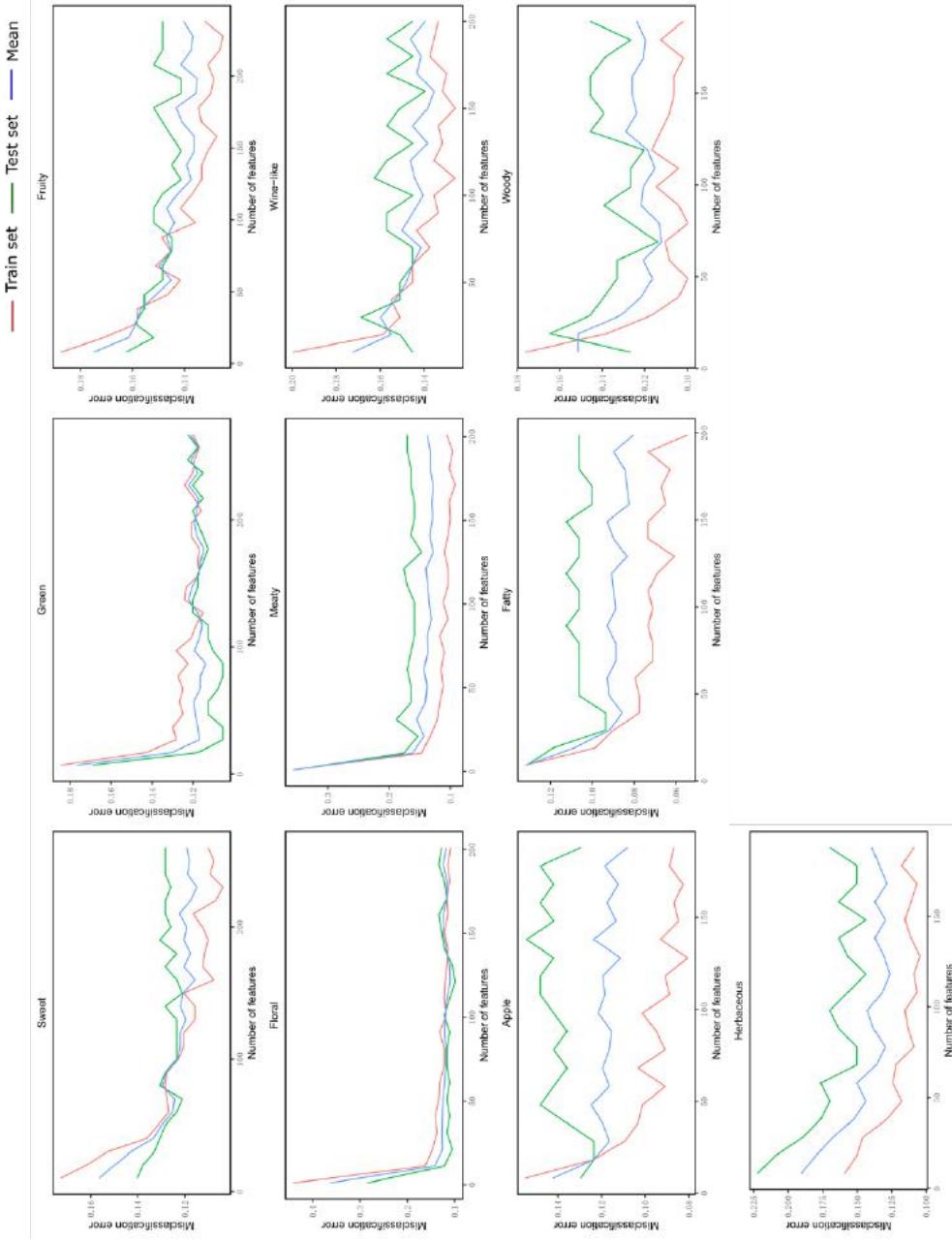
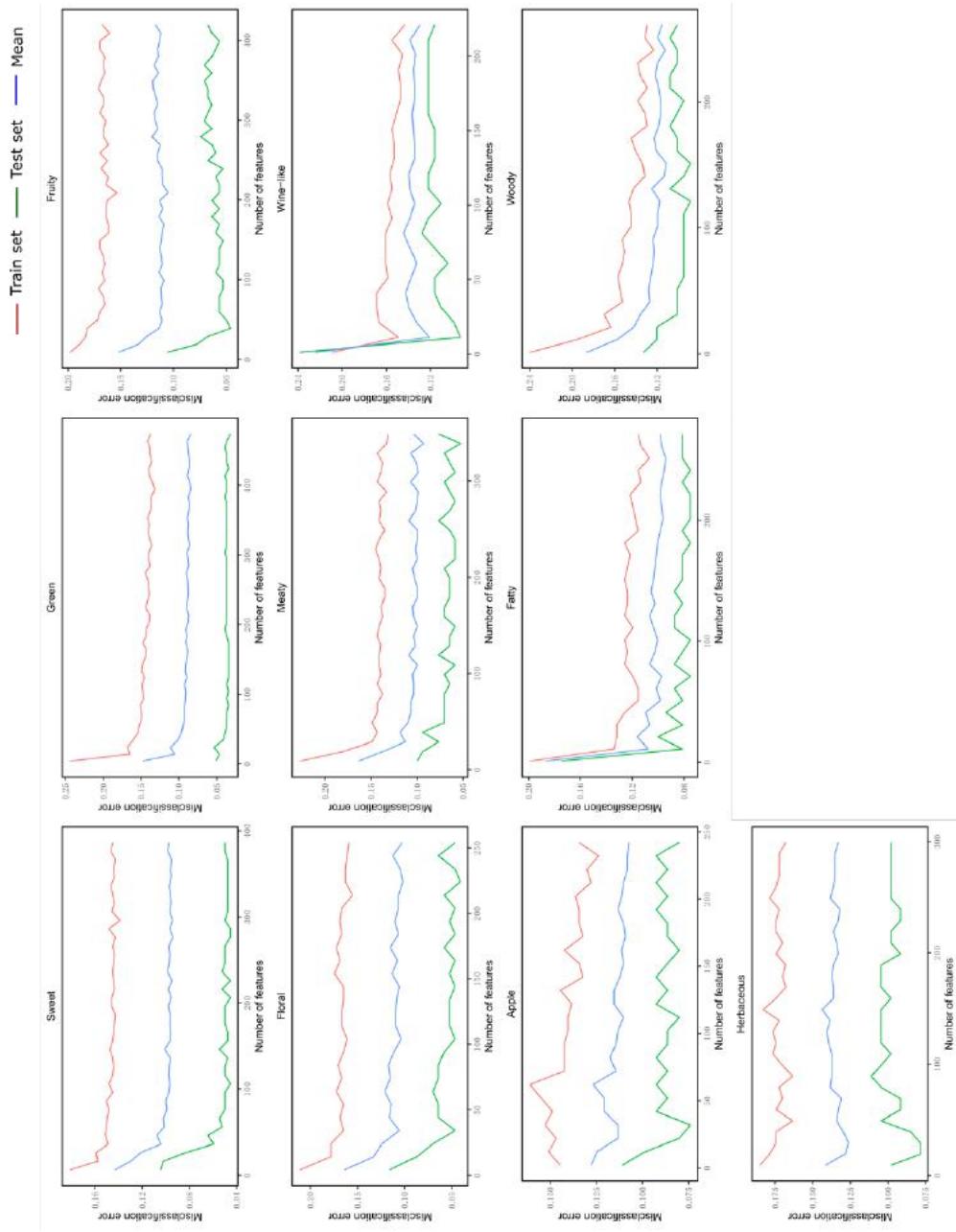


Fig. B.7 Impact of number of features on the misclassification error for RF models features obtained by PCA.



**Fig. B.8** Impact of number of features on the missclassification error for RF models under features obtained by BR-CT.

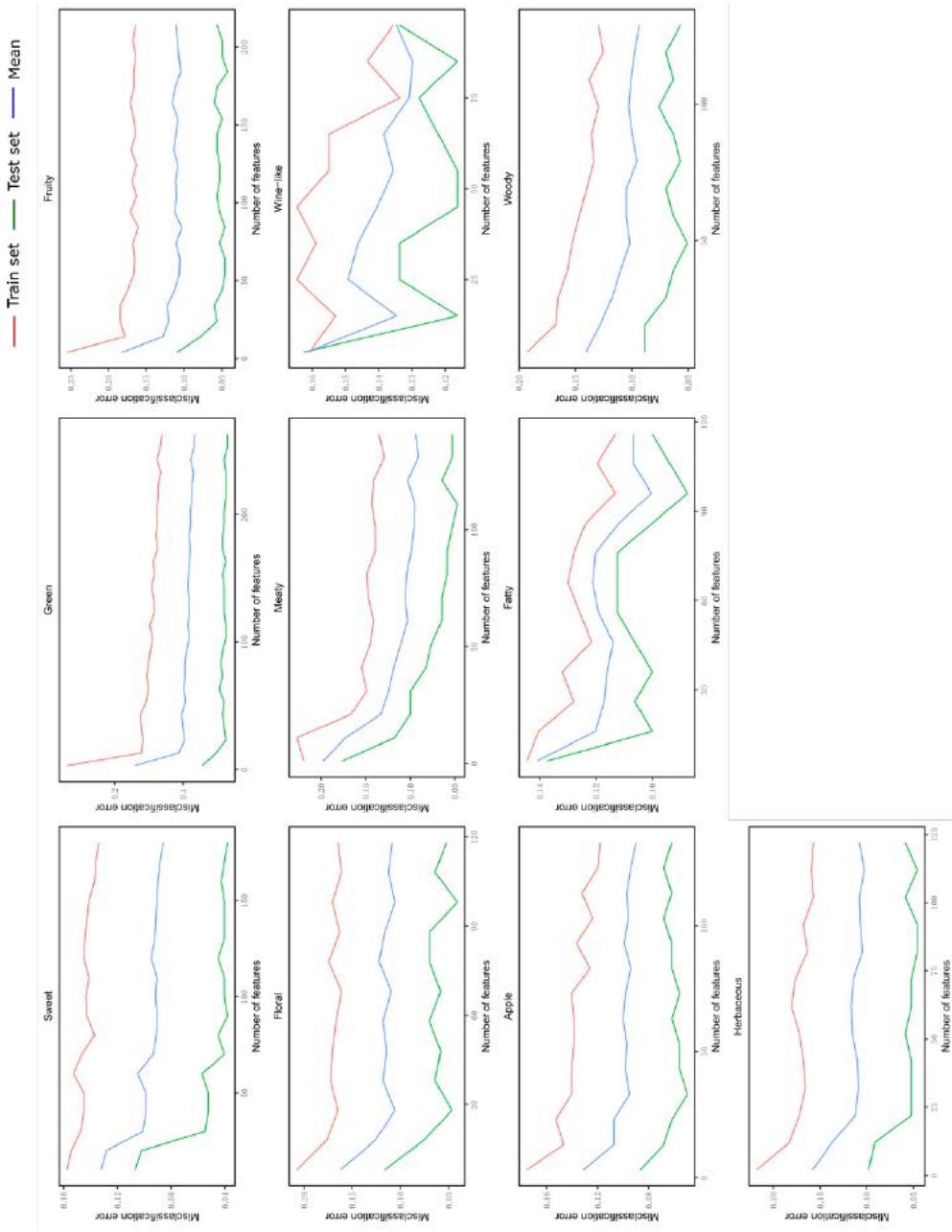


Fig. B.9 Impact of number of features on the misclassification error for RF models under features obtained by BR-C.

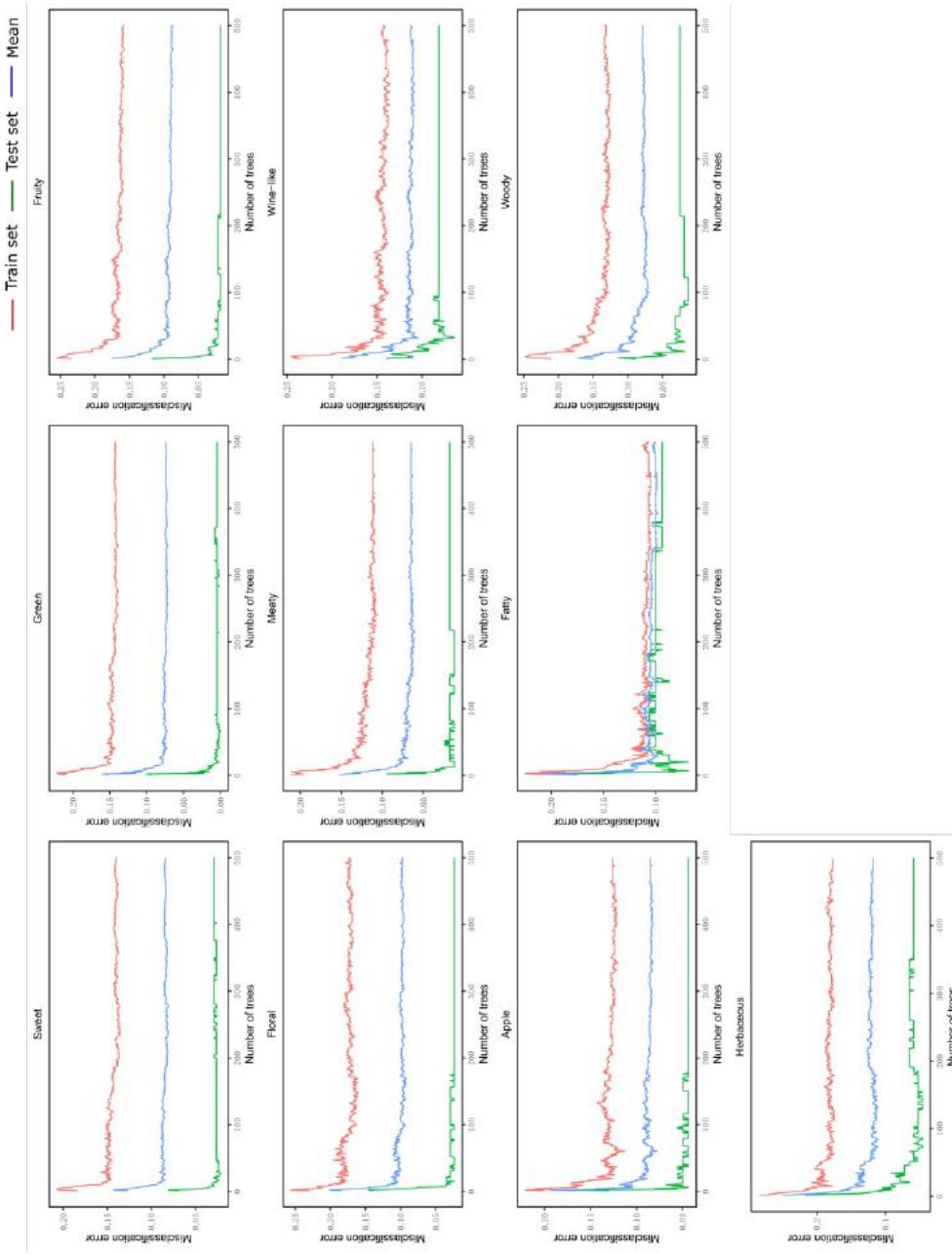


Fig. B.10 Impact of number of trees on the misclassification error for RF models under all parameters.

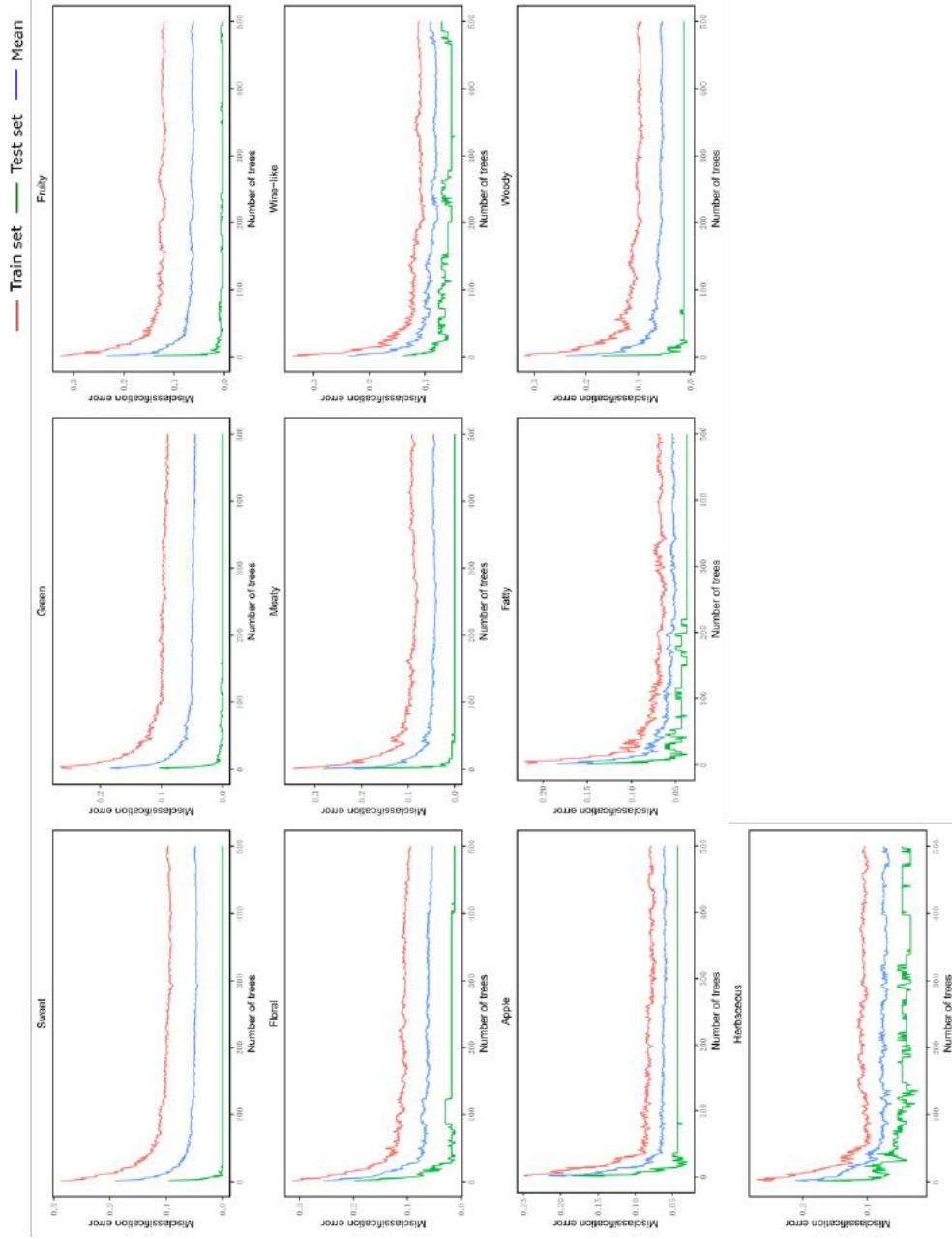


Fig. B.11 Impact of number of trees on the misclassification error for RF models features obtained by PCA.

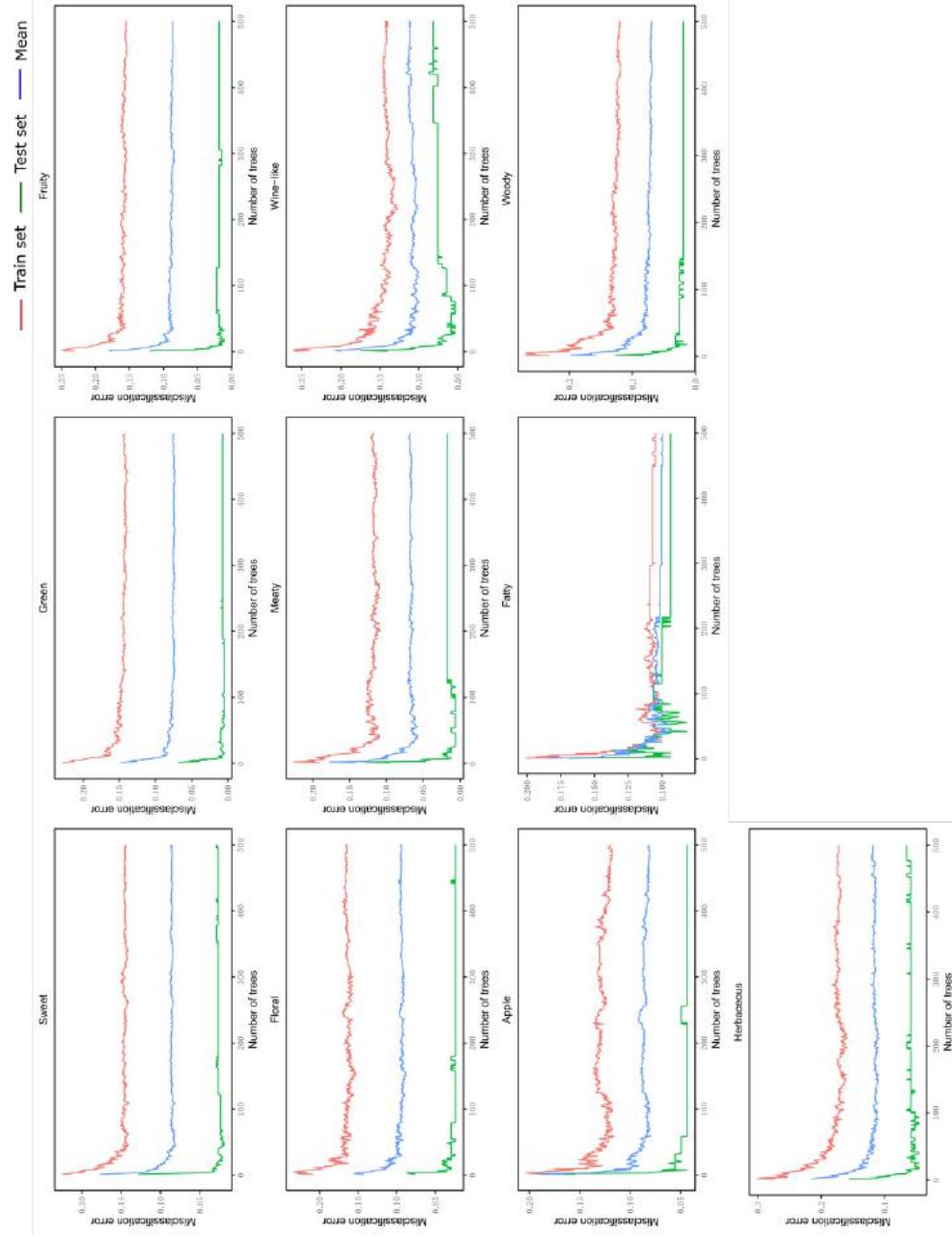


Fig. B.12 Impact of number of trees on the misclassification error for RF models under features obtained by BR-CT.

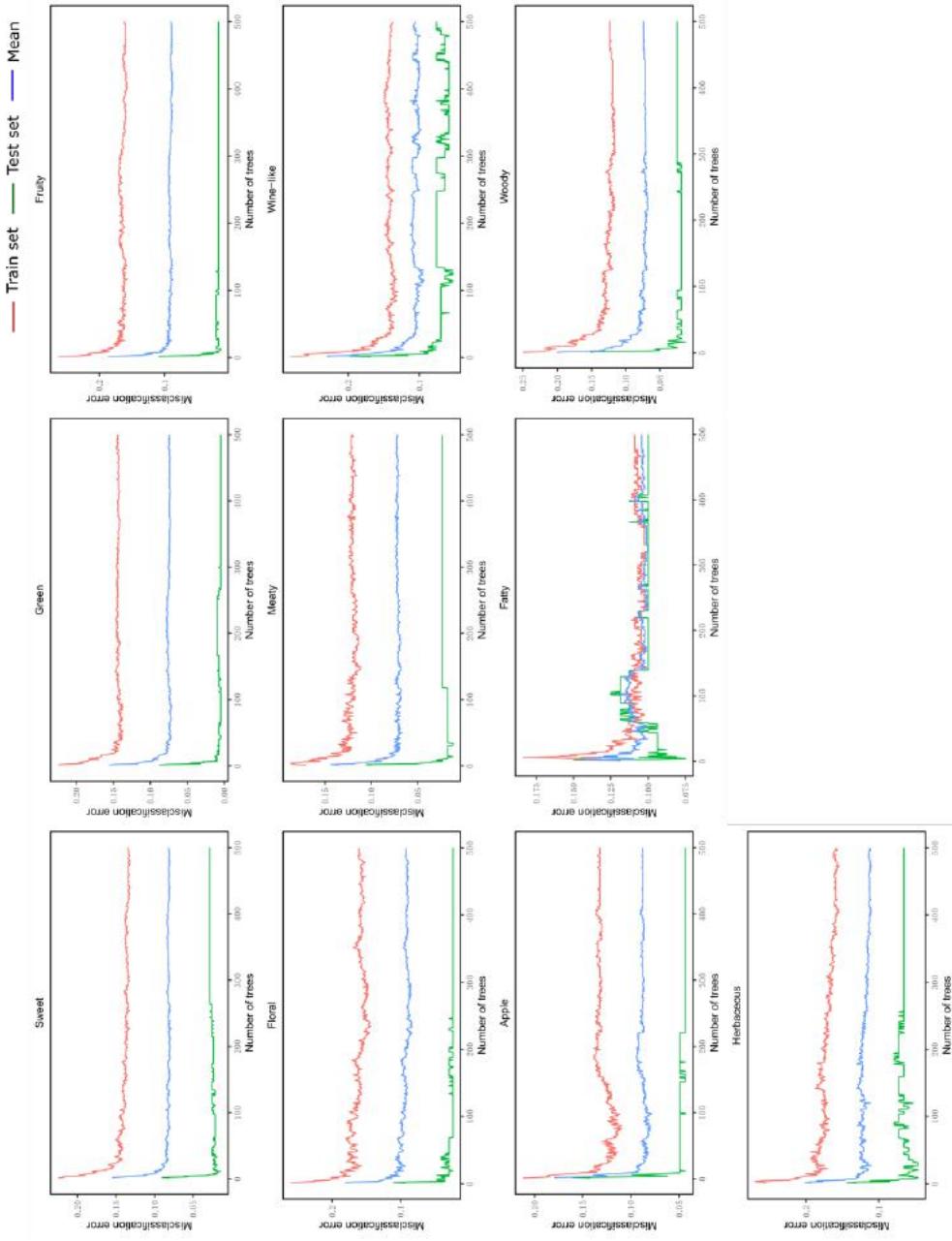


Fig. B.13 Impact of number of trees on the misclassification error for RF models under features obtained by BR-C.

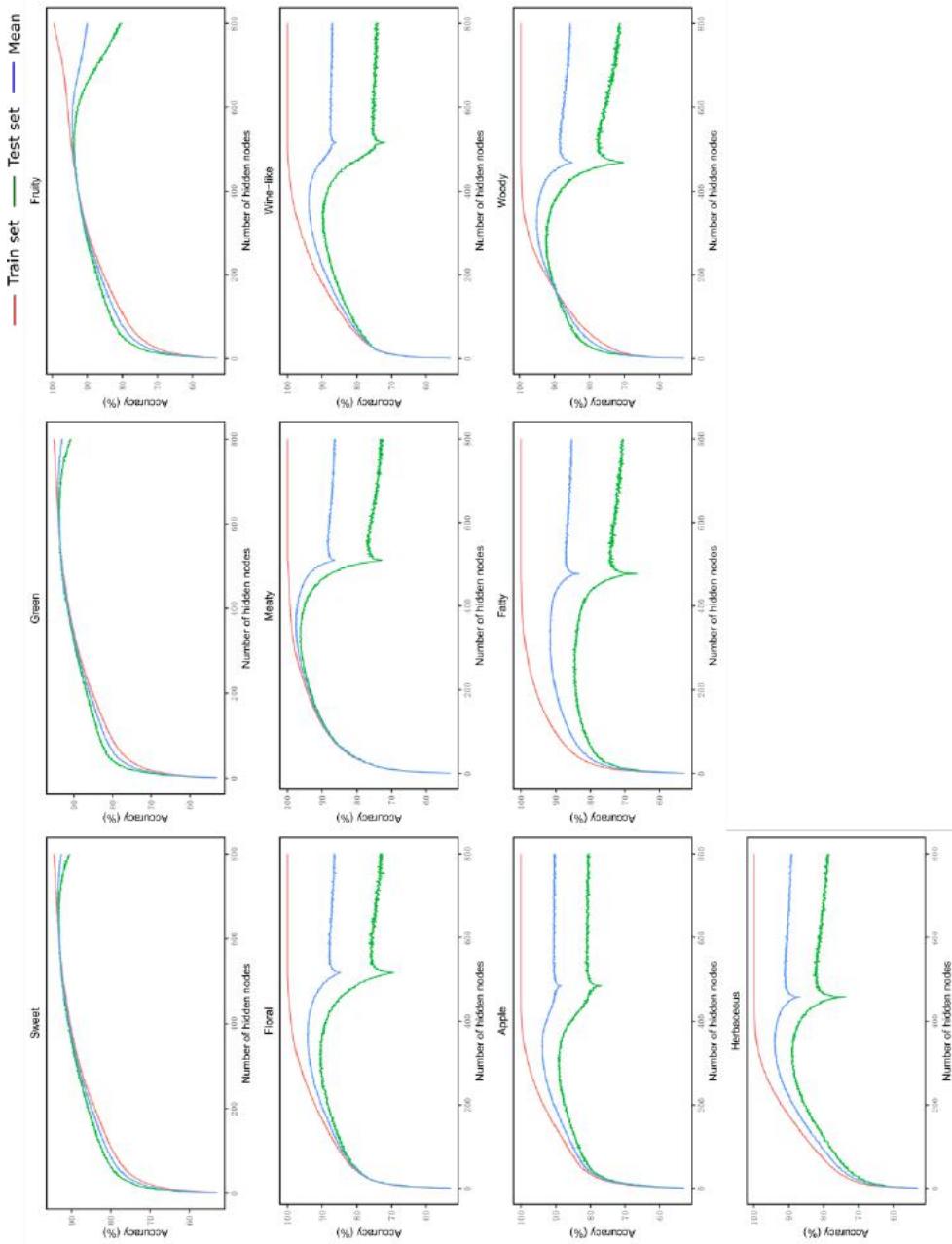


Fig. B.14 The accuracies under different numbers of hidden nodes for ELM models under all parameters.

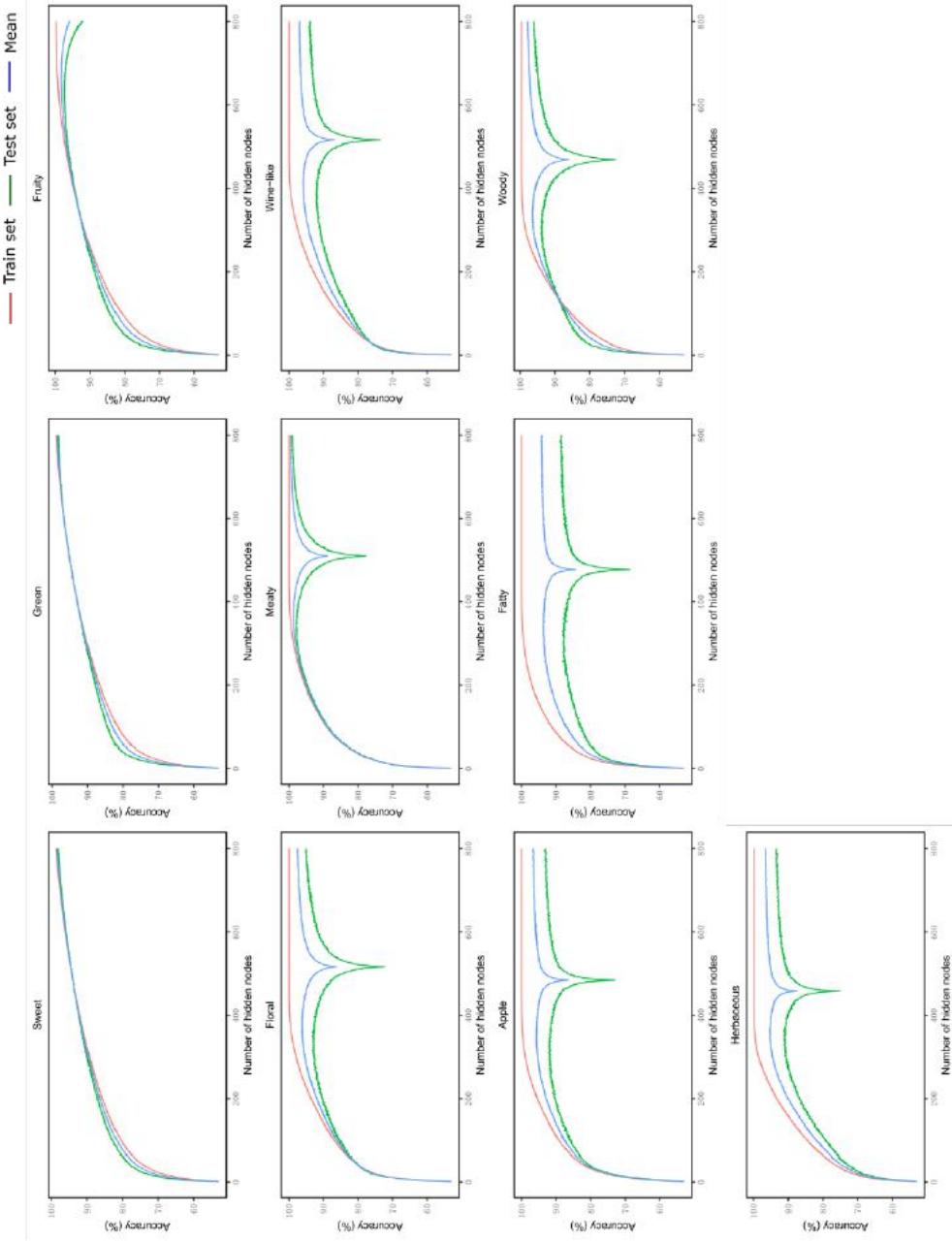


Fig. B.15 The accuracies under different numbers of hidden nodes for ELM models features obtained by PCA.

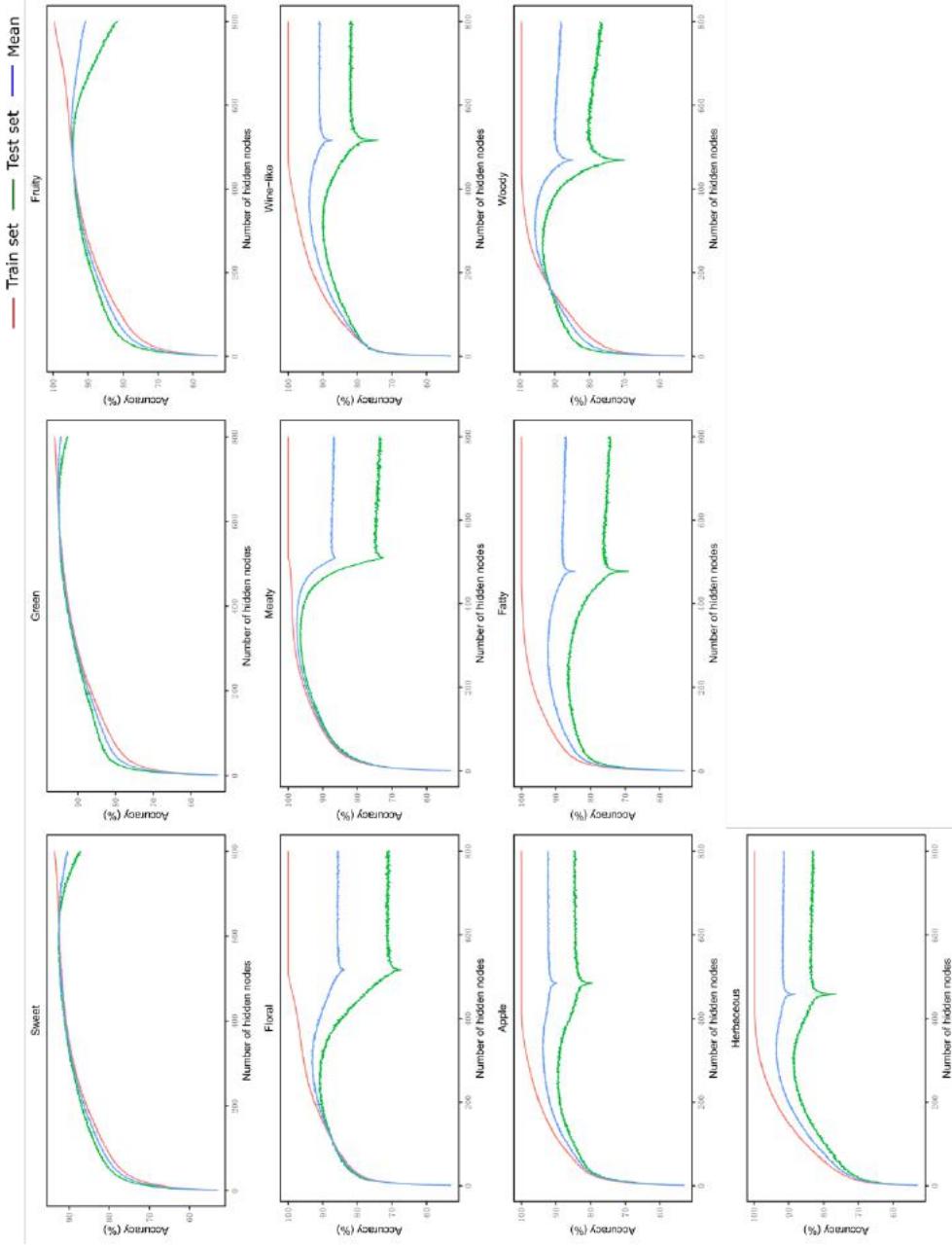


Fig. B.16 The accuracies under different numbers of hidden nodes for ELM models under features obtained by BR-CT.

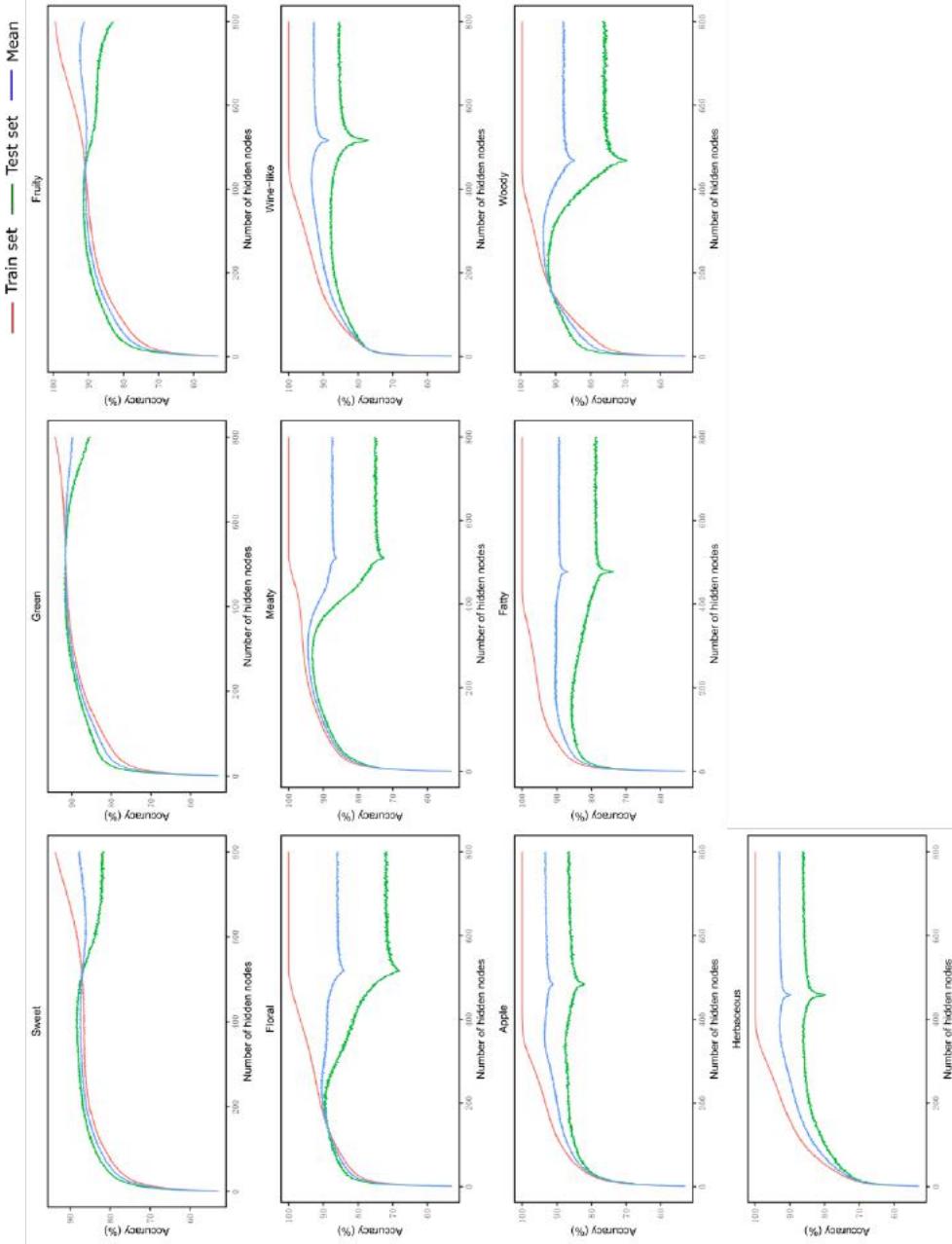


Fig. B.17 The accuracies under different numbers of hidden nodes for ELM models under features obtained by BR-C.

Table B.2 Comparison of identification results (train set) for ten OPs by SVM, RF and ELM models under AP, features extracted by PCA, BR-CT or BR-C.

Models	SVM						RF						ELM					
	Data set	AP	PCA	BR-CT	BR-C	AP	PCA	BR-CT	BR-C	AP	PCA	BR-CT	AP	PCA	BR-CT	BR-C	BR-CT	BR-C
Sweet	96.43%	97.23%	94.85%	94.12%	98.94%	85.10%	86.69%	85.34%	85.82%	93.57%	98.67%	92.37%	98.84%	95.17%	95.17%	91.29%	86.58%	
Green	97.39%	98.37%	97.39%	97.17%	99.65%	85.21%	90.20%	84.72%	85.78%	93.75%	98.84%	95.11%	94.96%	99.11%	94.96%	90.40%	90.40%	
Fruity	98.23%	98.58%	97.48%	97.48%	90.70%	82.76%	84.89%	82.88%	82.29%	94.97%	99.00%	93.59%	93.91%	97.41%	100.00%	90.97%	90.97%	
Floral	97.87%	97.87%	93.53%	98.24%	93.14%	85.49%	86.86%	87.65%	86.67%	98.47%	100.00%	98.43%	87.65%	98.47%	100.00%	95.67%	95.67%	
Meaty	93.73%	98.64%	98.84%	98.84%	99.59%	87.60%	90.29%	85.54%	85.95%	98.38%	100.00%	96.64%	85.27%	82.56%	100.00%	96.64%	96.64%	
Wine-like	98.76%	98.76%	98.76%	98.76%	98.76%	97.27%	87.63%	93.50%	87.63%	97.42%	97.70%	100.00%	96.68%	98.38%	100.00%	99.05%	99.05%	
Apple	98.95%	98.74%	98.74%	98.74%	98.73%	98.30%	98.73%	98.77%	89.38%	84.71%	86.41%	97.38%	99.58%	97.09%	97.22%	95.12%	95.12%	
Fatty	99.36%	99.78%	99.56%	99.56%	99.56%	91.70%	85.40%	82.79%	82.14%	98.76%	99.78%	98.30%	99.78%	98.30%	99.78%	93.96%	93.96%	
Woody	99.36%	99.78%	98.09%	98.08%	97.08%	95.76%	95.70%	96.30%	95.11%	96.80%	99.60%	96.05%	93.90%	99.60%	99.60%	93.90%	93.90%	
Herbaceous																		
Mean	97.91%	98.09%	98.08%	97.08%	97.66%	95.76%	95.70%	96.30%	95.11%	96.80%	99.60%	96.05%	93.90%	99.60%	99.60%	93.90%	93.90%	
SD.	1.66%	1.66%	0.97%	2.97%	2.72%	3.19%	1.84%	3.02%	1.85%	3.02%	0.50%	1.88%	3.90%	1.88%	1.88%	3.90%	3.90%	

Table B.3 Comparison of identification results (test set) for ten OPs by SVM, RF and ELM models under AP, features extracted by PCA, BR-CT or BR-C.

Models	SVM						RF						ELM					
	Data set	AP	PCA	BR-CT	BR-C	AP	PCA	BR-CT	BR-C	AP	PCA	BR-CT	AP	PCA	BR-CT	BR-C	BR-CT	BR-C
Sweet	95.49%	95.96%	95.49%	94.54%	94.54%	97.15%	99.52%	97.39%	97.86%	93.32%	98.20%	92.77%	98.84%	93.84%	98.36%	95.13%	91.95%	88.58%
Green	98.53%	99.02%	98.04%	99.51%	99.02%	99.75%	99.75%	99.75%	99.51%	99.51%	99.41%	99.13%	96.49%	96.49%	99.13%	91.58%	91.58%	
Fruity	98.23%	98.59%	97.53%	98.59%	97.88%	98.23%	97.88%	97.88%	97.88%	97.88%	97.90%	97.54%	96.51%	97.67%	90.74%	95.15%	91.24%	89.91%
Floral	98.26%	98.26%	97.67%	97.67%	90.70%	97.67%	99.42%	99.42%	96.51%	96.51%	97.67%	97.70%	99.13%	96.49%	96.49%	93.40%	93.40%	
Meaty	92.94%	92.94%	97.65%	97.65%	91.76%	97.65%	100.00%	100.00%	97.06%	97.06%	99.41%	99.41%	94.15%	94.15%	90.15%	88.14%	88.14%	
Wine-like	92.44%	92.44%	93.60%	93.60%	93.60%	92.44%	93.02%	93.60%	93.60%	93.60%	93.60%	93.60%	90.09%	90.09%	93.49%	89.76%	88.15%	
Apple	97.53%	97.53%	96.30%	96.91%	95.68%	93.83%	95.06%	96.30%	96.30%	96.30%	96.30%	96.30%	89.44%	89.44%	89.44%	88.93%	88.93%	
Fatty	90.63%	88.75%	92.50%	91.25%	94.38%	90.00%	89.38%	89.38%	89.38%	89.38%	91.46%	91.46%	91.46%	91.46%	91.46%	91.46%	91.46%	
Woody	98.10%	98.10%	97.47%	98.10%	100.00%	97.47%	98.10%	98.10%	98.10%	98.10%	92.68%	92.68%	92.68%	92.68%	92.68%	92.44%	92.44%	
Herbaceous	95.42%	95.42%	96.08%	96.73%	95.42%	95.42%	95.42%	95.42%	94.77%	94.77%	93.46%	93.46%	93.60%	93.60%	93.60%	86.47%	86.47%	
Mean	95.76%	95.70%	96.30%	95.11%	96.23%	97.36%	95.95%	96.32%	96.32%	96.32%	96.32%	96.32%	91.46%	91.46%	91.46%	91.99%	91.99%	
SD.	2.72%	3.19%	1.84%	3.02%	2.43%	2.71%	2.59%	3.05%	3.10%	2.96%	3.05%	3.05%	2.96%	2.96%	2.96%	2.89%	2.89%	2.44%

Table B.4 The list of 1006 molecular parameters in chapter3. All parameters were calculated by Dragoon 7.0.

No.	Name	Description	Block
1	MW	molecular weight	Constitutional indices
2	AMW	average molecular weight	Constitutional indices
3	Sv	sum of atomic van der Waals volumes (scaled on Carbon atom)	Constitutional indices
4	Se	sum of atomic Sanderson electronegativities (scaled on Carbon atom)	Constitutional indices
5	Sp	sum of atomic polarizabilities (scaled on Carbon atom)	Constitutional indices
6	Si	sum of first ionization potentials (scaled on Carbon atom)	Constitutional indices
7	Mv	mean atomic van der Waals volume (scaled on Carbon atom)	Constitutional indices
8	Me	mean atomic Sanderson electronegativity (scaled on Carbon atom)	Constitutional indices
9	Mp	mean atomic polarizability (scaled on Carbon atom)	Constitutional indices
10	Mi	mean first ionization potential (scaled on Carbon atom)	Constitutional indices
11	GD	graph density	Constitutional indices
12	nAT	number of atoms	Constitutional indices
13	nSK	number of non-H atoms	Constitutional indices
14	nTA	number of terminal atoms	Constitutional indices
15	nBT	number of bonds	Constitutional indices
16	nBO	number of non-H bonds	Constitutional indices
17	nBM	number of multiple bonds	Constitutional indices
18	SCBO	sum of conventional bond orders (H-depleted)	Constitutional indices
19	RBN	number of rotatable bonds	Constitutional indices
20	RBF	rotatable bond fraction	Constitutional indices
21	nDB	number of double bonds	Constitutional indices
22	nTB	number of triple bonds	Constitutional indices
23	nAB	number of aromatic bonds	Constitutional indices
24	nH	number of Hydrogen atoms	Constitutional indices
25	nC	number of Carbon atoms	Constitutional indices
26	nN	number of Nitrogen atoms	Constitutional indices
27	nO	number of Oxygen atoms	Constitutional indices
28	nP	number of Phosphorous atoms	Constitutional indices
29	nS	number of Sulfur atoms	Constitutional indices
30	nCL	number of Chlorine atoms	Constitutional indices
31	nHM	number of heavy atoms	Constitutional indices
32	nHet	number of heteroatoms	Constitutional indices
33	nX	number of halogen atoms	Constitutional indices
34	H%	percentage of H atoms	Constitutional indices
35	C%	percentage of C atoms	Constitutional indices
36	N%	percentage of N atoms	Constitutional indices
37	O%	percentage of O atoms	Constitutional indices
38	X%	percentage of halogen atoms	Constitutional indices
39	nCsp3	number of sp <sup>3</sup> hybridized Carbon atoms	Constitutional indices
40	nCsp2	number of sp <sup>2</sup> hybridized Carbon atoms	Constitutional indices
41	nCsp	number of sp hybridized Carbon atoms	Constitutional indices
42	nStructures	number of disconnected structures	Constitutional indices
43	nCIC	number of rings (cyclomatic number)	Ring descriptors
44	nCIR	number of circuits	Ring descriptors
45	TRS	total ring size	Ring descriptors
46	Rperim	ring perimeter	Ring descriptors
47	Rbrid	ring bridge count	Ring descriptors
48	MCD	molecular cyclized degree	Ring descriptors
49	RFD	ring fusion density	Ring descriptors
50	RCI	ring complexity index	Ring descriptors
51	NRS	number of ring systems	Ring descriptors
52	NNRS	normalized number of ring systems	Ring descriptors
53	nR03	number of 3-membered rings	Ring descriptors
54	nR04	number of 4-membered rings	Ring descriptors
55	nR05	number of 5-membered rings	Ring descriptors
56	nR06	number of 6-membered rings	Ring descriptors
57	nR07	number of 7-membered rings	Ring descriptors
58	nR09	number of 9-membered rings	Ring descriptors
59	nR10	number of 10-membered rings	Ring descriptors
60	nR11	number of 11-membered rings	Ring descriptors
61	nR12	number of 12-membered rings	Ring descriptors
62	nBnz	number of benzene-like rings	Ring descriptors
63	ARR	aromatic ratio	Ring descriptors
64	D/Dtr03	distance/detour ring index of order 3	Ring descriptors
65	D/Dtr04	distance/detour ring index of order 4	Ring descriptors
66	D/Dtr05	distance/detour ring index of order 5	Ring descriptors
67	D/Dtr06	distance/detour ring index of order 6	Ring descriptors
68	D/Dtr07	distance/detour ring index of order 7	Ring descriptors

No.	Name	Description	Block
69	D/Dtr09	distance/detour ring index of order 9	Ring descriptors
70	D/Dtr10	distance/detour ring index of order 10	Ring descriptors
71	D/Dtr11	distance/detour ring index of order 11	Ring descriptors
72	D/Dtr12	distance/detour ring index of order 12	Ring descriptors
73	Wap	all-path Wiener index	Topological indices
74	S1K	1-path Kier alpha-modified shape index	Topological indices
75	S2K	2-path Kier alpha-modified shape index	Topological indices
76	S3K	3-path Kier alpha-modified shape index	Topological indices
77	PHI	Kier flexibility index	Topological indices
78	PW2	path/walk 2 - Randic shape index	Topological indices
79	PW3	path/walk 3 - Randic shape index	Topological indices
80	PW4	path/walk 4 - Randic shape index	Topological indices
81	PW5	path/walk 5 - Randic shape index	Topological indices
82	MAXDN	maximal electrotopological negative variation	Topological indices
83	MAXDP	maximal electrotopological positive variation	Topological indices
84	DELS	molecular electrotopological variation	Topological indices
85	TIE	E-state topological parameter	Topological indices
86	MWC01	molecular walk count of order 1	Walk and path counts
87	MWC02	molecular walk count of order 2	Walk and path counts
88	MWC03	molecular walk count of order 3	Walk and path counts
89	MWC04	molecular walk count of order 4	Walk and path counts
90	MWC05	molecular walk count of order 5	Walk and path counts
91	MWC06	molecular walk count of order 6	Walk and path counts
92	MWC07	molecular walk count of order 7	Walk and path counts
93	MWC08	molecular walk count of order 8	Walk and path counts
94	MWC09	molecular walk count of order 9	Walk and path counts
95	MWC10	molecular walk count of order 10	Walk and path counts
96	SRW02	self-returning walk count of order 2	Walk and path counts
97	SRW03	self-returning walk count of order 3	Walk and path counts
98	SRW04	self-returning walk count of order 4	Walk and path counts
99	SRW05	self-returning walk count of order 5	Walk and path counts
100	SRW06	self-returning walk count of order 6	Walk and path counts
101	SRW07	self-returning walk count of order 7	Walk and path counts
102	SRW08	self-returning walk count of order 8	Walk and path counts
103	SRW09	self-returning walk count of order 9	Walk and path counts
104	SRW10	self-returning walk count of order 10	Walk and path counts
105	MPC01	molecular path count of order 1 (no. of non-H bonds)	Walk and path counts
106	MPC02	molecular path count of order 2 (Gordon-Scantlebury index)	Walk and path counts
107	MPC03	molecular path count of order 3	Walk and path counts
108	MPC04	molecular path count of order 4	Walk and path counts
109	MPC05	molecular path count of order 5	Walk and path counts
110	MPC06	molecular path count of order 6	Walk and path counts
111	MPC07	molecular path count of order 7	Walk and path counts
112	MPC08	molecular path count of order 8	Walk and path counts
113	MPC09	molecular path count of order 9	Walk and path counts
114	MPC10	molecular path count of order 10	Walk and path counts
115	piPC01	molecular multiple path count of order 1	Walk and path counts
116	piPC02	molecular multiple path count of order 2	Walk and path counts
117	piPC03	molecular multiple path count of order 3	Walk and path counts
118	piPC04	molecular multiple path count of order 4	Walk and path counts
119	piPC05	molecular multiple path count of order 5	Walk and path counts
120	piPC06	molecular multiple path count of order 6	Walk and path counts
121	piPC07	molecular multiple path count of order 7	Walk and path counts
122	piPC08	molecular multiple path count of order 8	Walk and path counts
123	piPC09	molecular multiple path count of order 9	Walk and path counts
124	piPC10	molecular multiple path count of order 10	Walk and path counts
125	TWC	total walk count	Walk and path counts
126	TPC	total path count	Walk and path counts
127	piID	conventional bond order ID number	Walk and path counts
128	PCD	difference between multiple path count and path count	Walk and path counts
129	CID	Randic ID number	Walk and path counts
130	BID	Balaban ID number	Walk and path counts
131	ISIZ	information index on molecular size	Information indices
132	IAC	total information index on atomic composition	Information indices
133	AAC	mean information index on atomic composition	Information indices
134	IDE	mean information content on the distance equality	Information indices
135	IDM	mean information content on the distance magnitude	Information indices
136	IDDE	mean information content on the distance degree equality	Information indices
137	IDDM	mean information content on the distance degree magnitude	Information indices
138	IDET	total information content on the distance equality	Information indices
139	IDMT	total information content on the distance magnitude	Information indices
140	IVDE	mean information content on the vertex degree equality	Information indices
141	IVDM	mean information content on the vertex degree magnitude	Information indices

No.	Name	Description	Block
142	Ges	Number of symmetry classes (based on electrotopological state)	Information indices
143	rGes	Relative number of symmetry classes (based on electrotopological state)	Information indices
144	S0K	Kier symmetry index	Information indices
145	HVcpx	graph vertex complexity index	Information indices
146	HDcpx	graph distance complexity index (log function)	Information indices
147	Uindex	Balaban U index	Information indices
148	Vindex	Balaban V index	Information indices
149	Xindex	Balaban X index	Information indices
150	Yindex	Balaban Y index	Information indices
151	IC0	Information Content index (neighborhood symmetry of 0-order)	Information indices
152	IC1	Information Content index (neighborhood symmetry of 1-order)	Information indices
153	IC2	Information Content index (neighborhood symmetry of 2-order)	Information indices
154	IC3	Information Content index (neighborhood symmetry of 3-order)	Information indices
155	IC4	Information Content index (neighborhood symmetry of 4-order)	Information indices
156	IC5	Information Content index (neighborhood symmetry of 5-order)	Information indices
157	TIC0	Total Information Content index (neighborhood symmetry of 0-order)	Information indices
158	TIC1	Total Information Content index (neighborhood symmetry of 1-order)	Information indices
159	TIC2	Total Information Content index (neighborhood symmetry of 2-order)	Information indices
160	TIC3	Total Information Content index (neighborhood symmetry of 3-order)	Information indices
161	TIC4	Total Information Content index (neighborhood symmetry of 4-order)	Information indices
162	TIC5	Total Information Content index (neighborhood symmetry of 5-order)	Information indices
163	SIC0	Structural Information Content index (neighborhood symmetry of 0-order)	Information indices
164	SIC1	Structural Information Content index (neighborhood symmetry of 1-order)	Information indices
165	SIC2	Structural Information Content index (neighborhood symmetry of 2-order)	Information indices
166	SIC3	Structural Information Content index (neighborhood symmetry of 3-order)	Information indices
167	SIC4	Structural Information Content index (neighborhood symmetry of 4-order)	Information indices
168	SIC5	Structural Information Content index (neighborhood symmetry of 5-order)	Information indices
169	CIC0	Complementary Information Content index (neighborhood symmetry of 0-order)	Information indices
170	CIC1	Complementary Information Content index (neighborhood symmetry of 1-order)	Information indices
171	CIC2	Complementary Information Content index (neighborhood symmetry of 2-order)	Information indices
172	CIC3	Complementary Information Content index (neighborhood symmetry of 3-order)	Information indices
173	CIC4	Complementary Information Content index (neighborhood symmetry of 4-order)	Information indices
174	CIC5	Complementary Information Content index (neighborhood symmetry of 5-order)	Information indices
175	BIC2	Bond Information Content index (neighborhood symmetry of 2-order)	Information indices
176	BIC3	Bond Information Content index (neighborhood symmetry of 3-order)	Information indices
177	BIC4	Bond Information Content index (neighborhood symmetry of 4-order)	Information indices
178	BIC5	Bond Information Content index (neighborhood symmetry of 5-order)	Information indices
179	ATS1m	Broto-Moreau autocorrelation of lag 1 (log function) weighted by mass	2D autocorrelations
180	ATS2m	Broto-Moreau autocorrelation of lag 2 (log function) weighted by mass	2D autocorrelations
181	ATS3m	Broto-Moreau autocorrelation of lag 3 (log function) weighted by mass	2D autocorrelations
182	ATS4m	Broto-Moreau autocorrelation of lag 4 (log function) weighted by mass	2D autocorrelations
183	ATS5m	Broto-Moreau autocorrelation of lag 5 (log function) weighted by mass	2D autocorrelations
184	ATS6m	Broto-Moreau autocorrelation of lag 6 (log function) weighted by mass	2D autocorrelations
185	ATS7m	Broto-Moreau autocorrelation of lag 7 (log function) weighted by mass	2D autocorrelations
186	ATS8m	Broto-Moreau autocorrelation of lag 8 (log function) weighted by mass	2D autocorrelations
187	ATS1v	Broto-Moreau autocorrelation of lag 1 (log function) weighted by van der Waals volume	2D autocorrelations
188	ATS2v	Broto-Moreau autocorrelation of lag 2 (log function) weighted by van der Waals volume	2D autocorrelations
189	ATS3v	Broto-Moreau autocorrelation of lag 3 (log function) weighted by van der Waals volume	2D autocorrelations
190	ATS4v	Broto-Moreau autocorrelation of lag 4 (log function) weighted by van der Waals volume	2D autocorrelations
191	ATS5v	Broto-Moreau autocorrelation of lag 5 (log function) weighted by van der Waals volume	2D autocorrelations
192	ATS6v	Broto-Moreau autocorrelation of lag 6 (log function) weighted by van der Waals volume	2D autocorrelations
193	ATS7v	Broto-Moreau autocorrelation of lag 7 (log function) weighted by van der Waals volume	2D autocorrelations
194	ATS8v	Broto-Moreau autocorrelation of lag 8 (log function) weighted by van der Waals volume	2D autocorrelations
195	ATS1e	Broto-Moreau autocorrelation of lag 1 (log function) weighted by Sanderson electronegativity	2D autocorrelations
196	ATS2e	Broto-Moreau autocorrelation of lag 2 (log function) weighted by Sanderson electronegativity	2D autocorrelations
197	ATS3e	Broto-Moreau autocorrelation of lag 3 (log function) weighted by Sanderson electronegativity	2D autocorrelations
198	ATS4e	Broto-Moreau autocorrelation of lag 4 (log function) weighted by Sanderson electronegativity	2D autocorrelations
199	ATS5e	Broto-Moreau autocorrelation of lag 5 (log function) weighted by Sanderson electronegativity	2D autocorrelations
200	ATS6e	Broto-Moreau autocorrelation of lag 6 (log function) weighted by Sanderson electronegativity	2D autocorrelations
201	ATS7e	Broto-Moreau autocorrelation of lag 7 (log function) weighted by Sanderson electronegativity	2D autocorrelations
202	ATS8e	Broto-Moreau autocorrelation of lag 8 (log function) weighted by Sanderson electronegativity	2D autocorrelations
203	ATS1p	Broto-Moreau autocorrelation of lag 1 (log function) weighted by polarizability	2D autocorrelations
204	ATS2p	Broto-Moreau autocorrelation of lag 2 (log function) weighted by polarizability	2D autocorrelations
205	ATS3p	Broto-Moreau autocorrelation of lag 3 (log function) weighted by polarizability	2D autocorrelations
206	ATS4p	Broto-Moreau autocorrelation of lag 4 (log function) weighted by polarizability	2D autocorrelations
207	ATS5p	Broto-Moreau autocorrelation of lag 5 (log function) weighted by polarizability	2D autocorrelations
208	ATS6p	Broto-Moreau autocorrelation of lag 6 (log function) weighted by polarizability	2D autocorrelations
209	ATS7p	Broto-Moreau autocorrelation of lag 7 (log function) weighted by polarizability	2D autocorrelations
210	ATS8p	Broto-Moreau autocorrelation of lag 8 (log function) weighted by polarizability	2D autocorrelations
211	ATS1i	Broto-Moreau autocorrelation of lag 1 (log function) weighted by ionization potential	2D autocorrelations
212	ATS2i	Broto-Moreau autocorrelation of lag 2 (log function) weighted by ionization potential	2D autocorrelations
213	ATS3i	Broto-Moreau autocorrelation of lag 3 (log function) weighted by ionization potential	2D autocorrelations
214	ATS4i	Broto-Moreau autocorrelation of lag 4 (log function) weighted by ionization potential	2D autocorrelations



No.	Name	Description	Block
288	MATS6p	Moran autocorrelation of lag 6 weighted by polarizability	2D autocorrelations
289	MATS7p	Moran autocorrelation of lag 7 weighted by polarizability	2D autocorrelations
290	MATS8p	Moran autocorrelation of lag 8 weighted by polarizability	2D autocorrelations
291	MATS1i	Moran autocorrelation of lag 1 weighted by ionization potential	2D autocorrelations
292	MATS2i	Moran autocorrelation of lag 2 weighted by ionization potential	2D autocorrelations
293	MATS3i	Moran autocorrelation of lag 3 weighted by ionization potential	2D autocorrelations
294	MATS4i	Moran autocorrelation of lag 4 weighted by ionization potential	2D autocorrelations
295	MATS5i	Moran autocorrelation of lag 5 weighted by ionization potential	2D autocorrelations
296	MATS6i	Moran autocorrelation of lag 6 weighted by ionization potential	2D autocorrelations
297	MATS7i	Moran autocorrelation of lag 7 weighted by ionization potential	2D autocorrelations
298	MATS8i	Moran autocorrelation of lag 8 weighted by ionization potential	2D autocorrelations
299	GATS1m	Geary autocorrelation of lag 1 weighted by mass	2D autocorrelations
300	GATS2m	Geary autocorrelation of lag 2 weighted by mass	2D autocorrelations
301	GATS3m	Geary autocorrelation of lag 3 weighted by mass	2D autocorrelations
302	GATS4m	Geary autocorrelation of lag 4 weighted by mass	2D autocorrelations
303	GATS5m	Geary autocorrelation of lag 5 weighted by mass	2D autocorrelations
304	GATS6m	Geary autocorrelation of lag 6 weighted by mass	2D autocorrelations
305	GATS7m	Geary autocorrelation of lag 7 weighted by mass	2D autocorrelations
306	GATS8m	Geary autocorrelation of lag 8 weighted by mass	2D autocorrelations
307	GATS1v	Geary autocorrelation of lag 1 weighted by van der Waals volume	2D autocorrelations
308	GATS2v	Geary autocorrelation of lag 2 weighted by van der Waals volume	2D autocorrelations
309	GATS3v	Geary autocorrelation of lag 3 weighted by van der Waals volume	2D autocorrelations
310	GATS4v	Geary autocorrelation of lag 4 weighted by van der Waals volume	2D autocorrelations
311	GATS5v	Geary autocorrelation of lag 5 weighted by van der Waals volume	2D autocorrelations
312	GATS6v	Geary autocorrelation of lag 6 weighted by van der Waals volume	2D autocorrelations
313	GATS7v	Geary autocorrelation of lag 7 weighted by van der Waals volume	2D autocorrelations
314	GATS8v	Geary autocorrelation of lag 8 weighted by van der Waals volume	2D autocorrelations
315	GATS1e	Geary autocorrelation of lag 1 weighted by Sanderson electronegativity	2D autocorrelations
316	GATS2e	Geary autocorrelation of lag 2 weighted by Sanderson electronegativity	2D autocorrelations
317	GATS3e	Geary autocorrelation of lag 3 weighted by Sanderson electronegativity	2D autocorrelations
318	GATS4e	Geary autocorrelation of lag 4 weighted by Sanderson electronegativity	2D autocorrelations
319	GATS5e	Geary autocorrelation of lag 5 weighted by Sanderson electronegativity	2D autocorrelations
320	GATS6e	Geary autocorrelation of lag 6 weighted by Sanderson electronegativity	2D autocorrelations
321	GATS7e	Geary autocorrelation of lag 7 weighted by Sanderson electronegativity	2D autocorrelations
322	GATS8e	Geary autocorrelation of lag 8 weighted by Sanderson electronegativity	2D autocorrelations
323	GATS1p	Geary autocorrelation of lag 1 weighted by polarizability	2D autocorrelations
324	GATS2p	Geary autocorrelation of lag 2 weighted by polarizability	2D autocorrelations
325	GATS3p	Geary autocorrelation of lag 3 weighted by polarizability	2D autocorrelations
326	GATS4p	Geary autocorrelation of lag 4 weighted by polarizability	2D autocorrelations
327	GATS5p	Geary autocorrelation of lag 5 weighted by polarizability	2D autocorrelations
328	GATS6p	Geary autocorrelation of lag 6 weighted by polarizability	2D autocorrelations
329	GATS7p	Geary autocorrelation of lag 7 weighted by polarizability	2D autocorrelations
330	GATS8p	Geary autocorrelation of lag 8 weighted by polarizability	2D autocorrelations
331	GATS1i	Geary autocorrelation of lag 1 weighted by ionization potential	2D autocorrelations
332	GATS2i	Geary autocorrelation of lag 2 weighted by ionization potential	2D autocorrelations
333	GATS3i	Geary autocorrelation of lag 3 weighted by ionization potential	2D autocorrelations
334	GATS4i	Geary autocorrelation of lag 4 weighted by ionization potential	2D autocorrelations
335	GATS5i	Geary autocorrelation of lag 5 weighted by ionization potential	2D autocorrelations
336	GATS6i	Geary autocorrelation of lag 6 weighted by ionization potential	2D autocorrelations
337	GATS7i	Geary autocorrelation of lag 7 weighted by ionization potential	2D autocorrelations
338	GATS8i	Geary autocorrelation of lag 8 weighted by ionization potential	2D autocorrelations
339	GGI1	topological charge index of order 1	2D autocorrelations
340	GGI2	topological charge index of order 2	2D autocorrelations
341	GGI3	topological charge index of order 3	2D autocorrelations
342	GGI4	topological charge index of order 4	2D autocorrelations
343	GGI5	topological charge index of order 5	2D autocorrelations
344	GGI6	topological charge index of order 6	2D autocorrelations
345	GGI7	topological charge index of order 7	2D autocorrelations
346	GGI8	topological charge index of order 8	2D autocorrelations
347	GGI9	topological charge index of order 9	2D autocorrelations
348	GGI10	topological charge index of order 10	2D autocorrelations
349	JGI1	mean topological charge index of order 1	2D autocorrelations
350	JGI2	mean topological charge index of order 2	2D autocorrelations
351	JGI3	mean topological charge index of order 3	2D autocorrelations
352	JGI4	mean topological charge index of order 4	2D autocorrelations
353	JGI5	mean topological charge index of order 5	2D autocorrelations
354	JGI6	mean topological charge index of order 6	2D autocorrelations
355	JGI7	mean topological charge index of order 7	2D autocorrelations
356	JGI8	mean topological charge index of order 8	2D autocorrelations
357	JGI9	mean topological charge index of order 9	2D autocorrelations
358	JGI10	mean topological charge index of order 10	2D autocorrelations
359	JGT	global topological charge index	2D autocorrelations
360	SpMax1_Bh(m)	largest eigenvalue n. 1 of Burden matrix weighted by mass	Burden eigenvalues



No.	Name	Description	Block
434	SpMin3_Bh(i)	smallest eigenvalue n. 3 of Burden matrix weighted by ionization potential	Burden eigenvalues
435	SpMin4_Bh(i)	smallest eigenvalue n. 4 of Burden matrix weighted by ionization potential	Burden eigenvalues
436	SpMin5_Bh(i)	smallest eigenvalue n. 5 of Burden matrix weighted by ionization potential	Burden eigenvalues
437	SpMin6_Bh(i)	smallest eigenvalue n. 6 of Burden matrix weighted by ionization potential	Burden eigenvalues
438	SpMin7_Bh(i)	smallest eigenvalue n. 7 of Burden matrix weighted by ionization potential	Burden eigenvalues
439	SpMin8_Bh(i)	smallest eigenvalue n. 8 of Burden matrix weighted by ionization potential	Burden eigenvalues
440	P_VSA_LogP_1	P_VSA-like on LogP, bin 1	P_VSA-like descriptors
441	P_VSA_LogP_2	P_VSA-like on LogP, bin 2	P_VSA-like descriptors
442	P_VSA_LogP_3	P_VSA-like on LogP, bin 3	P_VSA-like descriptors
443	P_VSA_LogP_4	P_VSA-like on LogP, bin 4	P_VSA-like descriptors
444	P_VSA_LogP_5	P_VSA-like on LogP, bin 5	P_VSA-like descriptors
445	P_VSA_LogP_6	P_VSA-like on LogP, bin 6	P_VSA-like descriptors
446	P_VSA_LogP_7	P_VSA-like on LogP, bin 7	P_VSA-like descriptors
447	P_VSA_LogP_8	P_VSA-like on LogP, bin 8	P_VSA-like descriptors
448	P_VSA_MR_1	P_VSA-like on Molar Refractivity, bin 1	P_VSA-like descriptors
449	P_VSA_MR_2	P_VSA-like on Molar Refractivity, bin 2	P_VSA-like descriptors
450	P_VSA_MR_3	P_VSA-like on Molar Refractivity, bin 3	P_VSA-like descriptors
451	P_VSA_MR_4	P_VSA-like on Molar Refractivity, bin 4	P_VSA-like descriptors
452	P_VSA_MR_5	P_VSA-like on Molar Refractivity, bin 5	P_VSA-like descriptors
453	P_VSA_MR_6	P_VSA-like on Molar Refractivity, bin 6	P_VSA-like descriptors
454	P_VSA_MR_7	P_VSA-like on Molar Refractivity, bin 7	P_VSA-like descriptors
455	P_VSA_MR_8	P_VSA-like on Molar Refractivity, bin 8	P_VSA-like descriptors
456	P_VSA_m_1	P_VSA-like on mass, bin 1	P_VSA-like descriptors
457	P_VSA_m_2	P_VSA-like on mass, bin 2	P_VSA-like descriptors
458	P_VSA_m_3	P_VSA-like on mass, bin 3	P_VSA-like descriptors
459	P_VSA_m_4	P_VSA-like on mass, bin 4	P_VSA-like descriptors
460	P_VSA_v_1	P_VSA-like on van der Waals volume, bin 1	P_VSA-like descriptors
461	P_VSA_v_2	P_VSA-like on van der Waals volume, bin 2	P_VSA-like descriptors
462	P_VSA_v_3	P_VSA-like on van der Waals volume, bin 3	P_VSA-like descriptors
463	P_VSA_e_1	P_VSA-like on Sanderson electronegativity, bin 1	P_VSA-like descriptors
464	P_VSA_e_2	P_VSA-like on Sanderson electronegativity, bin 2	P_VSA-like descriptors
465	P_VSA_e_3	P_VSA-like on Sanderson electronegativity, bin 3	P_VSA-like descriptors
466	P_VSA_e_4	P_VSA-like on Sanderson electronegativity, bin 4	P_VSA-like descriptors
467	P_VSA_e_5	P_VSA-like on Sanderson electronegativity, bin 5	P_VSA-like descriptors
468	P_VSA_p_1	P_VSA-like on polarizability, bin 1	P_VSA-like descriptors
469	P_VSA_p_2	P_VSA-like on polarizability, bin 2	P_VSA-like descriptors
470	P_VSA_p_3	P_VSA-like on polarizability, bin 3	P_VSA-like descriptors
471	P_VSA_p_4	P_VSA-like on polarizability, bin 4	P_VSA-like descriptors
472	P_VSA_i_1	P_VSA-like on ionization potential, bin 1	P_VSA-like descriptors
473	P_VSA_i_2	P_VSA-like on ionization potential, bin 2	P_VSA-like descriptors
474	P_VSA_i_3	P_VSA-like on ionization potential, bin 3	P_VSA-like descriptors
475	P_VSA_i_4	P_VSA-like on ionization potential, bin 4	P_VSA-like descriptors
476	P_VSA_s_1	P_VSA-like on I-state, bin 1	P_VSA-like descriptors
477	P_VSA_s_2	P_VSA-like on I-state, bin 2	P_VSA-like descriptors
478	P_VSA_s_3	P_VSA-like on I-state, bin 3	P_VSA-like descriptors
479	P_VSA_s_4	P_VSA-like on I-state, bin 4	P_VSA-like descriptors
480	P_VSA_s_5	P_VSA-like on I-state, bin 5	P_VSA-like descriptors
481	P_VSA_s_6	P_VSA-like on I-state, bin 6	P_VSA-like descriptors
482	P_VSA_ppp_L	P_VSA-like on potential pharmacophore points, L - lipophilic	P_VSA-like descriptors
483	P_VSA_ppp_P	P_VSA-like on potential pharmacophore points, P - positive	P_VSA-like descriptors
484	P_VSA_ppp_N	P_VSA-like on potential pharmacophore points, N - negative	P_VSA-like descriptors
485	P_VSA_ppp_D	P_VSA-like on potential pharmacophore points, D - hydrogen-bond donor	P_VSA-like descriptors
486	P_VSA_ppp_A	P_VSA-like on potential pharmacophore points, A - hydrogen-bond acceptor	P_VSA-like descriptors
487	P_VSA_ppp_ar	P_VSA-like on potential pharmacophore points, ar - aromatic atoms	P_VSA-like descriptors
488	P_VSA_ppp_con	P_VSA-like on potential pharmacophore points, con - conjugated atoms	P_VSA-like descriptors
489	P_VSA_ppp_hal	P_VSA-like on potential pharmacophore points, hal - halogen atoms	P_VSA-like descriptors
490	P_VSA_ppp_cyc	P_VSA-like on potential pharmacophore points, cyc - atoms belonging to cycles	P_VSA-like descriptors
491	P_VSA_ppp_ter	P_VSA-like on potential pharmacophore points, ter - terminal atoms	P_VSA-like descriptors
492	nCp	number of terminal primary C(sp3)	Functional group counts
493	nCs	number of total secondary C(sp3)	Functional group counts
494	nCt	number of total tertiary C(sp3)	Functional group counts
495	nCq	number of total quaternary C(sp3)	Functional group counts
496	nCrs	number of ring secondary C(sp3)	Functional group counts
497	nCrt	number of ring tertiary C(sp3)	Functional group counts
498	nCrq	number of ring quaternary C(sp3)	Functional group counts
499	nCar	number of aromatic C(sp2)	Functional group counts
500	nCbH	number of unsubstituted benzene C(sp2)	Functional group counts
501	nCb-	number of substituted benzene C(sp2)	Functional group counts
502	nConj	number of non-aromatic conjugated C(sp2)	Functional group counts
503	nR=Cp	number of terminal primary C(sp2)	Functional group counts
504	nR=Cs	number of aliphatic secondary C(sp2)	Functional group counts
505	nR=Ct	number of aliphatic tertiary C(sp2)	Functional group counts
506	nR		Functional group counts

No.	Name	Description	Block
507	nRNCS	number of isothiocyanates (aliphatic)	Functional group counts
508	nRCOOH	number of carboxylic acids (aliphatic)	Functional group counts
509	nArCOOH	number of carboxylic acids (aromatic)	Functional group counts
510	nRCOOR	number of esters (aliphatic)	Functional group counts
511	nArCOOR	number of esters (aromatic)	Functional group counts
512	nRCONH2	number of primary amides (aliphatic)	Functional group counts
513	nRCONHR	number of secondary amides (aliphatic)	Functional group counts
514	nRCONR2	number of tertiary amides (aliphatic)	Functional group counts
515	nRCOSH	number of thioacids (aliphatic)	Functional group counts
516	nRCOSR	number of thioesters (aliphatic)	Functional group counts
517	nArCOSR	number of thioesters (aromatic)	Functional group counts
518	nRCHO	number of aldehydes (aliphatic)	Functional group counts
519	nArCHO	number of aldehydes (aromatic)	Functional group counts
520	nRCO	number of ketones (aliphatic)	Functional group counts
521	nArCO	number of ketones (aromatic)	Functional group counts
522	nCONN	number of urea (-thio) derivatives	Functional group counts
523	nC(=N)N2	number of guanidine derivatives	Functional group counts
524	nRC=N	number of imines (aliphatic)	Functional group counts
525	nRNH2	number of primary amines (aliphatic)	Functional group counts
526	nArNH2	number of primary amines (aromatic)	Functional group counts
527	nRNHR	number of secondary amines (aliphatic)	Functional group counts
528	nArNHR	number of secondary amines (aromatic)	Functional group counts
529	nRNR2	number of tertiary amines (aliphatic)	Functional group counts
530	nN+	number of positively charged N	Functional group counts
531	nArNO2	number of nitro groups (aromatic)	Functional group counts
532	nN(CO)2	number of imides (-thio)	Functional group counts
533	nROH	number of hydroxyl groups	Functional group counts
534	nArOH	number of aromatic hydroxyls	Functional group counts
535	nOHP	number of primary alcohols	Functional group counts
536	nOHs	number of secondary alcohols	Functional group counts
537	nOHT	number of tertiary alcohols	Functional group counts
538	nROR	number of ethers (aliphatic)	Functional group counts
539	nArOR	number of ethers (aromatic)	Functional group counts
540	nO(C=O)2	number of anhydrides (-thio)	Functional group counts
541	nH2O	number of water molecules	Functional group counts
542	nSH	number of thiols	Functional group counts
543	nRSR	number of sulfides	Functional group counts
544	nRSSR	number of disulfides	Functional group counts
545	nSO	number of sulfoxides	Functional group counts
546	nSO2OH	number of sulfonic (thio-/dithio-) acids	Functional group counts
547	nPO4	number of phosphates/thiophosphates	Functional group counts
548	nOxiranes	number of Oxiranes	Functional group counts
549	nPyrrolidines	number of Pyrrolidines	Functional group counts
550	nOxolanes	number of Oxolanes	Functional group counts
551	nPyrroles	number of Pyrroles	Functional group counts
552	nImidazoles	number of Imidazoles	Functional group counts
553	nFuranes	number of Furanes	Functional group counts
554	nThiophenes	number of Thiophenes	Functional group counts
555	nOxazoles	number of Oxazoles	Functional group counts
556	nThiazoles	number of Thiazoles	Functional group counts
557	nPyridines	number of Pyridines	Functional group counts
558	nPyrimidines	number of Pyrimidines	Functional group counts
559	nPyrazines	number of Pyrazines	Functional group counts
560	nHDon	number of donor atoms for H-bonds (N and O)	Functional group counts
561	nHAcc	number of acceptor atoms for H-bonds (N,O,F)	Functional group counts
562	C-001	CH3R / CH4	Atom-centred fragments
563	C-002	CH2R2	Atom-centred fragments
564	C-003	CHR3	Atom-centred fragments
565	C-004	CR4	Atom-centred fragments
566	C-005	CH3X	Atom-centred fragments
567	C-006	CH2RX	Atom-centred fragments
568	C-007	CH2X2	Atom-centred fragments
569	C-008	CHR2X	Atom-centred fragments
570	C-009	CHRX2	Atom-centred fragments
571	C-010	CHX3	Atom-centred fragments
572	C-011	CR3X	Atom-centred fragments
573	C-012	CR2X2	Atom-centred fragments
574	C-015	=CH2	Atom-centred fragments
575	C-016	=CHR	Atom-centred fragments
576	C-017	=CR2	Atom-centred fragments
577	C-018	=CHX	Atom-centred fragments
578	C-019	=CRX	Atom-centred fragments
579	C-022		Atom-centred fragments

No.	Name	Description	Block
580	C-024	R-CH-R	Atom-centred fragments
581	C-025	R-CR-R	Atom-centred fragments
582	C-026	R-CX-R	Atom-centred fragments
583	C-027	R-CH-X	Atom-centred fragments
584	C-028	R-CR-X	Atom-centred fragments
585	C-029	R-CX-X	Atom-centred fragments
586	C-031	X-CR-X	Atom-centred fragments
587	C-033	R-CH..X	Atom-centred fragments
588	C-034	R-CR..X	Atom-centred fragments
589	C-035	R-CX..X	Atom-centred fragments
590	C-036	Al-CH=X	Atom-centred fragments
591	C-037	Ar-CH=X	Atom-centred fragments
592	C-038	Al-C(=X)-Al	Atom-centred fragments
593	C-039	Ar-C(=X)-R	Atom-centred fragments
594	C-040	R-C(=X)-X / R-C	Atom-centred fragments
595	C-041	X-C(=X)-X	Atom-centred fragments
596	C-042	X-CH..X	Atom-centred fragments
597	C-043	X-CR..X	Atom-centred fragments
598	C-044	X-CX..X	Atom-centred fragments
599	H-046	H attached to C0(sp3) no X attached to next C	Atom-centred fragments
600	H-047	H attached to C1(sp3)/C0(sp2)	Atom-centred fragments
601	H-048	H attached to C2(sp3)/C1(sp2)/C0(sp)	Atom-centred fragments
602	H-049	H attached to C3(sp3)/C2(sp2)/C3(sp2)/C3(sp)	Atom-centred fragments
603	H-050	H attached to heteroatom	Atom-centred fragments
604	H-051	H attached to alpha-C	Atom-centred fragments
605	H-052	H attached to C0(sp3) with 1X attached to next C	Atom-centred fragments
606	H-053	H attached to C0(sp3) with 2X attached to next C	Atom-centred fragments
607	H-054	H attached to C0(sp3) with 3X attached to next C	Atom-centred fragments
608	O-056	alcohol	Atom-centred fragments
609	O-057	phenol / enol / carboxyl OH	Atom-centred fragments
610	O-058	#NAME?	Atom-centred fragments
611	O-059	Al-O-Al	Atom-centred fragments
612	O-060	Al-O-Ar / Ar-O-Ar / R..O..R / R-O-C=X	Atom-centred fragments
613	O-061	O-	Atom-centred fragments
614	O-062	O- (negatively charged)	Atom-centred fragments
615	N-066	Al-NH2	Atom-centred fragments
616	N-067	Al2-NH	Atom-centred fragments
617	N-068	Al3-N	Atom-centred fragments
618	N-069	Ar-NH2 / X-NH2	Atom-centred fragments
619	N-070	Ar-NH-Al	Atom-centred fragments
620	N-072	RCO-N</>N-X=X	Atom-centred fragments
621	N-073	Ar2NH / Ar3N / Ar2N-Al / R..N..R	Atom-centred fragments
622	N-074	R	Atom-centred fragments
623	N-075	R-N-R / R-N-X	Atom-centred fragments
624	N-076	Ar-NO2 / R-N(-R)-O / RO-NO	Atom-centred fragments
625	Cl-090	Cl attached to C2(sp2)-C4(sp2)/C1(sp)/C4(sp3)/X	Atom-centred fragments
626	Cl-102	chloride ion	Atom-centred fragments
627	S-106	R-SH	Atom-centred fragments
628	S-107	R2S / RS-SR	Atom-centred fragments
629	S-108	R=S	Atom-centred fragments
630	S-109	R-SO-R	Atom-centred fragments
631	S-110	R-SO2-R	Atom-centred fragments
632	P-117	X3-P=X (phosphate)	Atom-centred fragments
633	SsCH3	Sum of sCH3 E-states	Atom-type E-state indices
634	SdCH2	Sum of dCH2 E-states	Atom-type E-state indices
635	SssCH2	Sum of ssCH2 E-states	Atom-type E-state indices
636	SdsCH	Sum of dsCH E-states	Atom-type E-state indices
637	SaaCH	Sum of aaCH E-states	Atom-type E-state indices
638	SsssCH	Sum of sssCH E-states	Atom-type E-state indices
639	SddC	Sum of ddC E-states	Atom-type E-state indices
640	StsC	Sum of tsC E-states	Atom-type E-state indices
641	SdssC	Sum of dssC E-states	Atom-type E-state indices
642	SaasC	Sum of aasC E-states	Atom-type E-state indices
643	SaaaC	Sum of aaaC E-states	Atom-type E-state indices
644	SssscC	Sum of ssssc E-states	Atom-type E-state indices
645	SsNH2	Sum of sNH2 E-states	Atom-type E-state indices
646	SssNH	Sum of ssNH E-states	Atom-type E-state indices
647	SsssN	Sum of sssN E-states	Atom-type E-state indices
648	SdsN	Sum of dsN E-states	Atom-type E-state indices
649	SaaN	Sum of aaN E-states	Atom-type E-state indices
650	SddsN	Sum of ddsN E-states	Atom-type E-state indices
651	SaasN	Sum of aasN E-states	Atom-type E-state indices
652	SaaNH	Sum of aaNH E-states	Atom-type E-state indices

No.	Name	Description	Block
653	SsOH	Sum of sOH E-states	Atom-type E-state indices
654	SdO	Sum of dO E-states	Atom-type E-state indices
655	SssO	Sum of ssO E-states	Atom-type E-state indices
656	SaaO	Sum of aaO E-states	Atom-type E-state indices
657	SdsssP	Sum of dsssP E-states	Atom-type E-state indices
658	SsSH	Sum of sSH E-states	Atom-type E-state indices
659	SdS	Sum of dS E-states	Atom-type E-state indices
660	SssS	Sum of ssS E-states	Atom-type E-state indices
661	SaaS	Sum of aaS E-states	Atom-type E-state indices
662	SdssS	Sum of dssS E-states	Atom-type E-state indices
663	SddssS	Sum of ddssS E-states	Atom-type E-state indices
664	NsCH3	Number of atoms of type sCH3	Atom-type E-state indices
665	NdCH2	Number of atoms of type dCH2	Atom-type E-state indices
666	NssCH2	Number of atoms of type ssCH2	Atom-type E-state indices
667	NdsCH	Number of atoms of type dsCH	Atom-type E-state indices
668	NaaCH	Number of atoms of type aaCH	Atom-type E-state indices
669	NsssCH	Number of atoms of type sssCH	Atom-type E-state indices
670	NddC	Number of atoms of type ddC	Atom-type E-state indices
671	NtsC	Number of atoms of type tsC	Atom-type E-state indices
672	NdssC	Number of atoms of type dssC	Atom-type E-state indices
673	NaasC	Number of atoms of type aasC	Atom-type E-state indices
674	NaaaC	Number of atoms of type aaaC	Atom-type E-state indices
675	NssssC	Number of atoms of type ssssC	Atom-type E-state indices
676	NsNH2	Number of atoms of type sNH2	Atom-type E-state indices
677	NssNH	Number of atoms of type ssNH	Atom-type E-state indices
678	NsssN	Number of atoms of type sssN	Atom-type E-state indices
679	NdsN	Number of atoms of type dN	Atom-type E-state indices
680	NaaN	Number of atoms of type aaN	Atom-type E-state indices
681	NddsN	Number of atoms of type ddsN	Atom-type E-state indices
682	NaasN	Number of atoms of type aasN	Atom-type E-state indices
683	NaaNH	Number of atoms of type aaNH	Atom-type E-state indices
684	NsOH	Number of atoms of type sOH	Atom-type E-state indices
685	NdO	Number of atoms of type dO	Atom-type E-state indices
686	NssO	Number of atoms of type ssO	Atom-type E-state indices
687	NaaO	Number of atoms of type aaO	Atom-type E-state indices
688	NdsssP	Number of atoms of type dsssP	Atom-type E-state indices
689	NsSH	Number of atoms of type sSH	Atom-type E-state indices
690	NdS	Number of atoms of type dS	Atom-type E-state indices
691	NssS	Number of atoms of type ssS	Atom-type E-state indices
692	NaaS	Number of atoms of type aaS	Atom-type E-state indices
693	NdssS	Number of atoms of type dssS	Atom-type E-state indices
694	NddssS	Number of atoms of type ddssS	Atom-type E-state indices
695	CATS2D_00_DD	CATS2D Donor-Donor at lag 00	CATS 2D
696	CATS2D_02_DD	CATS2D Donor-Donor at lag 02	CATS 2D
697	CATS2D_03_DD	CATS2D Donor-Donor at lag 03	CATS 2D
698	CATS2D_04_DD	CATS2D Donor-Donor at lag 04	CATS 2D
699	CATS2D_05_DD	CATS2D Donor-Donor at lag 05	CATS 2D
700	CATS2D_06_DD	CATS2D Donor-Donor at lag 06	CATS 2D
701	CATS2D_07_DD	CATS2D Donor-Donor at lag 07	CATS 2D
702	CATS2D_08_DD	CATS2D Donor-Donor at lag 08	CATS 2D
703	CATS2D_09_DD	CATS2D Donor-Donor at lag 09	CATS 2D
704	CATS2D_00_DA	CATS2D Donor-Acceptor at lag 00	CATS 2D
705	CATS2D_02_DA	CATS2D Donor-Acceptor at lag 02	CATS 2D
706	CATS2D_03_DA	CATS2D Donor-Acceptor at lag 03	CATS 2D
707	CATS2D_04_DA	CATS2D Donor-Acceptor at lag 04	CATS 2D
708	CATS2D_05_DA	CATS2D Donor-Acceptor at lag 05	CATS 2D
709	CATS2D_06_DA	CATS2D Donor-Acceptor at lag 06	CATS 2D
710	CATS2D_07_DA	CATS2D Donor-Acceptor at lag 07	CATS 2D
711	CATS2D_08_DA	CATS2D Donor-Acceptor at lag 08	CATS 2D
712	CATS2D_09_DA	CATS2D Donor-Acceptor at lag 09	CATS 2D
713	CATS2D_00_DP	CATS2D Donor-Positive at lag 00	CATS 2D
714	CATS2D_02_DP	CATS2D Donor-Positive at lag 02	CATS 2D
715	CATS2D_03_DP	CATS2D Donor-Positive at lag 03	CATS 2D
716	CATS2D_04_DP	CATS2D Donor-Positive at lag 04	CATS 2D
717	CATS2D_05_DP	CATS2D Donor-Positive at lag 05	CATS 2D
718	CATS2D_06_DP	CATS2D Donor-Positive at lag 06	CATS 2D
719	CATS2D_07_DP	CATS2D Donor-Positive at lag 07	CATS 2D
720	CATS2D_08_DP	CATS2D Donor-Positive at lag 08	CATS 2D
721	CATS2D_09_DP	CATS2D Donor-Positive at lag 09	CATS 2D
722	CATS2D_01_DN	CATS2D Donor-Negative at lag 01	CATS 2D
723	CATS2D_02_DN	CATS2D Donor-Negative at lag 02	CATS 2D
724	CATS2D_03_DN	CATS2D Donor-Negative at lag 03	CATS 2D
725	CATS2D_04_DN	CATS2D Donor-Negative at lag 04	CATS 2D

No.	Name	Description	Block
726	CATS2D_05_DN	CATS2D Donor-Negative at lag 05	CATS 2D
727	CATS2D_06_DN	CATS2D Donor-Negative at lag 06	CATS 2D
728	CATS2D_07_DN	CATS2D Donor-Negative at lag 07	CATS 2D
729	CATS2D_02_DL	CATS2D Donor-Lipophilic at lag 02	CATS 2D
730	CATS2D_03_DL	CATS2D Donor-Lipophilic at lag 03	CATS 2D
731	CATS2D_04_DL	CATS2D Donor-Lipophilic at lag 04	CATS 2D
732	CATS2D_05_DL	CATS2D Donor-Lipophilic at lag 05	CATS 2D
733	CATS2D_06_DL	CATS2D Donor-Lipophilic at lag 06	CATS 2D
734	CATS2D_07_DL	CATS2D Donor-Lipophilic at lag 07	CATS 2D
735	CATS2D_08_DL	CATS2D Donor-Lipophilic at lag 08	CATS 2D
736	CATS2D_09_DL	CATS2D Donor-Lipophilic at lag 09	CATS 2D
737	CATS2D_00_AA	CATS2D Acceptor-Acceptor at lag 00	CATS 2D
738	CATS2D_01_AA	CATS2D Acceptor-Acceptor at lag 01	CATS 2D
739	CATS2D_02_AA	CATS2D Acceptor-Acceptor at lag 02	CATS 2D
740	CATS2D_03_AA	CATS2D Acceptor-Acceptor at lag 03	CATS 2D
741	CATS2D_04_AA	CATS2D Acceptor-Acceptor at lag 04	CATS 2D
742	CATS2D_05_AA	CATS2D Acceptor-Acceptor at lag 05	CATS 2D
743	CATS2D_06_AA	CATS2D Acceptor-Acceptor at lag 06	CATS 2D
744	CATS2D_07_AA	CATS2D Acceptor-Acceptor at lag 07	CATS 2D
745	CATS2D_08_AA	CATS2D Acceptor-Acceptor at lag 08	CATS 2D
746	CATS2D_09_AA	CATS2D Acceptor-Acceptor at lag 09	CATS 2D
747	CATS2D_00_AP	CATS2D Acceptor-Positive at lag 00	CATS 2D
748	CATS2D_02_AP	CATS2D Acceptor-Positive at lag 02	CATS 2D
749	CATS2D_03_AP	CATS2D Acceptor-Positive at lag 03	CATS 2D
750	CATS2D_04_AP	CATS2D Acceptor-Positive at lag 04	CATS 2D
751	CATS2D_05_AP	CATS2D Acceptor-Positive at lag 05	CATS 2D
752	CATS2D_06_AP	CATS2D Acceptor-Positive at lag 06	CATS 2D
753	CATS2D_07_AP	CATS2D Acceptor-Positive at lag 07	CATS 2D
754	CATS2D_08_AP	CATS2D Acceptor-Positive at lag 08	CATS 2D
755	CATS2D_09_AP	CATS2D Acceptor-Positive at lag 09	CATS 2D
756	CATS2D_00_AN	CATS2D Acceptor-Negative at lag 00	CATS 2D
757	CATS2D_01_AN	CATS2D Acceptor-Negative at lag 01	CATS 2D
758	CATS2D_02_AN	CATS2D Acceptor-Negative at lag 02	CATS 2D
759	CATS2D_03_AN	CATS2D Acceptor-Negative at lag 03	CATS 2D
760	CATS2D_04_AN	CATS2D Acceptor-Negative at lag 04	CATS 2D
761	CATS2D_05_AN	CATS2D Acceptor-Negative at lag 05	CATS 2D
762	CATS2D_06_AN	CATS2D Acceptor-Negative at lag 06	CATS 2D
763	CATS2D_07_AN	CATS2D Acceptor-Negative at lag 07	CATS 2D
764	CATS2D_02_AL	CATS2D Acceptor-Lipophilic at lag 02	CATS 2D
765	CATS2D_03_AL	CATS2D Acceptor-Lipophilic at lag 03	CATS 2D
766	CATS2D_04_AL	CATS2D Acceptor-Lipophilic at lag 04	CATS 2D
767	CATS2D_05_AL	CATS2D Acceptor-Lipophilic at lag 05	CATS 2D
768	CATS2D_06_AL	CATS2D Acceptor-Lipophilic at lag 06	CATS 2D
769	CATS2D_07_AL	CATS2D Acceptor-Lipophilic at lag 07	CATS 2D
770	CATS2D_08_AL	CATS2D Acceptor-Lipophilic at lag 08	CATS 2D
771	CATS2D_09_AL	CATS2D Acceptor-Lipophilic at lag 09	CATS 2D
772	CATS2D_00_PP	CATS2D Positive-Positive at lag 00	CATS 2D
773	CATS2D_02_PP	CATS2D Positive-Positive at lag 02	CATS 2D
774	CATS2D_04_PP	CATS2D Positive-Positive at lag 04	CATS 2D
775	CATS2D_05_PP	CATS2D Positive-Positive at lag 05	CATS 2D
776	CATS2D_06_PP	CATS2D Positive-Positive at lag 06	CATS 2D
777	CATS2D_07_PP	CATS2D Positive-Positive at lag 07	CATS 2D
778	CATS2D_02_PN	CATS2D Positive-Negative at lag 02	CATS 2D
779	CATS2D_03_PN	CATS2D Positive-Negative at lag 03	CATS 2D
780	CATS2D_04_PN	CATS2D Positive-Negative at lag 04	CATS 2D
781	CATS2D_05_PN	CATS2D Positive-Negative at lag 05	CATS 2D
782	CATS2D_06_PN	CATS2D Positive-Negative at lag 06	CATS 2D
783	CATS2D_07_PN	CATS2D Positive-Negative at lag 07	CATS 2D
784	CATS2D_02_PL	CATS2D Positive-Lipophilic at lag 02	CATS 2D
785	CATS2D_03_PL	CATS2D Positive-Lipophilic at lag 03	CATS 2D
786	CATS2D_04_PL	CATS2D Positive-Lipophilic at lag 04	CATS 2D
787	CATS2D_05_PL	CATS2D Positive-Lipophilic at lag 05	CATS 2D
788	CATS2D_06_PL	CATS2D Positive-Lipophilic at lag 06	CATS 2D
789	CATS2D_07_PL	CATS2D Positive-Lipophilic at lag 07	CATS 2D
790	CATS2D_08_PL	CATS2D Positive-Lipophilic at lag 08	CATS 2D
791	CATS2D_00_NN	CATS2D Negative-Negative at lag 00	CATS 2D
792	CATS2D_01_NN	CATS2D Negative-Negative at lag 01	CATS 2D
793	CATS2D_02_NN	CATS2D Negative-Negative at lag 02	CATS 2D
794	CATS2D_03_NN	CATS2D Negative-Negative at lag 03	CATS 2D
795	CATS2D_04_NN	CATS2D Negative-Negative at lag 04	CATS 2D
796	CATS2D_05_NN	CATS2D Negative-Negative at lag 05	CATS 2D
797	CATS2D_06_NN	CATS2D Negative-Negative at lag 06	CATS 2D
798	CATS2D_00_NL	CATS2D Negative-Lipophilic at lag 00	CATS 2D

No.	Name	Description	Block
799	CATS2D_01_NL	CATS2D Negative-Lipophilic at lag 01	CATS 2D
800	CATS2D_02_NL	CATS2D Negative-Lipophilic at lag 02	CATS 2D
801	CATS2D_03_NL	CATS2D Negative-Lipophilic at lag 03	CATS 2D
802	CATS2D_04_NL	CATS2D Negative-Lipophilic at lag 04	CATS 2D
803	CATS2D_05_NL	CATS2D Negative-Lipophilic at lag 05	CATS 2D
804	CATS2D_06_NL	CATS2D Negative-Lipophilic at lag 06	CATS 2D
805	CATS2D_07_NL	CATS2D Negative-Lipophilic at lag 07	CATS 2D
806	CATS2D_08_NL	CATS2D Negative-Lipophilic at lag 08	CATS 2D
807	CATS2D_09_NL	CATS2D Negative-Lipophilic at lag 09	CATS 2D
808	CATS2D_00_LL	CATS2D Lipophilic-Lipophilic at lag 00	CATS 2D
809	CATS2D_01_LL	CATS2D Lipophilic-Lipophilic at lag 01	CATS 2D
810	CATS2D_02_LL	CATS2D Lipophilic-Lipophilic at lag 02	CATS 2D
811	CATS2D_03_LL	CATS2D Lipophilic-Lipophilic at lag 03	CATS 2D
812	CATS2D_04_LL	CATS2D Lipophilic-Lipophilic at lag 04	CATS 2D
813	CATS2D_05_LL	CATS2D Lipophilic-Lipophilic at lag 05	CATS 2D
814	CATS2D_06_LL	CATS2D Lipophilic-Lipophilic at lag 06	CATS 2D
815	CATS2D_07_LL	CATS2D Lipophilic-Lipophilic at lag 07	CATS 2D
816	CATS2D_08_LL	CATS2D Lipophilic-Lipophilic at lag 08	CATS 2D
817	CATS2D_09_LL	CATS2D Lipophilic-Lipophilic at lag 09	CATS 2D
818	T(N..N)	sum of topological distances between N..N	2D Atom Pairs
819	T(N..O)	sum of topological distances between N..O	2D Atom Pairs
820	T(N..S)	sum of topological distances between N..S	2D Atom Pairs
821	T(O..O)	sum of topological distances between O..O	2D Atom Pairs
822	T(O..S)	sum of topological distances between O..S	2D Atom Pairs
823	T(S..S)	sum of topological distances between S..S	2D Atom Pairs
824	B01[C-C]	Presence/absence of C - C at topological distance 1	2D Atom Pairs
825	B01[C-N]	Presence/absence of C - N at topological distance 1	2D Atom Pairs
826	B01[C-O]	Presence/absence of C - O at topological distance 1	2D Atom Pairs
827	B01[C-S]	Presence/absence of C - S at topological distance 1	2D Atom Pairs
828	B01[N-O]	Presence/absence of N - O at topological distance 1	2D Atom Pairs
829	B01[O-S]	Presence/absence of O - S at topological distance 1	2D Atom Pairs
830	B01[O-P]	Presence/absence of O - P at topological distance 1	2D Atom Pairs
831	B01[S-S]	Presence/absence of S - S at topological distance 1	2D Atom Pairs
832	B02[C-C]	Presence/absence of C - C at topological distance 2	2D Atom Pairs
833	B02[C-N]	Presence/absence of C - N at topological distance 2	2D Atom Pairs
834	B02[C-O]	Presence/absence of C - O at topological distance 2	2D Atom Pairs
835	B02[C-S]	Presence/absence of C - S at topological distance 2	2D Atom Pairs
836	B02[N-N]	Presence/absence of N - N at topological distance 2	2D Atom Pairs
837	B02[N-O]	Presence/absence of N - O at topological distance 2	2D Atom Pairs
838	B02[N-S]	Presence/absence of N - S at topological distance 2	2D Atom Pairs
839	B02[O-O]	Presence/absence of O - O at topological distance 2	2D Atom Pairs
840	B02[O-S]	Presence/absence of O - S at topological distance 2	2D Atom Pairs
841	B02[S-S]	Presence/absence of S - S at topological distance 2	2D Atom Pairs
842	B03[C-C]	Presence/absence of C - C at topological distance 3	2D Atom Pairs
843	B03[C-N]	Presence/absence of C - N at topological distance 3	2D Atom Pairs
844	B03[C-O]	Presence/absence of C - O at topological distance 3	2D Atom Pairs
845	B03[C-S]	Presence/absence of C - S at topological distance 3	2D Atom Pairs
846	B03[N-N]	Presence/absence of N - N at topological distance 3	2D Atom Pairs
847	B03[N-O]	Presence/absence of N - O at topological distance 3	2D Atom Pairs
848	B03[N-S]	Presence/absence of N - S at topological distance 3	2D Atom Pairs
849	B03[O-O]	Presence/absence of O - O at topological distance 3	2D Atom Pairs
850	B03[O-S]	Presence/absence of O - S at topological distance 3	2D Atom Pairs
851	B03[S-S]	Presence/absence of S - S at topological distance 3	2D Atom Pairs
852	B04[C-C]	Presence/absence of C - C at topological distance 4	2D Atom Pairs
853	B04[C-N]	Presence/absence of C - N at topological distance 4	2D Atom Pairs
854	B04[C-O]	Presence/absence of C - O at topological distance 4	2D Atom Pairs
855	B04[C-S]	Presence/absence of C - S at topological distance 4	2D Atom Pairs
856	B04[N-N]	Presence/absence of N - N at topological distance 4	2D Atom Pairs
857	B04[N-O]	Presence/absence of N - O at topological distance 4	2D Atom Pairs
858	B04[N-S]	Presence/absence of N - S at topological distance 4	2D Atom Pairs
859	B04[O-O]	Presence/absence of O - O at topological distance 4	2D Atom Pairs
860	B04[O-S]	Presence/absence of O - S at topological distance 4	2D Atom Pairs
861	B04[S-S]	Presence/absence of S - S at topological distance 4	2D Atom Pairs
862	B05[C-C]	Presence/absence of C - C at topological distance 5	2D Atom Pairs
863	B05[C-N]	Presence/absence of C - N at topological distance 5	2D Atom Pairs
864	B05[C-O]	Presence/absence of C - O at topological distance 5	2D Atom Pairs
865	B05[C-S]	Presence/absence of C - S at topological distance 5	2D Atom Pairs
866	B05[N-N]	Presence/absence of N - N at topological distance 5	2D Atom Pairs
867	B05[N-O]	Presence/absence of N - O at topological distance 5	2D Atom Pairs
868	B05[N-S]	Presence/absence of N - S at topological distance 5	2D Atom Pairs
869	B05[O-O]	Presence/absence of O - O at topological distance 5	2D Atom Pairs
870	B05[O-S]	Presence/absence of O - S at topological distance 5	2D Atom Pairs
871	B05[S-S]	Presence/absence of S - S at topological distance 5	2D Atom Pairs

No.	Name	Description	Block
872	B06[C-C]	Presence/absence of C - C at topological distance 6	2D Atom Pairs
873	B06[C-N]	Presence/absence of C - N at topological distance 6	2D Atom Pairs
874	B06[C-O]	Presence/absence of C - O at topological distance 6	2D Atom Pairs
875	B06[C-S]	Presence/absence of C - S at topological distance 6	2D Atom Pairs
876	B06[N-N]	Presence/absence of N - N at topological distance 6	2D Atom Pairs
877	B06[N-O]	Presence/absence of N - O at topological distance 6	2D Atom Pairs
878	B06[N-S]	Presence/absence of N - S at topological distance 6	2D Atom Pairs
879	B06[O-O]	Presence/absence of O - O at topological distance 6	2D Atom Pairs
880	B06[O-S]	Presence/absence of O - S at topological distance 6	2D Atom Pairs
881	B06[S-S]	Presence/absence of S - S at topological distance 6	2D Atom Pairs
882	B07[C-C]	Presence/absence of C - C at topological distance 7	2D Atom Pairs
883	B07[C-N]	Presence/absence of C - N at topological distance 7	2D Atom Pairs
884	B07[C-O]	Presence/absence of C - O at topological distance 7	2D Atom Pairs
885	B07[C-S]	Presence/absence of C - S at topological distance 7	2D Atom Pairs
886	B07[N-N]	Presence/absence of N - N at topological distance 7	2D Atom Pairs
887	B07[N-O]	Presence/absence of N - O at topological distance 7	2D Atom Pairs
888	B07[O-O]	Presence/absence of O - O at topological distance 7	2D Atom Pairs
889	B07[S-S]	Presence/absence of S - S at topological distance 7	2D Atom Pairs
890	B08[C-C]	Presence/absence of C - C at topological distance 8	2D Atom Pairs
891	B08[C-N]	Presence/absence of C - N at topological distance 8	2D Atom Pairs
892	B08[C-O]	Presence/absence of C - O at topological distance 8	2D Atom Pairs
893	B08[C-S]	Presence/absence of C - S at topological distance 8	2D Atom Pairs
894	B08[N-O]	Presence/absence of N - O at topological distance 8	2D Atom Pairs
895	B08[O-O]	Presence/absence of O - O at topological distance 8	2D Atom Pairs
896	B09[C-C]	Presence/absence of C - C at topological distance 9	2D Atom Pairs
897	B09[C-N]	Presence/absence of C - N at topological distance 9	2D Atom Pairs
898	B09[C-O]	Presence/absence of C - O at topological distance 9	2D Atom Pairs
899	B09[C-S]	Presence/absence of C - S at topological distance 9	2D Atom Pairs
900	B09[N-O]	Presence/absence of N - O at topological distance 9	2D Atom Pairs
901	B09[O-O]	Presence/absence of O - O at topological distance 9	2D Atom Pairs
902	B09[S-S]	Presence/absence of S - S at topological distance 9	2D Atom Pairs
903	B10[C-C]	Presence/absence of C - C at topological distance 10	2D Atom Pairs
904	B10[C-N]	Presence/absence of C - N at topological distance 10	2D Atom Pairs
905	B10[C-O]	Presence/absence of C - O at topological distance 10	2D Atom Pairs
906	B10[N-O]	Presence/absence of N - O at topological distance 10	2D Atom Pairs
907	B10[O-O]	Presence/absence of O - O at topological distance 10	2D Atom Pairs
908	B10[S-S]	Presence/absence of S - S at topological distance 10	2D Atom Pairs
909	F01[C-C]	Frequency of C - C at topological distance 1	2D Atom Pairs
910	F01[C-N]	Frequency of C - N at topological distance 1	2D Atom Pairs
911	F01[C-O]	Frequency of C - O at topological distance 1	2D Atom Pairs
912	F01[C-S]	Frequency of C - S at topological distance 1	2D Atom Pairs
913	F01[N-O]	Frequency of N - O at topological distance 1	2D Atom Pairs
914	F01[O-S]	Frequency of O - S at topological distance 1	2D Atom Pairs
915	F01[O-P]	Frequency of O - P at topological distance 1	2D Atom Pairs
916	F01[S-S]	Frequency of S - S at topological distance 1	2D Atom Pairs
917	F02[C-C]	Frequency of C - C at topological distance 2	2D Atom Pairs
918	F02[C-N]	Frequency of C - N at topological distance 2	2D Atom Pairs
919	F02[C-O]	Frequency of C - O at topological distance 2	2D Atom Pairs
920	F02[C-S]	Frequency of C - S at topological distance 2	2D Atom Pairs
921	F02[N-N]	Frequency of N - N at topological distance 2	2D Atom Pairs
922	F02[N-O]	Frequency of N - O at topological distance 2	2D Atom Pairs
923	F02[N-S]	Frequency of N - S at topological distance 2	2D Atom Pairs
924	F02[O-O]	Frequency of O - O at topological distance 2	2D Atom Pairs
925	F02[O-S]	Frequency of O - S at topological distance 2	2D Atom Pairs
926	F02[S-S]	Frequency of S - S at topological distance 2	2D Atom Pairs
927	F03[C-C]	Frequency of C - C at topological distance 3	2D Atom Pairs
928	F03[C-N]	Frequency of C - N at topological distance 3	2D Atom Pairs
929	F03[C-O]	Frequency of C - O at topological distance 3	2D Atom Pairs
930	F03[C-S]	Frequency of C - S at topological distance 3	2D Atom Pairs
931	F03[N-N]	Frequency of N - N at topological distance 3	2D Atom Pairs
932	F03[N-O]	Frequency of N - O at topological distance 3	2D Atom Pairs
933	F03[N-S]	Frequency of N - S at topological distance 3	2D Atom Pairs
934	F03[O-O]	Frequency of O - O at topological distance 3	2D Atom Pairs
935	F03[O-S]	Frequency of O - S at topological distance 3	2D Atom Pairs
936	F03[S-S]	Frequency of S - S at topological distance 3	2D Atom Pairs
937	F04[C-C]	Frequency of C - C at topological distance 4	2D Atom Pairs
938	F04[C-N]	Frequency of C - N at topological distance 4	2D Atom Pairs
939	F04[C-O]	Frequency of C - O at topological distance 4	2D Atom Pairs
940	F04[C-S]	Frequency of C - S at topological distance 4	2D Atom Pairs
941	F04[N-N]	Frequency of N - N at topological distance 4	2D Atom Pairs
942	F04[N-O]	Frequency of N - O at topological distance 4	2D Atom Pairs
943	F04[N-S]	Frequency of N - S at topological distance 4	2D Atom Pairs
944	F04[O-O]	Frequency of O - O at topological distance 4	2D Atom Pairs

No.	Name	Description	Block
945	F04[O-S]	Frequency of O - S at topological distance 4	2D Atom Pairs
946	F04[S-S]	Frequency of S - S at topological distance 4	2D Atom Pairs
947	F05[C-C]	Frequency of C - C at topological distance 5	2D Atom Pairs
948	F05[C-N]	Frequency of C - N at topological distance 5	2D Atom Pairs
949	F05[C-O]	Frequency of C - O at topological distance 5	2D Atom Pairs
950	F05[C-S]	Frequency of C - S at topological distance 5	2D Atom Pairs
951	F05[N-N]	Frequency of N - N at topological distance 5	2D Atom Pairs
952	F05[N-O]	Frequency of N - O at topological distance 5	2D Atom Pairs
953	F05[N-S]	Frequency of N - S at topological distance 5	2D Atom Pairs
954	F05[O-O]	Frequency of O - O at topological distance 5	2D Atom Pairs
955	F05[O-S]	Frequency of O - S at topological distance 5	2D Atom Pairs
956	F05[S-S]	Frequency of S - S at topological distance 5	2D Atom Pairs
957	F06[C-C]	Frequency of C - C at topological distance 6	2D Atom Pairs
958	F06[C-N]	Frequency of C - N at topological distance 6	2D Atom Pairs
959	F06[C-O]	Frequency of C - O at topological distance 6	2D Atom Pairs
960	F06[C-S]	Frequency of C - S at topological distance 6	2D Atom Pairs
961	F06[N-N]	Frequency of N - N at topological distance 6	2D Atom Pairs
962	F06[N-O]	Frequency of N - O at topological distance 6	2D Atom Pairs
963	F06[N-S]	Frequency of N - S at topological distance 6	2D Atom Pairs
964	F06[O-O]	Frequency of O - O at topological distance 6	2D Atom Pairs
965	F06[O-S]	Frequency of O - S at topological distance 6	2D Atom Pairs
966	F06[S-S]	Frequency of S - S at topological distance 6	2D Atom Pairs
967	F07[C-C]	Frequency of C - C at topological distance 7	2D Atom Pairs
968	F07[C-N]	Frequency of C - N at topological distance 7	2D Atom Pairs
969	F07[C-O]	Frequency of C - O at topological distance 7	2D Atom Pairs
970	F07[C-S]	Frequency of C - S at topological distance 7	2D Atom Pairs
971	F07[N-N]	Frequency of N - N at topological distance 7	2D Atom Pairs
972	F07[N-O]	Frequency of N - O at topological distance 7	2D Atom Pairs
973	F07[O-O]	Frequency of O - O at topological distance 7	2D Atom Pairs
974	F07[S-S]	Frequency of S - S at topological distance 7	2D Atom Pairs
975	F08[C-C]	Frequency of C - C at topological distance 8	2D Atom Pairs
976	F08[C-N]	Frequency of C - N at topological distance 8	2D Atom Pairs
977	F08[C-O]	Frequency of C - O at topological distance 8	2D Atom Pairs
978	F08[C-S]	Frequency of C - S at topological distance 8	2D Atom Pairs
979	F08[N-O]	Frequency of N - O at topological distance 8	2D Atom Pairs
980	F08[O-O]	Frequency of O - O at topological distance 8	2D Atom Pairs
981	F09[C-C]	Frequency of C - C at topological distance 9	2D Atom Pairs
982	F09[C-N]	Frequency of C - N at topological distance 9	2D Atom Pairs
983	F09[C-O]	Frequency of C - O at topological distance 9	2D Atom Pairs
984	F09[C-S]	Frequency of C - S at topological distance 9	2D Atom Pairs
985	F09[N-O]	Frequency of N - O at topological distance 9	2D Atom Pairs
986	F09[O-O]	Frequency of O - O at topological distance 9	2D Atom Pairs
987	F09[S-S]	Frequency of S - S at topological distance 9	2D Atom Pairs
988	F10[C-C]	Frequency of C - C at topological distance 10	2D Atom Pairs
989	F10[C-N]	Frequency of C - N at topological distance 10	2D Atom Pairs
990	F10[C-O]	Frequency of C - O at topological distance 10	2D Atom Pairs
991	F10[N-O]	Frequency of N - O at topological distance 10	2D Atom Pairs
992	F10[O-O]	Frequency of O - O at topological distance 10	2D Atom Pairs
993	F10[S-S]	Frequency of S - S at topological distance 10	2D Atom Pairs
994	Uc	unsaturation count	Molecular properties
995	Ui	unsaturation index	Molecular properties
996	Hy	hydrophilic factor	Molecular properties
997	TPSA(NO)	topological polar surface area using N,O polar contributions	Molecular properties
998	TPSA(Tot)	topological polar surface area using N,O,S,P polar contributions	Molecular properties
999	SAacc	surface area of acceptor atoms from P_VSA-like descriptors	Molecular properties
1000	SAdon	surface area of donor atoms from P_VSA-like descriptors	Molecular properties
1001	Vx	McGowan volume	Molecular properties
1002	VvdwMG	van der Waal volume from McGowan volume	Molecular properties
1003	Ro5	Lipinski Rule of 5	Drug-like indices
1004	cRo5	Complementary Lipinski Alert index	Drug-like indices
1005	DLS_05	modified drug-like score from Zheng et al. (2 rules)	Drug-like indices
1006	DLS_07	modified drug-like score from Veber et al. (2 rules)	Drug-like indices



# Appendix C

## Chapter 6 support information

Table C.1 Standard performance measures of LDA, KNN, and NBC models for train set (%).

Models	PVOCs	Sensitivity	Specificity	Precision	Recall	F1 score	Accuracy
KNN	cis-Jasmone	96.55±12.23	97.09±4.56	77.5±35.8	96.55±12.23	91.11±15.66	97.54±6.63
	α-Pinene	82.89±22.24	99.07±2.95	93±22.5	82.89±22.24	88.72±13.54	91.1±11.69
	Limonene	95.58±12.1	98.81±2.85	91±21.77	95.58±12.1	92.86±13.37	97.31±6.29
	γ-Terpiene	96±12.34	99.2±2.18	94±16.33	96±12.34	93.47±12.68	97.6±6.13
	cis-Jasmone+α-Pinene	89.61±21.69	95.74±4.12	68±31.4	89.61±21.69	77.72±17.84	92.98±11.17
	cis-Jasmone+Limonene	98.23±7.52	98.92±3.24	91.5±25.68	98.23±7.52	97.16±8.55	98.94±3.79
	α-Pinene+Limonene	80.17±21.15	99.65±1.52	97.5±10.95	80.17±21.15	86.2±15.21	89.91±10.69
	Limonene+γ-Terpiene	98.96±5.83	99.1±2.86	93±22.5	98.96±5.83	97.29±8.68	99.27±2.99
LDA	cis-Jasmone	99.88±1.25	99.98±0.2	99.86±1.43	99.88±1.25	99.86±1.01	99.93±0.63
	α-Pinene	80.19±8.24	99.77±0.66	98.43±4.49	80.19±8.24	88.15±5.54	89.98±4.2
	Limonene	99.57±2.45	98±1.35	85.57±9.84	99.57±2.45	91.73±5.85	98.78±1.4
	γ-Terpiene	100±0	99.74±0.68	98.14±4.83	100±0	99±2.6	99.87±0.34
	cis-Jasmone+α-Pinene	78.01±8.15	98.25±0.77	88±5.26	78.01±8.15	82.56±6.15	88.13±4.32
	cis-Jasmone+Limonene	99.13±3.21	99.98±0.2	99.86±1.43	99.13±3.21	99.46±1.86	99.55±1.6
	α-Pinene+Limonene	98.7±4.5	96.55±1.31	74.86±9.76	98.7±4.5	84.7±6.07	97.62±2.2
	Limonene+γ-Terpiene	98.75±3.77	100±0	100±0	98.75±3.77	99.33±2.01	99.38±1.88
NBC	cis-Jasmone	95.55±6.86	98.4±1.17	88.57±8.37	95.55±6.86	91.63±5.75	96.98±3.53
	α-Pinene	79.36±8.74	97.73±1.34	84.14±9.5	79.36±8.74	81.28±6.89	88.55±4.55
	Limonene	99.75±1.76	99.9±0.44	99.29±3.13	99.75±1.76	99.48±1.9	99.83±0.9
	γ-Terpiene	99.88±1.25	100±0	100±0	99.88±1.25	99.93±0.67	99.94±0.63
	cis-Jasmone+α-Pinene	94.54±8.22	98.42±0.84	88.86±5.95	94.54±8.22	91.42±5.87	96.48±4.29
	cis-Jasmone+Limonene	100±0	100±0	100±0	100±0	100±0	100±0
	α-Pinene+Limonene	89.61±5.95	99.19±1.24	94.29±8.85	89.61±5.95	91.61±5.65	94.4±3.11
	Limonene+γ-Terpiene	100±0	100±0	100±0	100±0	100±0	100±0

Table C.2 Standard performance measures of LDA, KNN, and NBC models for test set (%).

Models	PVOCs	Sensitivity	Specificity	Precision	Recall	F1 score	Accuracy
KNN	cis-Jasmone	95.55±6.86	98.4±1.17	88.57±8.37	95.55±6.86	91.63±5.75	96.98±3.53
	α-Pinene	79.36±8.74	97.73±1.34	84.14±9.5	79.36±8.74	81.28±6.89	88.55±4.55
	Limonene	99.75±1.76	99.9±0.44	99.29±3.13	99.75±1.76	99.48±1.9	99.83±0.9
	γ-Terpiene	99.88±1.25	100±0	100±0	99.88±1.25	99.93±0.67	99.94±0.63
	cis-Jasmone+α-Pinene	94.54±8.22	98.42±0.84	88.86±5.95	94.54±8.22	91.42±5.87	96.48±4.29
	cis-Jasmone+Limonene	100±0	100±0	100±0	100±0	100±0	100±0
	α-Pinene+Limonene	89.61±5.95	99.19±1.24	94.29±8.85	89.61±5.95	91.61±5.65	94.4±3.11
	Limonene+γ-Terpiene	100±0	100±0	100±0	100±0	100±0	100±0
LDA	cis-Jasmone	98.64±8.18	98.87±2.8	91.5±21.39	98.64±8.18	94.9±12.51	98.87±4.48
	α-Pinene	74.74±24.83	98.72±3.37	90.5±25.32	74.74±24.83	83.05±15.4	86.85±13.13
	Limonene	89.82±23.48	96.79±4.26	76±32.16	89.82±23.48	85.04±18.39	93.55±12.81
	γ-Terpiene	100±0	98.16±3.34	86±25.7	100±0	92.44±14.03	99.24±1.4
	cis-Jasmone+α-Pinene	75.29±22.66	98.33±3.12	88±22.61	75.29±22.66	78.63±17.65	86.86±11.59
	cis-Jasmone+Limonene	95.62±11.32	99.74±1.55	98±12.14	95.62±11.32	96.7±8.05	97.74±5.65
	α-Pinene+Limonene	98.24±8.14	95.12±4.46	62.5±35.09	98.24±8.14	81.45±16.67	97.35±4.28
	Limonene+γ-Terpiene	94.67±12.28	100±0	100±0	94.67±12.28	96.8±7.37	97.33±6.14
NBC	cis-Jasmone	86.29±21.78	97.09±3.74	78.5±27.76	86.29±21.78	80.66±17.79	91.79±11.3
	α-Pinene	69.39±29.74	95.74±4.48	69±33.17	69.39±29.74	71.32±18.07	82.74±15.86
	Limonene	89.65±18.38	98.83±3.37	91±26	89.65±18.38	92.3±12.05	94.54±9.6
	γ-Terpiene	97.94±9.08	98.16±3.34	86±25.7	97.94±9.08	91.13±14.52	98.21±4.58
	cis-Jasmone+α-Pinene	89.52±16.37	98.36±3.25	87.5±25	89.52±16.37	87.08±13.94	94.11±7.87
	cis-Jasmone+Limonene	100±0	100±0	100±0	100±0	100±0	100±0
	α-Pinene+Limonene	89.29±20.2	97.91±3.51	84.5±26.3	89.29±20.2	86.53±17.51	93.71±10.77
	Limonene+γ-Terpiene	99.75±1.76	98±14.07	100±0	100±0	100±0	100±0