

การวิเคราะห์เชิงสถิติหลายตัวแปรในออปติคัลสเปกโตรสโคปิกเซ็นเซอร์แอร์เรย์เพื่อจำแนกตัวอย่าง
สารอินทรีย์และสารชีวภาพ



นางสาวรติมาดา มุ่งการดี

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีชีวภาพ
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2557
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

MULTIVARIATE STATISTICAL ANALYSES IN OPTICAL SPECTROSCOPIC SENSOR ARRAY
FOR DISCRIMINATING ORGANIC AND BIOLOGICAL SAMPLES

Miss Radeemada Mungkarndee



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Biotechnology

Faculty of Science

Chulalongkorn University

Academic Year 2014

Copyright of Chulalongkorn University

Thesis Title	MULTIVARIATE STATISTICAL ANALYSES IN OPTICAL SPECTROSCOPIC SENSOR ARRAY FOR DISCRIMINATING ORGANIC AND BIOLOGICAL SAMPLES
By	Miss Radeemada Mungkamdee
Field of Study	Biotechnology
Thesis Advisor	Associate Professor Mongkol Sukwattanasinitt, Ph.D.
Thesis Co-Advisor	Gamolwan Tumcharern, Ph.D. Ittipon Techakriengkrai, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.....Dean of the Faculty of Science
(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

.....Chairman
(Assistant Professor Varawut Tangpasuthadol, Ph.D.)

.....Thesis Advisor
(Associate Professor Mongkol Sukwattanasinitt, Ph.D.)

.....Thesis Co-Advisor
(Gamolwan Tumcharern, Ph.D.)

.....Thesis Co-Advisor
(Ittipon Techakriengkrai, Ph.D.)

.....Examiner
(Assistant Professor Kanet Wongravee, Ph.D.)

.....Examiner
(Assistant Professor Kanoktip Packdibamrung, Ph.D.)

.....External Examiner
(Prasat Kittakoop, Ph.D.)

รติมาดา มุ่งการดี : การวิเคราะห์เชิงสถิติหลายตัวแปรในออฟติคัลสเปกโตรสโคปิกเซ็นเซอร์แออร์เรย์เพื่อจำแนกตัวอย่างสารอินทรีย์และสารชีวภาพ (MULTIVARIATE STATISTICAL ANALYSES IN OPTICAL SPECTROSCOPIC SENSOR ARRAY FOR DISCRIMINATING ORGANIC AND BIOLOGICAL SAMPLES) อ.ที่ปรึกษาวิทยานิพนธ์
 หลัก: รศ. ดร.มงคล สุขวัฒนาสินธิ์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ดร.กมลวรรณ ธรรมเจริญ, ดร.อิทธิพล เตชะเกรียงไกร, หน้า.

เนื่องจากความกังวลทางด้านสิ่งแวดล้อมและสาธารณสุข ส่งผลให้การตรวจหาสารอินทรีย์ในตัวอย่างทางชีวภาพมีความสำคัญทั้งทางด้านเศรษฐกิจและกฎหมาย เช่น การตรวจหาโปรตีนที่อยู่ในผลิตภัณฑ์จากนม การตรวจหาสารลดแรงตึงผิวที่ใช้ในผลิตภัณฑ์เครื่องสำอางและยา การตรวจหาสารประกอบอินทรีย์ระเหยที่ปนเปื้อนในเครื่องดื่มปราศจากแอลกอฮอล์ และการตรวจหาเชื้อจุลินทรีย์ก่อโรคในอาหาร การประยุกต์ใช้การวิเคราะห์เชิงสถิติหลายตัวแปรกับการตอบสนองของสเปกโตรสโคปิกเชิงแสงที่ถูกเหนี่ยวนำโดยอันตรกิริยาระหว่างองค์ประกอบเซ็นเซอร์กับสารอินทรีย์ทำให้สามารถจัดจำแนกตัวอย่างทางชีวภาพได้ การวิเคราะห์การรู้จำแบบของการตอบสนองฟลูออเรสเซนซ์ ที่มีการเติมตัวอย่างนมพาณิชย์กับสารเรืองแสง ด้วยการวิเคราะห์การจำแนกเชิงเส้น (linear discriminant analysis) สามารถจำแนกตัวอย่างนมที่ผ่านกระบวนการทางความร้อน ได้แก่ นมพาสเจอร์ไรซ์ นมสเตอริไรซ์ นมสดยูเอชที และนมคั้นรูป (นมสดยูเอชทีที่มีนมผงผสมอยู่) ได้ที่ความถูกต้อง 100% วิธีการเช่นเดียวกันให้ผลสำเร็จในการนำมาประยุกต์ใช้กับการจำแนกเชื้อจุลินทรีย์ก่อโรคในอาหาร 8 ชนิด นอกจากนี้ การจำแนกสารลดแรงตึงผิวชนิดประจุลบสามารถทำได้โดยการวิเคราะห์องค์ประกอบหลัก (principal component analysis) ต่อร้อยละของการตอบสนองเชิงสีที่ได้จากสัญญาณการดูดกลืนแสงของพอลิไดแอเซทิลีนสามารถได้ การวิเคราะห์องค์ประกอบหลักดังกล่าวยังสามารถวิเคราะห์ค่าการเปลี่ยนแปลงสี RGB ที่ได้จากรูปภาพเซ็นเซอร์แออร์เรย์บนกระดาษที่เกิดจากการเปลี่ยนสีของพอลิไดแอเซทิลีน เพื่อใช้ในการตรวจสอบและแยกสารประกอบอินทรีย์ระเหยได้

สาขาวิชา เทคโนโลยีชีวภาพ
 ปีการศึกษา 2557

ลายมือชื่อนิสิต

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ลายมือชื่อ อ.ที่ปรึกษาร่วม

ลายมือชื่อ อ.ที่ปรึกษาร่วม

5173927323 : MAJOR BIOTECHNOLOGY

KEYWORDS: BIOSENSOR / MULTIVARIATE STATISTICAL ANALYSIS / PRINCIPAL COMPONENT ANALYSIS (PCA) / LINEAR DISCRIMINANT ANALYSIS (LDA) / FLUORESCENCE SENSOR / COLORIMETRIC SENSOR

RADEEMADA MUNGKARNDEE: MULTIVARIATE STATISTICAL ANALYSES IN OPTICAL SPECTROSCOPIC SENSOR ARRAY FOR DISCRIMINATING ORGANIC AND BIOLOGICAL SAMPLES. ADVISOR: ASSOC. PROF. MONGKOL SUKWATTANASINITT, Ph.D., CO-ADVISOR: GAMOLWAN TUMCHARERN, Ph.D., ITTIPON TECHAKRIENGKRAI, Ph.D., pp.

Due to environmental and public health concerns, detection of organic materials in biological samples, such as proteins contained in dairy products, surfactants extensively used in cosmetic and pharmaceutical products, volatile organic compounds (VOCs) contaminated in soft drinks, and foodborne pathogens, is of importance for economic and legal issues. Multivariate statistical analyses applied to the optical spectroscopic response which is induced by interactions between sensing element and organic materials allow a possible discrimination of biological samples. The pattern recognition of fluorescence responses upon addition of commercial milk samples to fluorophore solutions are analyzed by linear discriminant analysis (LDA). Milk samples according to their thermal processes; i.e. paturized milk, sterilized milk, UHT fresh milk and recombined milk (UHT milk having milk powder) are discriminated with 100% classification accuracy. LDA is also successfully applied for the discrimination of eight foodborne pathogens. Anionic surfactants are successfully identified by using principal component analysis (PCA) applied to percentage of colorimetric response obtained from the UV-Vis spectra of polymerized diacetylenes in the presence of anionic surfactants. PCA is also applied to the RGB color changes profile plot obtained from images of paper-based polydiacetylene colorimetric sensor array developed for the detection and identification of VOCs.

Field of Study: Biotechnology

Academic Year: 2014

Student's Signature

Advisor's Signature

Co-Advisor's Signature

Co-Advisor's Signature

ACKNOWLEDGEMENTS

The work has been conducted under the supervision of Associate Professor Dr. Mongkol Sukwattanasinitt. I would like to express my utmost gratitude to Dr. Mongkol for his supervision and support along this work. His great expertise and positive attitude push and motivate me to overcome all obstacles have come in my way. Thank you for his patience and for encouraging me over the past years. One could not have had asked for a better advisor.

My sincere appreciation is given to my co-advisors, Dr. Gamolwan Tumcharern and Dr. Ittipon Techakriengkrai for their kind attention and valuable suggestion. This research is completely impossible to succeed without their helpfulness.

I am also greatly grateful to Associate Professor Dr. Paitoon Rashatasakhon Assistance Professor and Dr. Sumrit Wacharasindhu for their suggestion, generousness during the course of this research.

Special thanks go to my co-workers, Dr. Nakorn Niamnont, Miss Thichamporn Eaidkong and Miss Wanwisa Thongmalai for our discussions and collaborations. Mr. Akachai, Eakkaphon also deserve special thanks for the fluorescent compound and suggestion he provided to me. I have learned a lot from them.

My appreciation is also given to everyone in MAPS Group, Tilleke and Gibbins and PTT Global Chemical whom I could not list all of their names here, for great friendships, smiles, good wishes and helps.

Finally a big great thanks to my beloved family for their encouragement and for listening to all my complaints when the plan did not go as planned, especially my sister who always stands by my side during both of my pleasant and hard time. Without them, this wouldn't have been possible. This thesis is heartily dedicated to my mother who took the lead to heaven before the completion of this work.

CONTENTS

	Page
THAI ABSTRACT	iv
ENGLISH ABSTRACT	v
ACKNOWLEDGEMENTS	vi
CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER I INTRODUCTION	1
1.1 Overview	1
1.2 Chemometrics.....	2
1.2.1. Pattern recognition.....	4
1.2.1.1 Unsupervised patter recognition.....	5
1.2.1.2 Supervised patter recognition.....	7
1.2.2 Feature selection.....	11
1.3 Optical Biosensor Array.....	12
1.3.1 Fluorescence Biosensor.....	12
1.3.2 Colorimetric Sensor.....	14
1.4 Literature reviews.....	17
1.4.1 Fluorescence biosensor array.....	17
1.4.2 Colorimetric sensor array.....	22
1.5 Objectives and scope of the dissertation	25
CHAPTER II EXPERIMENTAL	27
2.1 Fluorescence sensor array.....	27

2.1.1 Materials.....	27
2.1.2 Fluorescence sensor study	28
2.1.2.1 Fluorescence measurement of proteins	28
2.1.2.2 Fluorescence measurement of commercial milk samples via cuvette	29
2.1.2.3 Fluorescence measurement of commercial milk samples via microplate	29
2.1.2.4 Fluorescence measurement of foodborne pathogens via microplate	29
2.1.3 Data preparation	29
2.2 Colorimetric sensor array.....	30
2.2.1 Materials.....	30
2.2.1 Colorimetric vesicle-based sensor array	30
2.2.1.1 Colorimetric detection of surfactants	30
2.2.1.2 Data preparation for vesicle-based colorimetric response.....	31
2.2.2 Colorimetric paper-based sensor array	31
2.2.2.1 Fabrication of PDAs coated paper	31
2.2.2.2 Preparation of saturated analytes vapor	32
2.2.2.3 Image processing methods for screening	32
2.2.2.4 Data preparation.....	32
2.3 Data analysis using multivariate statistical analysis.....	33
2.3.1 Descriptive data analysis	33
2.3.2 Predictive data analysis	33
2.3.3 Feature Selection.....	34

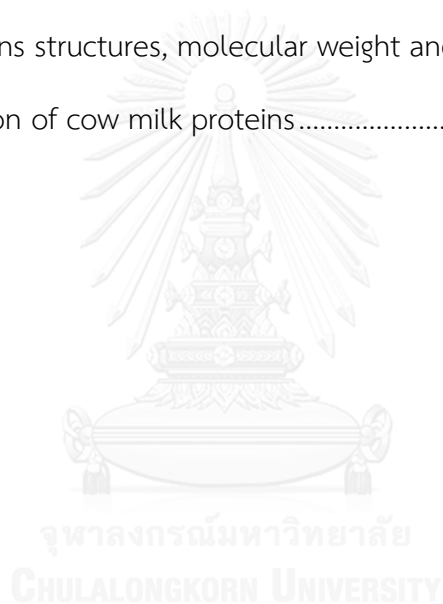
	Page
CHAPTER III RESULTS AND DISCUSSION.....	35
3.1 Fluorescence Sensor Array.....	35
3.1.1 Fluorescence sensor array for protein discrimination.....	35
3.1.1.1 Multivariate statistical analysis.....	37
3.1.1.1.1 Feature selection.....	40
3.1.2 Fluorescence sensor array for milk discrimination.....	43
3.1.2.1 Discrimination of pure milk proteins.....	43
3.1.2.2 Discrimination of commercial milk from different sources.....	45
3.1.2.3 Discrimination of cow milk according to the thermal processes..	46
3.1.2.4 Variable selection and reduction.....	52
3.1.3 Fluorescence sensor array for pathogen discrimination.....	56
3.1.3.1 Variable selection and reduction.....	60
3.2 Colorimetric Sensor Array.....	61
3.2.1 Colorimetric vesicle-based sensor array.....	62
3.2.1.1 Multivariate statistical analysis.....	67
3.2.2 Colorimetric paper-based sensor array for VOCs discrimination.....	67
3.2.2.1 Multivariate Data analysis.....	70
3.2.2.2 Variable selection and reduction.....	72
3.2.2.3 Colorimetric paper-based sensor array for automotive fuel discrimination.....	74
CHAPTER IV CONCLUSION.....	79
.....	81
REFERENCES.....	81

VITA..... 99



LIST OF TABLES

Table 1.1. The colors of the visible light spectrum	15
Table 3.1. Variance contribution of the first two PCs and classification accuracy obtained from PCA and FDA on the fluorescent intensities (ΔI) measured at various wavelengths.....	40
Table 3.2. Pathogenic microorganisms responsible for foodborne illness.....	57
Table A1. Proteins structures, molecular weight and isoelectric point	91
Table A2. Milk proteins structures, molecular weight and isoelectric point	92
Table A3. Composition of cow milk proteins.....	92



LIST OF FIGURES

Figure 1.1 Elements and selected components of a typical biosensor.....	2
Figure 1.2. Principle of Chemometrics	3
Figure 1.3. A principal component analysis (PCA) model approximates the variation in a multi-dimensional data table by a lower dimensional model plane. The directions in the PC score plot correspond to directions in the loading plot, and vice versa	7
Figure 1.4. (a) Good class separation. (b) Bad class separation.	11
Figure 1.5. 2-D 2-class data, along with first LDA basis vector and first PCA basis vector	11
Figure 1.6. Effects of optical density on the fluorescence intensity of quinine sulfate.	14
Figure 1.7. Various geometric arrangements for fluorescence measurement.	14
Figure 1.8. The electromagnetic spectrum which represents the complete range of electromagnetic radiation	15
Figure 1.9. Subtractive colors combine to form black (left) and additive colors reduce to produce white (right)	16
Figure 1.10. A comparison of the chromaticities enclosed by color spaces	17
Figure 1.11. PCA similarity map defined by the principal components 1 and 2 for the data table including tryptophan emission, retinol excitation and ANS excitation spectra.	18
Figure 1.12. Chemical structures of PPE polymers (P1-P6).....	18
Figure 1.13. Canonical score plot for the first three factors of fluorescence response patterns obtained with PPE polymer array against 17 protein analytes.....	19

Figure 1.14. Chemical structure of cationic gold nanoparticles (NP1–NP6) and anionic fluorescent polymer PPE-CO ₂ (m ≈ 12, where m refers to the number of repeated units in the polymer).	20
Figure 1.15. Displacement of quenched fluorescent polymer to restore the fluorescence signal (dark green strips, fluorescence off; light green strips, fluorescence on) by protein analyte (in blue).....	20
Figure 1.16. LDA score plot for the first two factors of fluorescence response patterns obtained from the nanoparticle–poly(<i>p</i> -phenyleneethynylene) assembly arrays against 7 proteins with identical absorption values of A = 0.005 at 280 nm (left) and the proteins at concentration of 5 μM (right).	21
Figure 1.17. Design of the PFBT conjugated q-MNP system.....	21
Figure 1.18. (left) LDA score plot of q-MNP polymer complex in the presence of bacteria (10 ⁷ cfu mL ⁻¹); (right) LDA score plot of q-MNP polymer complex in the presence of bacteria (OD600 = 0.2).	22
Figure 1.19. Image of the 36-dye colorimetric sensor array before exposure (left) and after exposure to decylamine (middle) after equilibration at full vapor pressure at 295 K.	23
Figure 1.20. Two principal components of the colorimetric sensor array from the response data averages of the 100 VOCs at 295 K, at their full vapor pressure.	23
Figure 1.21. (A) Color difference maps for 10 μM individual and mixed proteins; (B) PCA score plot for seven individual and mixed proteins.	24
Figure 1.22. (A) Color difference maps for 10 μM three natural and denatured proteins; (B) PCA score plots for six native and denatured proteins.	25
Figure 2.1. Structures of dendritic polyelectrolyte fluorescence compounds 1-6	28
Figure 2.2. Structures of diacetylene monomers 1-8	30

- Figure 3.1.** Cropped photographic images of the fluorophore solutions (2.0 μM) in phosphate buffer saline (10 mM, pH 7.4) mixed with each of protein solution ($A_{280} = 0.1$) under black light. 36
- Figure 3.2.** Histogram of logarithmic values of relative intensity ($\log I/I_0$) at 460 nm ($\lambda_{\text{ex}} = 375$ nm) of fluorophore solutions (0.20 μM) in phosphate buffer saline (10 mM, pH 7.4) mixed with each of protein solution ($A_{280} = 0.01$). 37
- Figure 3.3.** PCA score plot of ΔI obtained from the data set of 5 fluorophores \times 8 protein samples \times 9 replicates measured at 430 nm (a), 460 nm (b), 490 nm (c), 500 nm (d), 510 nm (e) and 520 nm (f). 38
- Figure 3.4.** PCA loading plot of ΔI measured at 500 nm obtained from the data set of 5 fluorophores \times 8 protein samples \times 9 replicates. 41
- Figure 3.5.** PCA score plot of ΔI measured at the excitation wavelength of 500 nm obtained from the dataset of the reduced sensing elements composing of (a) 2 and 5, and (b) 1 and 5. 42
- Figure 3.6.** (a) PCA score plot and (b) PCA loading plot of ΔI values of fluorophores 1-5 upon mixing with each pure milk protein. 45
- Figure 3.7.** PCA score plot of ΔI values in the range of 400-600 nm for 6 types of commercial milk. 46
- Figure 3.8.** Fluorescence responses (ΔI) of fluorophore (a) 1 and (b) 5 upon addition of pasteurized (P), sterilized (S), UHT (U) and recombined (R) milk; (c) PCA score plot of the fluorescence responses of 9 measurements (3 packages \times 3 replicates) of each type of milk. Oval outlines indicate groups of milk samples of the same thermal treatment. 48
- Figure 3.9.** Fluorescence responses (ΔI) of fluorophore (a) 2 and (b) 6 upon addition of pasteurized (P), sterilized (S), UHT (U) and recombined (R) milk; (c) PCA score plot of the fluorescence responses of 9 observations (3 packages \times 3 replicates) of each type of milk. Oval outlines indicate groups of the milk samples of the same thermal treatment. 50

- Figure 3.10.** PCA score plot of the fluorescence responses of 108 milk samples from pasteurized (P), UHT (U), sterilized (S) and recombined (R) milks..... 51
- Figure 3.11.** LDA score plot of first two discriminant factors (LD1 and LD2) obtained from fluorescence responses data of 108 milk samples from pasteurized (P), UHT (U), sterilized (S) and recombined (R) milk. Oval outlines indicate groups of milk samples of the same thermal treatment at 99% confidence level. 52
- Figure 3.12.** LDA correlation plot of (a) first discriminant factor (LD1) and (b) second discriminant factor (LD2) of fluorescence responses of **1, 2, 5** and **6** to milk samples at each emission wavelength. 53
- Figure 3.13.** Comparison of classification accuracy obtained from leave one out cross-validation method of the LDA of fluorescence responses of the sensor arrays consisting of four (**1+2+5+6**), three (**1+5+6**) and two (**5+6**) fluorophores..... 54
- Figure 3.14.** LDA score plot obtained from fluorescence responses of (a) **1+5+6** array and (b) 1,8-ANS in emission range of 420-560 nm. Oval outlines indicate groups of the milk samples of same thermal treatment at 99% confidence level..... 55
- Figure 3.15.** PCA score plot of ΔI values of fluorophores **1-3** upon mixing with each bacteria sample 59
- Figure 3.16.** LDA score plot of first three discriminant factors (F1, F2 and F3) obtained from fluorescence responses data of bacteria samples (a); LDA loading plot of ΔI values of fluorophores **1, 2, 5** upon mixing with each bacteria sample (b). 60
- Figure 3.17.** LDA score plot of first two discriminant factors (F1 and F2) obtained from ΔI values of fluorophores **2** and **5** upon mixing with each bacteria sample. Oval outlines indicate groups of bacteria sample at 95% confidence level 61
- Figure 3.18.** Photograph images of PDA sols (0.1 mM) in the absence (Blank) and presence of each surfactant at a concentration of 50 mM..... 63

- Figure 3.19.** UV/Vis spectra and photographs of PDA sols (0.1 mM) derived from a) **4**, b) **5** and c) **1+6** mixture in the presence of various amounts of SDC, SDS, and SDBS at room temperature showing their diverse colorimetric transitions..... 64
- Figure 3.20.** UV-visible spectrums of **4**, **5**, and **1+6** with TTAB, DTAB, and HTAB 65
- Figure 3.21.** UV-visible spectrums of **4**, **5**, and **1+6** with Tween-20®, Brij®58P, and Triton®X-100 65
- Figure 3.22.** Histogram of % CR obtained from nine replicated measurements of nine mixtures of **4**, **5** and **1+6** (0.10 mM) and SDC, SDS, SDBS, TTAB, DTAB, HTAB, Tween-20®, Brij®58P and Triton® X-100 (50 µM)..... 66
- Figure 3.23.** PCA score plot of % CR obtained from nine replicated measurements of nine mixtures of **4**, **5** and **1+6** (0.10 mM) and SDC, SDS, SDBS, TTAB, DTAB, HTAB, Tween-20®, Brij®58P and Triton® X-100 (50 µM)..... 67
- Figure 3.24.** (a) Scanned images and (b) photographs of the paper-based PDA sensor array prepared from diacetylene monomers **1-8** before and after 60 min exposure to saturated vapors of volatile organic solvents at 30 °C. All images were acquired at 150 dpi..... 69
- Figure 3.25.** Quantitative color change profile of paper-based PDA sensor array prepared from diacetylene monomers **1-8** after 60 min exposure to saturated vapors of volatile organic solvents at 30 °C obtained from (a) scanner and (b) digital camera. Error bars represent standard deviations of the intensity of each polymer. 70
- Figure 3.26.** PCA score plot of paper-based **1-8** PDAs sensor array derived from (a) scanner and (b) digital camera, upon exposure to 18 solvents. Each data point represents an average of RGB value obtained from 12 replications. 71
- Figure 3.27.** PCA loading plots of the eight element sensors in the developed arrays..... 72

Figure 3.28. PCA score plot of RGB color changes obtained from the paper-based PDA sensor array prepared from (a) 2 and 8 ; (b) 3 and 8 ; (c) 2 and 3 after exposure to 18 VOC vapors.....	73
Figure 3.29. The commercial gasoline in various grades from Petroleum Authority of Thailand (A) and Esso company (B).	74
Figure 3.30. Scanned images of the PDA-based paper colorimetric sensor array prepared from diacetylene monomers 1-8 after exposure to vapors of gasoline.	75
Figure 3.31. RGB color change profile of paper-based PDA sensor array prepared from diacetylene monomers 1-8 after exposure to saturated vapors of automotive fuels.....	76
Figure 3.32. PCA score plot of Δ RGB data set obtained from paper-based PDA sensor array derived from diacetylene monomers 1-8 upon exposure to 10 gasoline. Each data point represents an average of RGB value obtained from 9 replications.....	77
Figure 3.33. PCA score plot of RGB color changes obtained from the paper-based PDA sensor array derived from 2 and 3 upon exposure to automotive fuels (2 PDAs \times 3 replicates \times 10 automotive fuels \times 3 measurements of each RGB).....	78
Figure A1 LDA score plot obtained from fluorescence responses in emission range of 420-560 nm of (a) 6 at 99.07 % correct classification and (b) 1,8-ANS at 97.84 % correct classification. Oval outlines indicate groups of the milk samples of same thermal treatment at 99% confidence level.....	93
Figure A2 LDA score plot of first two discriminant factors (F1 and F2) obtained from Δ I values of fluorophores 1 , 2 and 5 upon mixing with each bacteria sample	94
Figure A3. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 10 μ M	94
Figure A4. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 20 μ M	95

Figure A5. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 30 μM	95
Figure A6. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 40 μM	96
Figure A7. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 50 μM	96
Figure A8. PCA score plot of paper-based 1-8 PDAs sensor array derived from (a) RGB value and (b) CMYK value from scanner upon exposure to 18 solvents. Each data point represents an average of color values obtained from 12 replications.	97
Figure A9. PCA score plot of paper-based 1-8 PDAs sensor array derived from (a) RGB value and (b) CMYK value from digital camera upon exposure to 18 solvents. Each data point represents an average of color values obtained from 12 replications.....	98

CHAPTER I

INTRODUCTION

1.1 Overview

Biosensor applications have recently expanded as the promising tool generating a rapid, inexpensive and sensitive response based on colorimetric or fluorescent changes, and having multiplex capabilities where many traditional sensing systems incapable. It is a tool that converts a biological response from a stimulus into a measurable and readable output signal [1] The basic principle of a biosensor is to detect a biological recognition and transform it into a type of signal via a transducer. **Figure 1.1** shows typical elements of biosensor [2]: a) bioreceptors that specifically bind to the analyte via utilization of a biochemical mechanism for recognition; b) an interface where signal is generated and recognized by the interaction between the bioreceptors and the analyte; c) the physical transducer where the transducer signal is converted to a measurable signal, as an electrical signal, for data acquisition and processing; d) computer software for the data analysis which convert the raw data into an easily understandable form; e) the visual display of an outcome - an interface to the human operator. Biosensors can be applied to a large variety of samples including body fluids, food samples, cell cultures and environmental samples. A bioreceptor can be a biological molecular species (such as enzyme, protein, antibody, nucleic acid or chemical compounds in biological samples) or a living biological system (such as a tissue, microorganism, organelle or cell).The transduction may be optical, electrochemical, thermometric, piezoelectric, magnetic and micromechanical or combinations thereof.

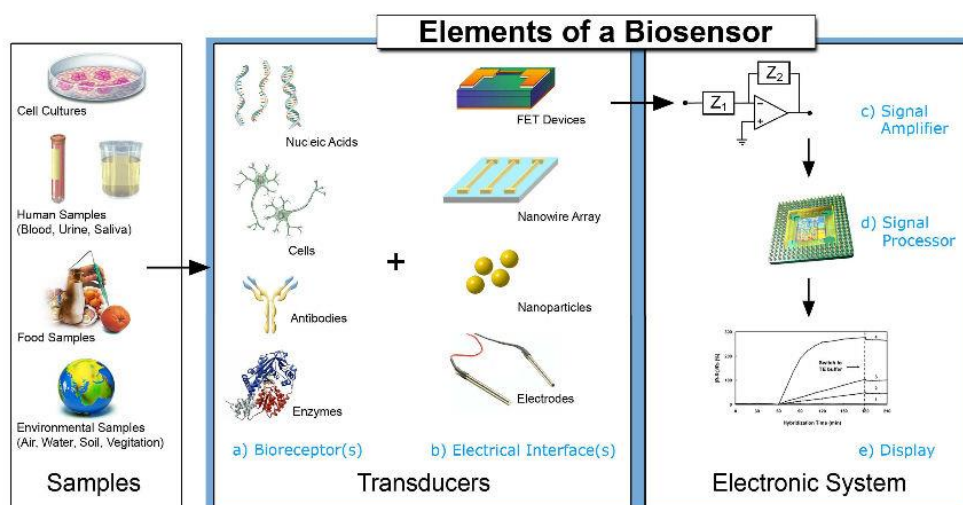


Figure 1.1 Elements and selected components of a typical biosensor

The biosensor array using spectroscopic techniques is designed to detect multiple samples in the presence of several analytes consisting of several hundred detectable signals throughout the dataset. Especially with emerging technology, the capacity of analytical instruments has been improved to acquire large amounts of data rapidly. This leads to very complex datasets which requires powerful data acquisition and analysis systems. Thus, the challenge of biosensor array is the way to improve discriminatory powers capable to differentiate constituents of interest in complex systems. Chemometrics could help to perform data mining to extract useful information from the complex spectral datasets.

1.2 Chemometrics

Most chemical measurements are inherently multivariate in which data sets consisted of many variables with overlapped information obtained from one sample in time. The spectroscopy is an obvious example since a spectrum at hundreds of wavelengths can be measured on a single sample. The chemical information contained in the spectral data is hidden in the band position, the band intensities and the bandwidths. Whereas the band positions have information about the chemical structures, the intensities of the bands are related to the quantity of the compounds in a mixture [3]. The easiest way to determine the content of a chemical

compounds is to measure the change of the intensity of a well-resolved band that clearly belongs to this compound which is only possible for a pure component. However, biological samples retain numerous components giving complex spectroscopic spectra which typically composed of broad overlapping spectral bands containing chemical, physical, and structural information of all sample components, the analytical information contained in spectra is multivariate in nature and, sometimes, non-selective. Traditional approaches as *univariate* can miss much information because each variable is considered individually, for example only one measurement is carried out on a candidate compound (bond lengths, dipole moments, bond angles, etc.). In this instance, chemometric with multivariate statistical analysis allows us to explore multiple variables and their correlation simultaneously [4].

The term “chemometrics” was first introduced by a Swede researcher, Svante Wold, and an American researcher, Bruce R. Kowalski in the early 1970s [5] and the International Chemometrics Society was established shortly thereafter by them, the two pioneers in the field. Since then, chemometrics has been defined in broad terms as the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods [6].

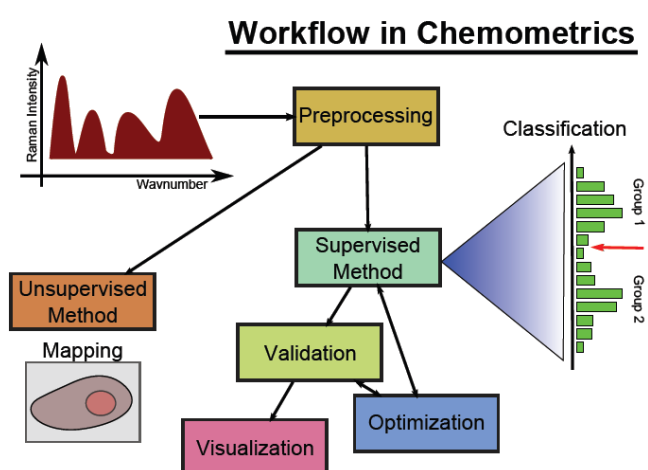


Figure 1.2. Principle of Chemometrics [7]

Chemometric analysis, the mathematical method of extracting information from chemical systems by data-driven means, is a powerful method to extract highly correlated data from a large amount of correlated and non-correlated data and to group the samples with similar characteristics and then to establish classification for samples. By this analysis, both descriptive and predictive issues can be solved. In the descriptive approach, properties of the investigated data are extracted in order to learn the underlying interrelationships hidden in variables. In the predictive approach, the ability of the discriminant to predict the class of an unknown sample will be tested by validation technique [8].

The analytical information contained in spectroscopic data can be extracted by using various multivariate analysis techniques that relate analytical signals to the properties of the analytes. These methods encompass descriptive techniques such as the principal component analysis (PCA) and canonical correlation analysis (CCA) and predictive techniques such as linear discriminant analysis (LDA) and principal component regression (PCR). Several multivariate statistical analyses can be applied to the spectroscopic dataset as a multivariate data matrix of substances in real samples to demonstrate the discriminating ability of the resulting array. To discover the relationships among all responses and variables obtained from a sample and distinguish them from the interfering signals efficiently, all of the data must be processed simultaneously.

1.2.1. Pattern recognition

Human's brain normally has the ability to recognize either tangible or intangible objects in the environment and learn to distinguish patterns of interest, such as a human face, fingerprint image, a speech signal, or food flavor from what they have previously learned or perceived. Pattern recognition in the context of chemometrics is an approach to machine intelligence which is based on statistical modeling of high-dimensional data that may not be visualized by human's perception. For example, spectroscopy sensor array is a group of sensors (consisting of many sensing elements and/or many wavelengths) which provides a dataset of

interrelated spectral bands. This spectral dataset has multiple measured variables and is beyond human perception. The computer-based pattern recognition can be divided into two techniques [9, 10]: (1) supervised classification in which the input is identified as a member of a predefined class, and (2) unsupervised classification in which the input is assigned to an unknown class.

1.2.1.1 Unsupervised pattern recognition

An unsupervised method does not require any a priori knowledge about the group structure in the data, but instead produce the grouping, i.e. clustering, themselves based on similarities and dissimilarities between samples. This type of analysis is often very useful at an early stage of the investigation to explore subpopulations in a data set, e.g. different freshness of products. Cluster analysis can be performed for better visualization using a method such as Principal Component Analysis (PCA).

Principal component analysis (PCA)

The first step of chemometrics analysis begins with the descriptive approach to obtain an overview of all the information in the dataset by performing an exploratory analysis in the multivariate space. Principal Component Analysis (PCA) is probably the most widespread multivariate statistical technique used in this approach [4]. Since PCA is unsupervised pattern recognition, it does not require any a priori knowledge about the group structure in the data [11, 12], but instead self-organize the grouping, i.e. natural clustering. This type of analysis is often very useful at an early stage of the investigation to explore subpopulations in a data set, e.g. different freshness of a product.

PCA is the appropriate choice for data investigation by calculating orthogonal eigenvectors (principal components, PCs) that lie in the direction of the maximum variance within that data set. The first PC contains the highest degree of variance and other PCs follow in the order of decreasing

variance. PCA may be used to test the quality of the data because it will equally and statistically treat all of the data points. Therefore, PCA will define the clusters based on their data similarity and would attest to the high discriminatory ability of the array.

The main goal of PCA is to reduce a large number of variables to a much smaller number of principal components (PCs) that capture the vast majority of variance in the data [13, 14]. This reduces the dimensionality of the data considerably, enabling effective perception, visualization, regression and classification of multivariate data. With this technique, relationships between different parameters (samples and variables) and/or the detection of possible clusters within the samples and/or variables can be found. PCA performed on biosensor response data makes it possible to draw the similarity maps of the samples and to get the response patterns.

PCA decomposes the data matrix as follows [11]:

$$\mathbf{X}_{ij} = \mathbf{T}_A \cdot \mathbf{V}_A^T + \mathbf{E}_{ij} \dots\dots\dots(1.1)$$

where A is the number of components; $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A]$ is the score vectors which give the coordinates of samples in the PC space; $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_A]$ is the loadings vectors which indicate the importance of each original variable contributing to the principal components; \mathbf{E} is the residuals or noise, the part of the data which was not explained by the model.

The steps of the PCs selection are as follows:

1. The center of the data cloud is determined as the zeroth principal component (PC0).
2. The direction of the maximum data variation is relatively sought toward from the centered origin, PC0. This is the PC1.

3. If the data are not fully described, more directions (PC2, PC3...) orthogonal to the first PC axis (or its earlier PC) are determined in order to describe the maximum variation in the residuals.

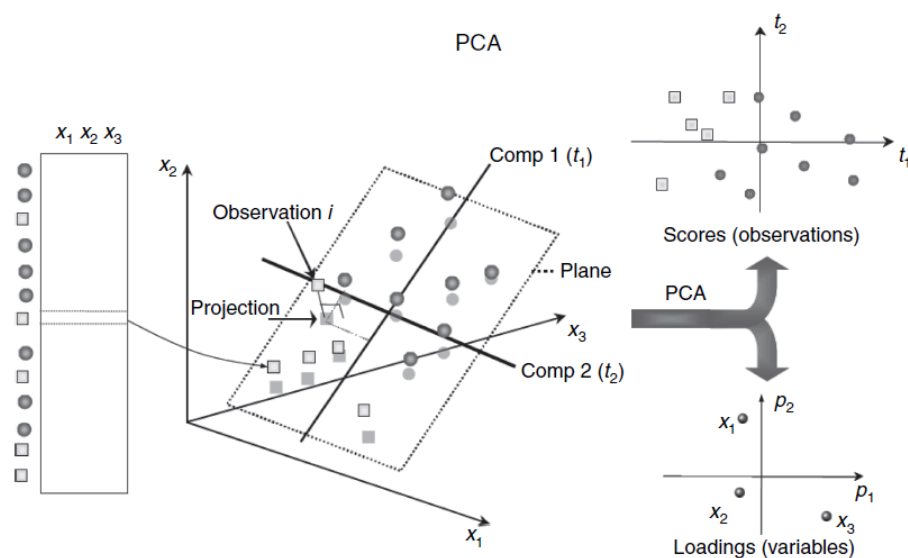


Figure 1.3. A principal component analysis (PCA) model approximates the variation in a multi-dimensional data table by a lower dimensional model plane. The directions in the PC score plot correspond to directions in the loading plot, and vice versa [12].

1.2.1.2 Supervised pattern recognition

In these methods, each data point is formerly assigned to a definite class. The distance between classes characterizes the partition obtained. The validity of the method can be verified by comparison of the distances. The distances between classes means have to be clearly superior to the distances within classes. Another way to validate the discrimination is to test it by cross-validation method which the data will be separated into two sets: a training set to elaborate the method (calibration), and a test set to validate it. The classification rules are later used for allocating new or unknown samples to the most probable subclass.

Linear discriminant analysis (LDA)

LDA is useful whenever there are grounds to postulate the existence of a number of groups (categories) into which the samples may be classified by modeling the similarities between the data corresponding to the same cluster by introducing the group classification of the trials into the data set. The data are then used as a training set to generate a linear discriminant (LD) function, which describes the best fit parameters to separate different clusters (analyses).

LDA is supervised where the class membership has to be known in advance for the analysis. LDA is similar to PCA because they are both considered as a dimension reduction method by determining a smaller dimension hyperplane on which the points will be projected from the higher dimension. However, whereas PCA selects a direction that retains maximal structure among the data in a lower dimension, LDA selects a direction that achieves maximum separation among the given classes [15]. LDA is based on the determination of linear discriminant functions, which maximize the ratio of between-class variance and minimize the ratio of within-class variance, by seeking a set of optimal vectors, denoted by $W = [w_1, w_2, \dots, w_i]$.

$$J(W) = \text{tr} \left(\frac{W^T S_b W}{W^T S_w W} \right) \dots\dots\dots(1.2)$$

where S_B is the “between-class scatter matrix”, the distance between the centroids of different classes; and S_W is the “within-class scatter matrix”, the accumulated distance of an instance to the centroid of its class.

The definitions of the scatter matrices are:

$$S_b = \sum_{i=1}^C p_i (m_i - m)(m_i - m)^T$$

$$S_W = \sum_{i=1}^C \sum_{j=1}^{n_i} p_i (x_{ij} - m_i)(x_{ij} - m_i)^T$$

where p_i is the prior probability of the i^{th} class, m_i is the centroid of class i , x_{ij} is the j^{th} sample of class i and n_i is the number of training samples from class i , m is the centroid of the global centroid.

Listed below are the general steps for performing a linear discriminant analysis [16]:

1. Seeking the d-dimensional mean vectors;
2. Computing the Scatter Matrices S_B and S_W ;
3. Computing the eigenvectors and corresponding eigenvalues for the scatter matrices.
 - Solving the generalized eigenvalue problem for the matrix $S_W^{-1} S_B$ to obtain the linear discriminants;
 - Checking whether the eigenvector-eigenvalue calculation is correct and satisfy
4. Selecting linear discriminants (LD) for the new feature subspace;
 - Sorting the eigenvectors by decreasing eigenvalues
 - Choosing k eigenvectors with the largest eigenvalues
5. Transforming the samples onto the new subspace.
 - $y = W^T \times X$ where x is a $d \times 1$ -dimensional vector representing one sample, and y is the transformed $k \times 1$ -dimensional sample in the new subspace.

In LDA, classes are supposed to follow a multivariate normal distribution and be linearly separated. LDA can be considered, as PCA, as a feature reduction method in the sense that both, LDA and PCA, determine a smaller dimension hyperplane on

which the points will be projected from the higher dimension. However, whereas PCA selects a direction that aims to retain maximal structure among the data, LDA selects a direction that aims to achieve maximum separation among the sample classes [17, 18] (see **Figure 1.4**).



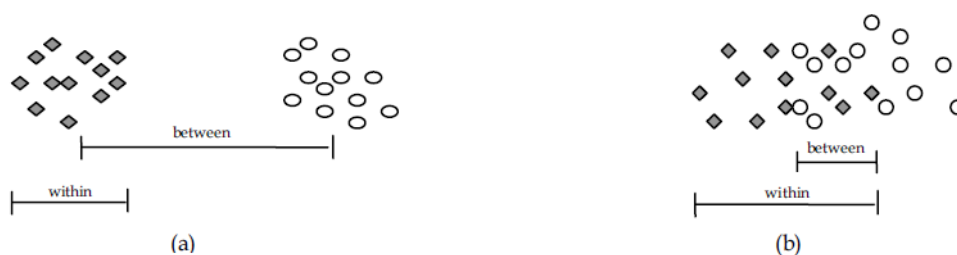


Figure 1.4. (a) Good class separation. (b) Bad class separation.

In these methods, each data point is formerly assigned to a definite class. The distance between classes characterizes the partition obtained. The validity of the method can be verified by comparison of the distances. The distances between classes means have to be clearly superior to the distances within classes. As shown in **Figure 1.5** [19], two classes (red and blue colors) are well separated by a projection onto the first LDA basis vector, but poorly separated by a projection onto the first PCA basis vector.

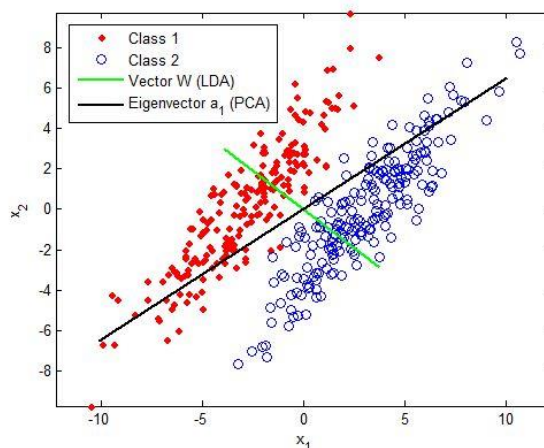


Figure 1.5. 2-D 2-class data, along with first LDA basis vector and first PCA basis vector

1.2.2 Feature selection

Spectroscopic measurements carried out in chemistry or biotechnology usually contain both meaningful variables and meaningless noise [20]. Noisy variables

increase chances of false classification. The noise reduction by feature selection should be taken into account to remove the number of irrelevant or redundant variables for improvement of the overall sensor's quality. The feature selection aims to gain a better understanding of the essential features that play an important contribution in governing the sensor behavior under investigation [21].

The feature selection is usually based on the cross-validation [22] which is a procedure to estimate the performance of a model which the data training set is divided into two sets: a training set to elaborate the method, and a test set to validate it. Leave-one-out routine is a type of the cross-validation where procedure consists of removing one sample out of the training set at the time, constructing the predictor on the basis only of the rest of the data which will be used as the training data, then testing on the removed sample. Since a priori information of sample groups is required to perform the predictive analysis, only supervised method is able to do.

1.3 Optical Biosensor Array

Biosensors can be classified by either a type of bioreceptor or a transducer employed. For the optical biosensor, the sensor is based on various technologies of optical phenomena (i.e. fluorometric or colorimetric changes) which are a result of both interactions of a recognition center binding with target analytes and chromophore or fluorophore coupling to the recognition center [1, 23].

1.3.1 Fluorescence Biosensor

Fluorescence is a process in which sensitive molecules emit light from electronically excited states induced by absorption of light. The tremendous improvements in fluorescence-based techniques can be observed over the last decades. Fluorescence sensor techniques can be categorized into three main classes: (1) intrinsic fluorescence such as aromatic amino acids, neurotransmitters, porphyrins, and green fluorescent protein, are those that occur naturally; (2) extrinsic fluorescence in which synthetic dyes or modified biochemicals are added to a

specimen to produce fluorescence with specific spectral properties; and (3) displacement or differential probes that is the displacement of a fluorogenic indicator from an indicator–receptor complex by the analyte [1].

Some biological samples rely on their intrinsic fluorescence characteristics. For example, milk contains intrinsic fluorophores such as vitamin A in fats and amino acid residues present and proteins [24, 25]. These Intrinsic fluorophores are sensitive to environment and are able to provide information on conformational changes of proteins [3, 26]. In addition, tryptophan residues in hydrophobic interior of protein-containing sample provide fluorescent properties different from residues on a hydrophilic surface. However, it is not always the case the intrinsic fluorescence is adequate for the desired investigation, extrinsic fluorescent dyes may be used in this case.

There are at least two parameters governing the appearance of fluorescence intensity and spectral distribution i.e. the optical density of the sample and the geometry of sample illumination. The fluorescence intensities are proportional to the sample concentration only in a limited range of optical densities. **Figure 1.6** shows the effects of optical density on the fluorescence intensity of quinine sulfate obtained in a 1-cm² cuvette that was centrally illuminated [27]. The solid line (—) is the measured intensities, while the dashed line (– –) is the corrected intensities. High optical densities of the sample (the absorbance of the sample exceeds 0.1) can distort the emission spectra as well as the apparent intensities when it is recorded by the conventional right-angle fluorescence measurement from the center of a centrally illuminated cuvette (**Figure 1.7**, top left). A dilution of turbid samples is sometime not a proper choice of solution since the organization of the sample matrix can be lost, especially in food products. To avoid these problems, front-face illumination geometry with the incidence angle of the excitation radiation set at 30^o to 60^o performed using either triangular cuvettes or square cuvettes is performed to reduce light scattering effects [27] (**Figure 1.7**).

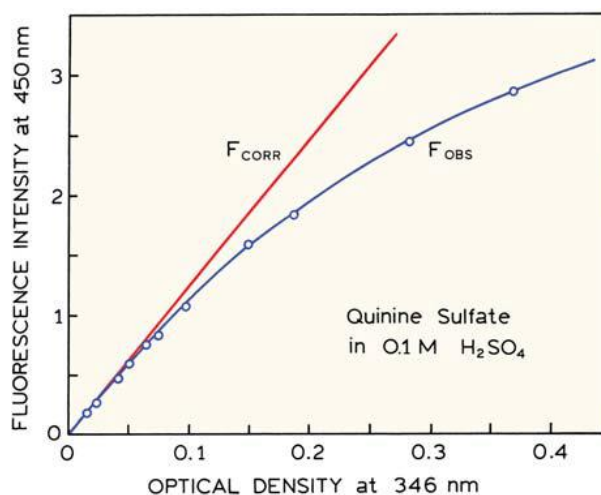


Figure 1.6. Effects of optical density on the fluorescence intensity of quinine sulfate.

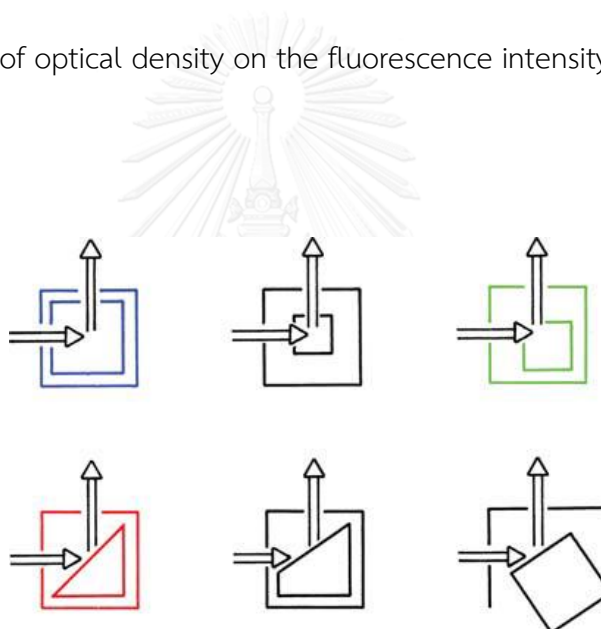


Figure 1.7. Various geometric arrangements for fluorescence measurement.

1.3.2 Colorimetric Sensor

Human's eye perceive variety of colors derived from the spectrum of light. The chemical basis of human vision depends upon the absorption of light (electromagnetic radiation) by pigments in the eye. Rod in human's eye is analogous to transducers which transform electromagnetic energy into the chemical energy and stimulate an impulse and carry to the brain via the optic nerves [28]. This visible light

region for human consists of a spectrum of wavelengths ranging from approximately 700 nm to approximately 400 nm (Figure 1.8).

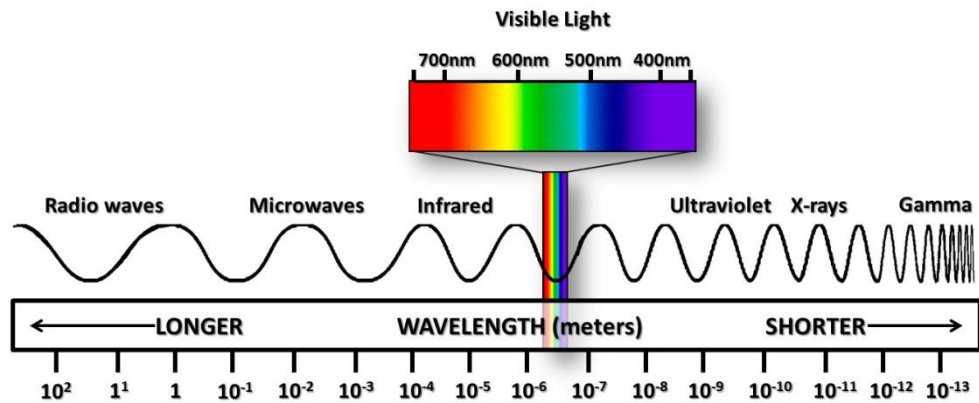


Figure 1.8. The electromagnetic spectrum which represents the complete range of electromagnetic radiation [29].

Table 1.1. The colors of the visible light spectrum [30]

Color	Wavelength Interval	Frequency Interval
Red	~ 700–635 nm	~ 430–480 THz
Orange	~ 635–590 nm	~ 480–510 THz
Yellow	~ 590–560 nm	~ 510–540 THz
Green	~ 560–490 nm	~ 540–610 THz
Blue	~ 490–450 nm	~ 610–670 THz
Violet	~ 450–400 nm	~ 670–750 THz

Primary colors are sets of fundamental colors that can be simultaneously combined to make a useful range of colors. Primary colors (**Figure 1.9**) can be divided into two types: additive and subtractive.



Figure 1.9. Subtractive colors combine to form black (left) and additive colors reduce to produce white (right) [31].

The additive primaries are those which are obtained by light: red (R), green (G) and blue (B). The sets of RGB colors combine to form white and form the basis of screen color such as computer screen, television, scanners, digital cameras and other devices that give light. Subtractive primaries are those obtained by the subtraction of light: cyan (C), magenta (M) and yellow (Y). These colors combine to form black (or K) and is the basis of printing that reflect light rather than emit light. RGB / CMYK color models are based on different principle, the spaces of colors in CMYK and RGB are different as shown in **Figure 1.10**.

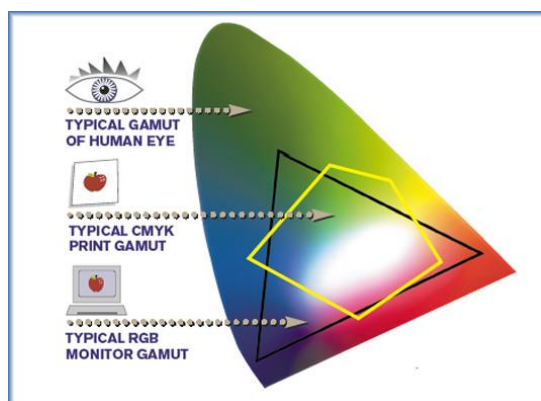


Figure 1.10. A comparison of the chromaticities enclosed by color spaces [32].

1.4 Literature reviews

1.4.1 Fluorescence biosensor array

In 1997, Dufour et al. [33] applied PCA for discrimination of raw, heated and homogenized milks using fluorimetric data. The fluorescence spectra of raw (NHO), heated (NHP), homogenized (HOM) and homogenized + heated (HOP) milks were recorded using a variable angle front-surface accessory having the incidence angle of the excitation radiation set at 56° , in the presence of 1-anilinonaphthalene-8-sulfonic acid (ANS) as extrinsic fluorophore. The emission spectra of tryptophan (305-400 nm), retinol (350-500 nm) and ANS (400-600nm) were recorded with excitation wavelengths set at 290, 321 and 370 nm, respectively, and the excitation spectra of retinol (260-350 nm) and ANS (250-450 nm) were recorded with emission wavelengths set at 410 and 466 nm, respectively. Principal component analysis was applied to the normalized fluorescence spectral data in order to distinguish between milk samples. It was shown that the map defined by principal components 1 and 2 (**Figure 1.11**) discriminated NHO, NHP, HOM and HOP samples as a function of homogenization and heating, respectively. Notably, the measurements by recording the excitation and emission spectra on different wavelengths as used in this study are inconvenient for a real application.

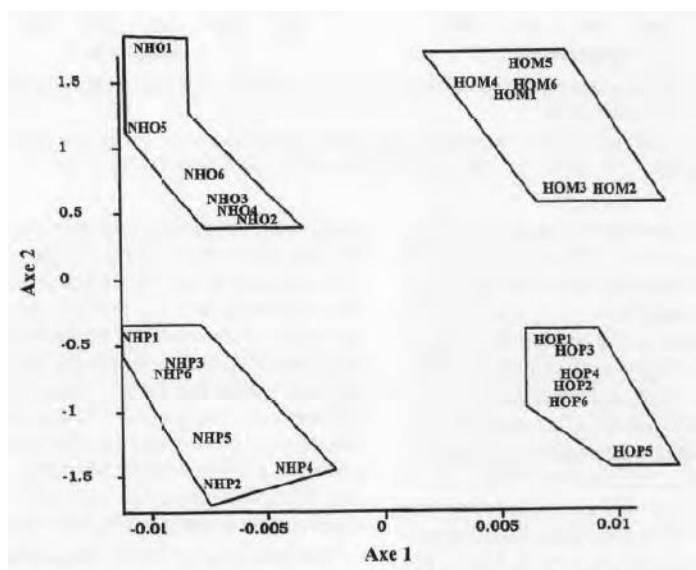


Figure 1.11. PCA similarity map defined by the principal components 1 and 2 for the data table including tryptophan emission, retinol excitation and ANS excitation spectra.

In 2007, Miranda et al. [34] used functionalized poly(p-phenyleneethynylene)s (PPEs) to build a protein sensor array for proteins. These highly fluorescent polymers possess various charge characteristics and molecular sizes (**Figure 1.12**).

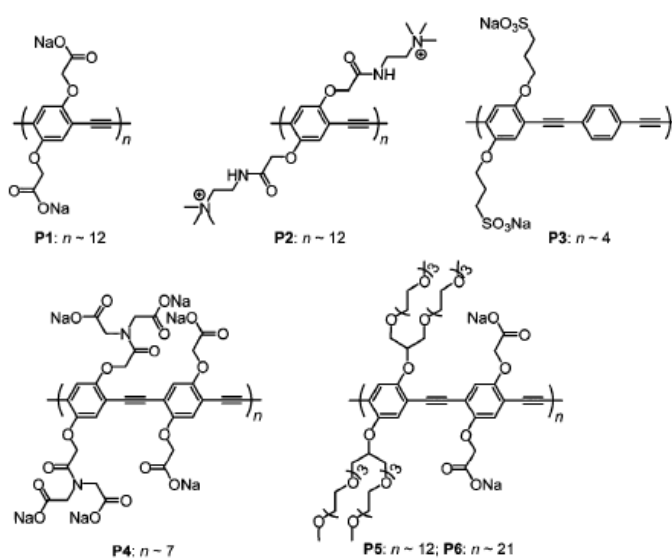


Figure 1.12. Chemical structures of PPE polymers (P1-P6)

However, they display substantial overlap in their absorption and emission spectra. Therefore, the fluorescence response patterns were subjected to linear discriminant analysis (LDA). The first three canonical factors contain 65.0, 20.8, and 7.3% of the variation, respectively, occupying 93.1% of total variation. The patterns are clustered into 17 different groups corresponding to the numbers of proteins as shown in **Figure 1.13**, and the cross-validation reveals 100% classification accuracy.

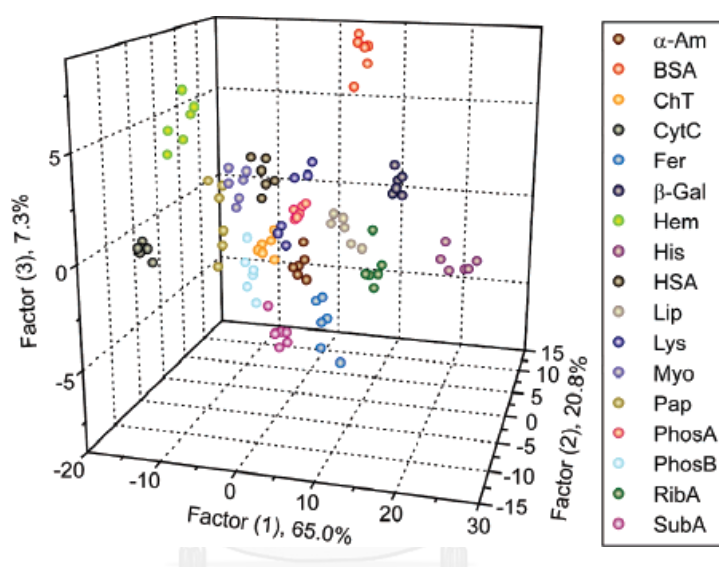


Figure 1.13. Canonical score plot for the first three factors of fluorescence response patterns obtained with PPE polymer array against 17 protein analytes.

In 2007, You et al. [35] developed a displacement fluorescent sensor array containing six non-covalent gold nanoparticle–fluorescent polymer conjugates (**Figure 1.14**) for the identification of proteins.

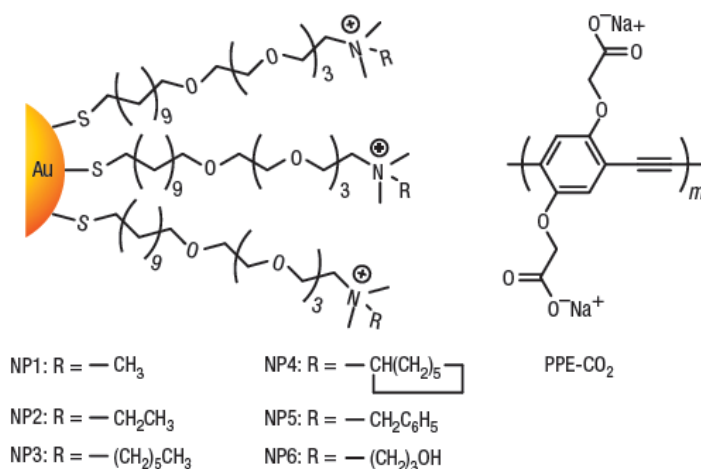


Figure 1.14. Chemical structure of cationic gold nanoparticles (NP1–NP6) and anionic fluorescent polymer PPE-CO₂ ($m \approx 12$, where m refers to the number of repeated units in the polymer).

Seven proteins having diverse molecular weight and isoelectric point (pI) were selected as the target analytes in this study. The nanoparticles associated with fluorescent dye produce the fluorescence "switch-off" and the subsequent binding of protein analytes displaces the dyes to generate the fluorescence "switch-on" as shown in **Figure 1.15**.

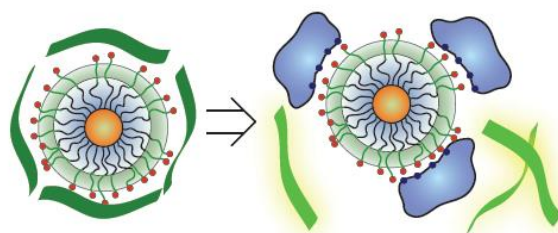


Figure 1.15. Displacement of quenched fluorescent polymer to restore the fluorescence signal (dark green strips, fluorescence off; light green strips, fluorescence on) by protein analyte (in blue).

A fluorogenic response patterns were generated from the sensor array obtained from the addition of proteins ($5 \mu\text{M}$ or $A_{280} = 0.005$) to the polymer conjugates. LDA was applied to the raw data of fluorogenic response patterns. As shown in **Figure**

1.16, LDA can distinguish each protein response pattern with an accuracy of 100% and 94.2% respectively.

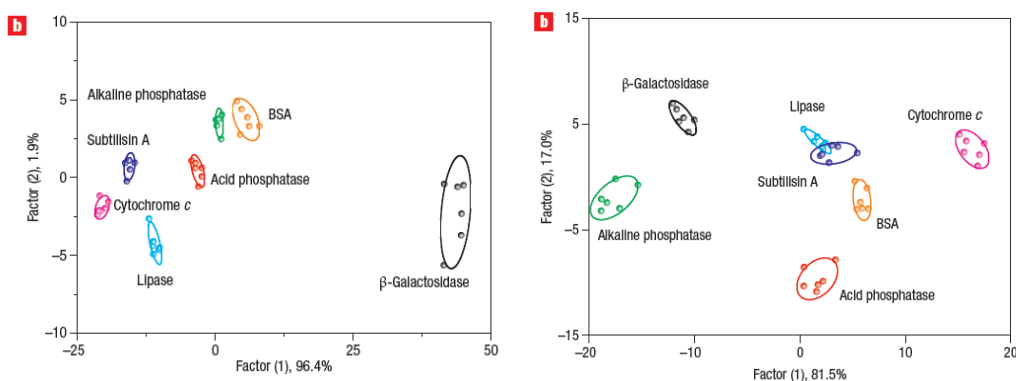


Figure 1.16. LDA score plot for the first two factors of fluorescence response patterns obtained from the nanoparticle–poly(*p*-phenyleneethynylene) assembly arrays against 7 proteins with identical absorption values of $A = 0.005$ at 280 nm (left) and the proteins at concentration of $5 \mu\text{M}$ (right).

In 2014, Wan et al. [36] designed a ‘chemical nose’ sensor array containing three quaternized magnetic nanoparticles (q-MNPs)–fluorescent polymer systems for bacteria detection. The fluorescence intensity at 538 nm was measured with an excitation wavelength of 486 nm. The response intensity of the array differentiates upon the level of displacement determined by the relative q-MNP–fluorescent polymer binding strength and bacteria cells–MNP interaction (**Figure 1.17**).

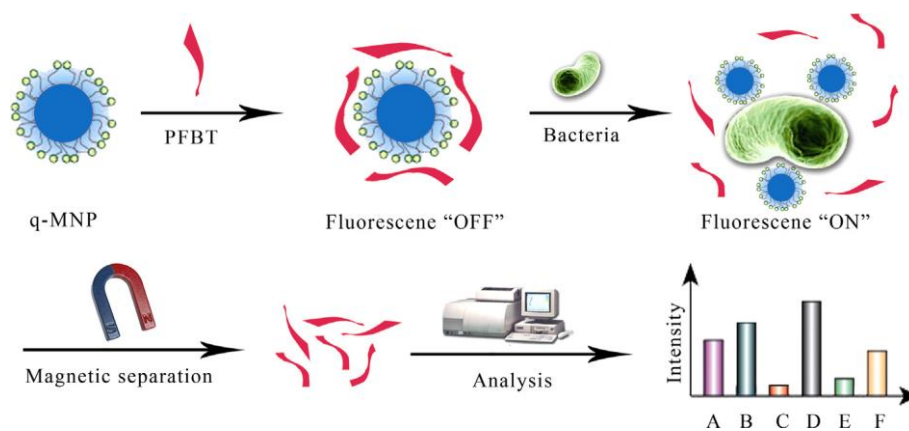


Figure 1.17. Design of the PFBT conjugated q-MNP system.

Two types of bacteria concentrations, 10^7 cfu mL⁻¹ and OD600 = 0.2 were prepared for comparison of the sensor's effectiveness. The fluorescence response data matrices were analyzed by LDA technique relatively low accuracy of 96.8% and 87.5% respectively (Figure 1.18).

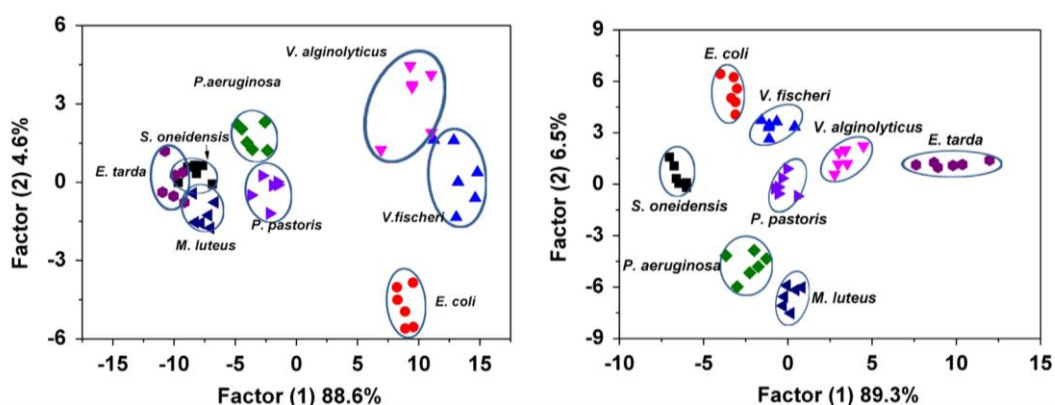


Figure 1.18. (left) LDA score plot of q-MNP polymer complex in the presence of bacteria (10^7 cfu mL⁻¹); (right) LDA score plot of q-MNP polymer complex in the presence of bacteria (OD600 = 0.2).

1.4.2 Colorimetric sensor array

In 2006, Janzen et al. [37] developed a colorimetric sensor array for the detection and identification of volatile organic compounds. The array composed of chemically responsive dyes which produce digital images as a distinct map showing a characteristic fingerprint for each analyte. The digital images are commonly achieved with an ordinary flatbed scanner. Figure 1.20 shows the color map which is generated by RGB subtraction of the array image before and after exposure for red (R), green (G) and blue (B) values. Every spot in the array is described by RGB color values; for an eight-bit color scanner, this spans the range of 0-255: i.e., black color is (0, 0, 0) and white color is (255, 255, 255). The subtraction of the “before” and “after” images yields a vector in 108 dimensions (i.e., 36 changes from six by six array in RGB color values). Each vector is ranging from a number of -510 to +510 and usefully visualized

as shown in the difference map (Figure 1.19, right), which shows the absolute values of the color changes.

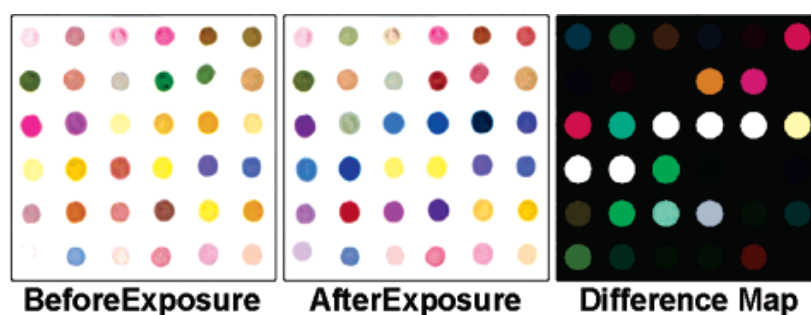


Figure 1.19. Image of the 36-dye colorimetric sensor array before exposure (left) and after exposure to decylamine (middle) after equilibration at full vapor pressure at 295 K.

Figure 1.20 shows 53% of the discriminatory ability of the array captured by the first two components. As an inherent consequence, only fair spatial discrimination among chemical classes can be observed. The clusters of hydrocarbons, thiols, ketones, esters, and aldehydes are significantly overlapped, while phosphines cannot be distinguished from aromatic amines. The discrimination of analytes into their chemical classes is limited.

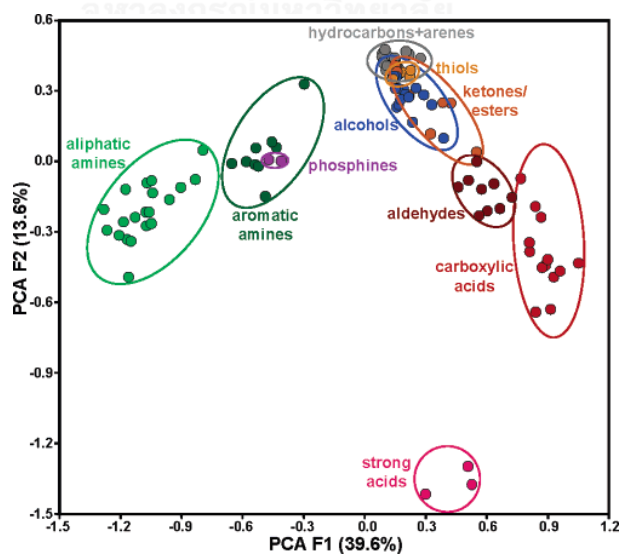


Figure 1.20. Two principal components of the colorimetric sensor array from the response data averages of the 100 VOCs at 295 K, at their full vapor pressure.

In 2011, Hou et al. [38] developed an artificial tongue system through a low-cost and simple colorimetric sensor array consisting of porphyrin, porphyrin derivatives (mainly metalloporphyrins) and chemically responsive dyes capable of rapid interaction with proteins. The array produced color patterns in response to each protein differently. A colorful difference map was obtained by comparing the RGB values of “initial” and “final” images. The color patterns coupled with PCA allow the identification of the pure and mixed proteins as shown in **Figure 1.21** and denatured proteins as shown in **Figure 1.22**.

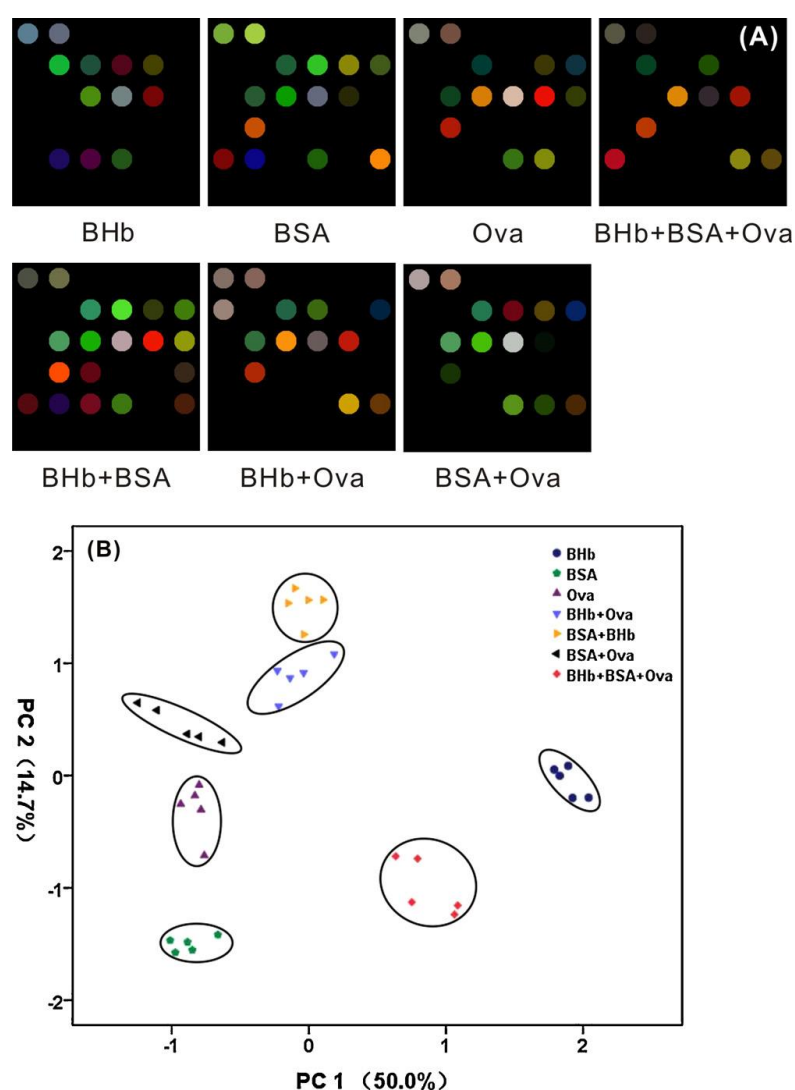


Figure 1.21. (A) Color difference maps for 10 μ M individual and mixed proteins; (B) PCA score plot for seven individual and mixed proteins.

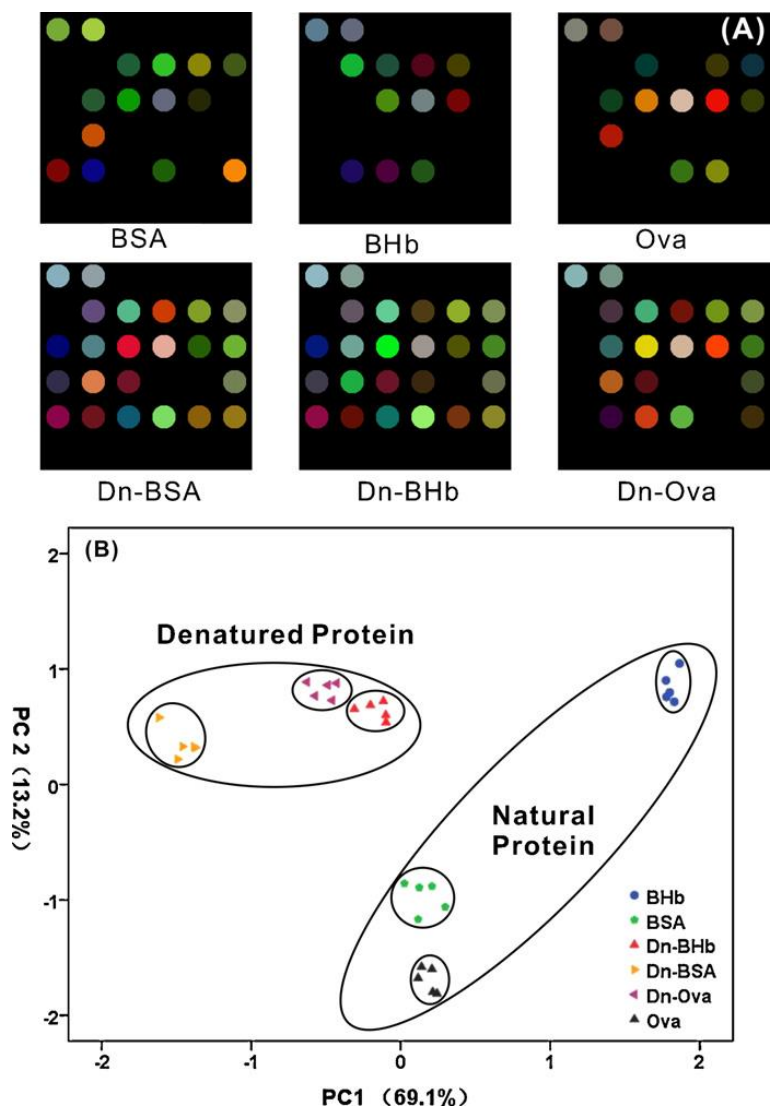


Figure 1.22. (A) Color difference maps for 10 μ M three natural and denatured proteins; (B) PCA score plots for six native and denatured proteins.

1.5 Objectives and scope of the dissertation

1. To apply chromophore or fluorophore compounds as sensing arrays for organic substances in biological samples
2. To obtain the optimum sensing array through multivariate statistical analyses

To achieve the above objectives, the scope of work includes:-

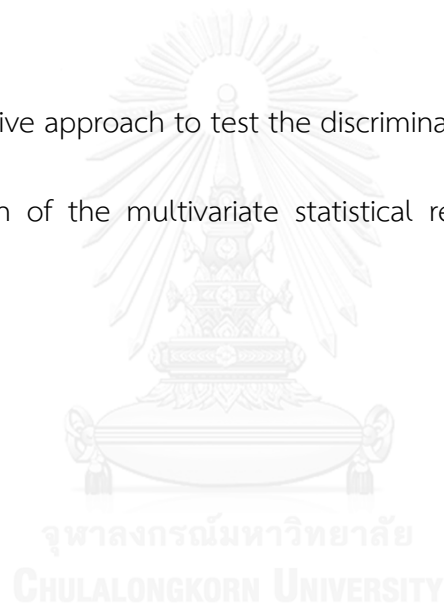
1. Spectroscopy measurements of the chromophores or fluorophores in responses to samples

2. Chemometric data acquisition and analysis using multivariate statistical analyses.

2.1 A descriptive approach to obtain an overview of all the information in the data set

2.2 A predictive approach to test the discriminating ability of sensor array

3. Interpretation of the multivariate statistical results obtained from sensing array



CHAPTER II

EXPERIMENTAL

2.1 Fluorescence sensor array

2.1.1 Materials

Concanavalin A (ConA, from Jack bean), cytochrome C (CytC, from equine heart), histone (His, from calf thymus, type III-S), human serum albumin (HSA), lysozyme (Lys, from chicken egg white), myoglobin (Myo, from equine heart) and papain (Pap, from papaya latex), α -Casein (α -CN), β -Casein (β -CN), α -Lactalbumin (α -LA), β -Lactoglobulin (β -LG) and 1-Anilinonaphthalene-8-sulfonic acid (1,8-ANS) were purchased from Sigma. Bovine serum albumin (BSA) was purchased from Fluka. All commercial milk products were purchased from local supermarkets and used well before the expiration dates. However, life period of the products was not taken into account. All solutions were prepared by using Milli-Q water (18.1 M Ω) as the solvent. All chemicals were used as received without further purification.

Vibrio cholera, *Shigella flexneri*, *Listeria monocytogenes*, *Bacillus cereus* ATCC 14579, *Salmonella Typhimurium* ATCC 11331, *Staphylococcus aureus* ATCC 25923 and *Escherichia coli* ATCC 25922 were kindly provided by Associate Professor Chanpen Wiwat, Faculty of Pharmacy, Mahidol University, Thailand. *Enterotoxigenic Escherichia coli* (ETEC) 290 was kindly provided by Assistant Professor Potjane Srimanote, Faculty of Allied Health Science, Thammasat University, Thailand. All solutions were prepared by using PBS buffer (pH 7.4) as the solvent. All chemicals were used as received without further purification. Fluorescence spectra were acquired from a SpectraMax M2 microplate reader (Molecular Devices, Sunnyvale, CA) using black polystyrene 96-well microplates.

Dendritic polyelectrolyte fluorescence compounds were prepared according to previous reports [39-41].

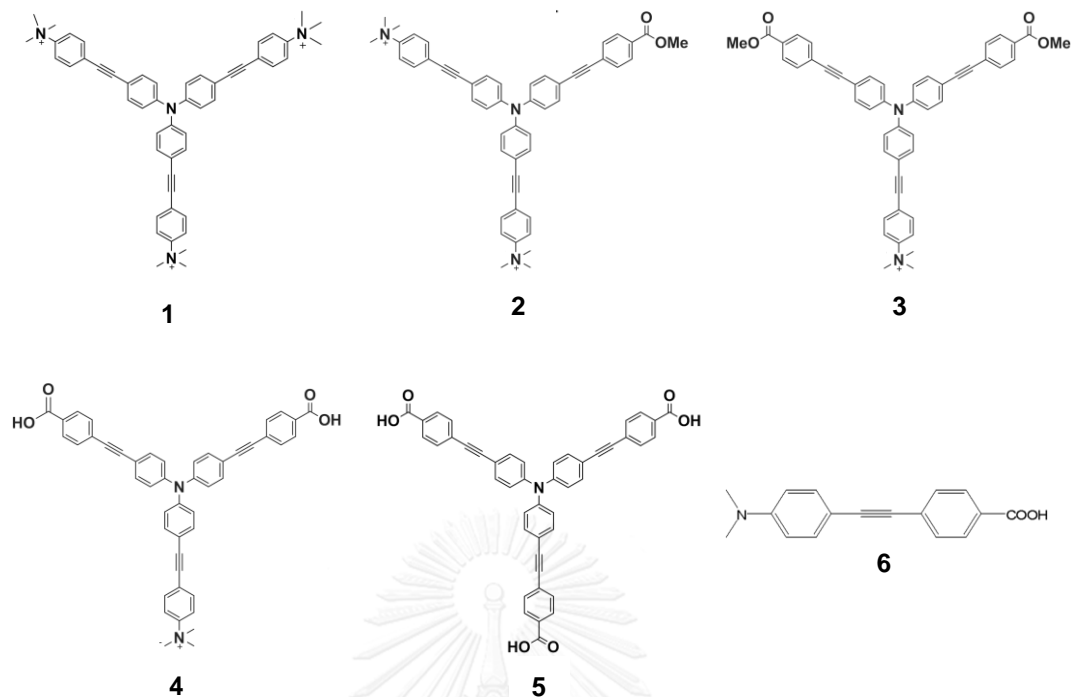


Figure 2.1. Structures of dendritic polyelectrolyte fluorescence compounds **1-6**

2.1.2 Fluorescence sensor study

2.1.2.1 Fluorescence measurement of proteins

The stock solutions of all fluorophores (5 μM) were prepared in 10 mM sodium phosphate buffer saline (PBS) pH 7.4. The stock solutions of BSA, α -CN, β -CN, α -LA, and β -LG were prepared in PBS to have the absorbance at 280nm (A_{280}) equal to 0.4. For the measurement of protein in a cuvette, the protein and fluorophore mixture was prepared by mixing the corresponding stock solutions and diluting with PBS to afford the final concentration of the fluorophore of 0.2 μM and the protein concentration with calculated A_{280} of 0.01. Fluorescence spectra were recorded at ambient temperature from 400 to 600 nm at the excitation wavelength of 375 nm.

2.1.2.2 Fluorescence measurement of commercial milk samples via cuvette

The commercial milk sample (2 μL) was diluted 10,000 times with 10 mM PBS prior to the addition of the fluorophore stock solution. The final concentration of the fluorophore in the milk samples was 0.01 μM . Fluorescence spectra were recorded at ambient temperature from 400 to 600 nm at the excitation wavelength of 375 nm.

2.1.2.3 Fluorescence measurement of commercial milk samples via microplate

The fluorophore stock solution was added directly to the commercial milk samples (70 μL) in the microplate wells to furnish its final concentration of 1 μM . The mixtures were shaken at ambient temperature for 15 min before acquiring the spectra. Fluorescence spectra were recorded at ambient temperature from 400 to 600 nm at the excitation wavelength of 375 nm.

2.1.2.4 Fluorescence measurement of foodborne pathogens via microplate

The number of bacterial counts (CFU/mL) was estimated from the optical density (OD) at 600nm nm, using a calibration line related to the plate counts. For the fluorescence measurement, the number of bacterial counts was controlled at 10^8 CFU/mL. Fluorescence spectra were recorded at ambient temperature from 400 to 700 nm at the excitation wavelength of 375 nm.

2.1.3 Data preparation

ΔI was calculated from the difference of the fluorescence intensity, $I_E - I_0$, where I_E and I_0 were the fluorescence intensity in the presence and absence of the samples, respectively.

2.2 Colorimetric sensor array

2.2.1 Materials

Eight diacetylene monomers (DA), i.e. amphiphilic (**1–6**) and bolaamphiphilic (**7–8**) compounds were utilized as precursors for preparation of PDA sensing elements in the colorimetric sensor array (**Figure 2.2**). 10,12-pentacosanoic acid (PCDA), 10,12-tricosadiynoic acid (TCDA) and 10-undecynoic acid were purchased from GFS Chemicals (USA). 6,8-nonadecadiynoic acid (6,8-19DA, **3**) was prepared according to previous report [42].

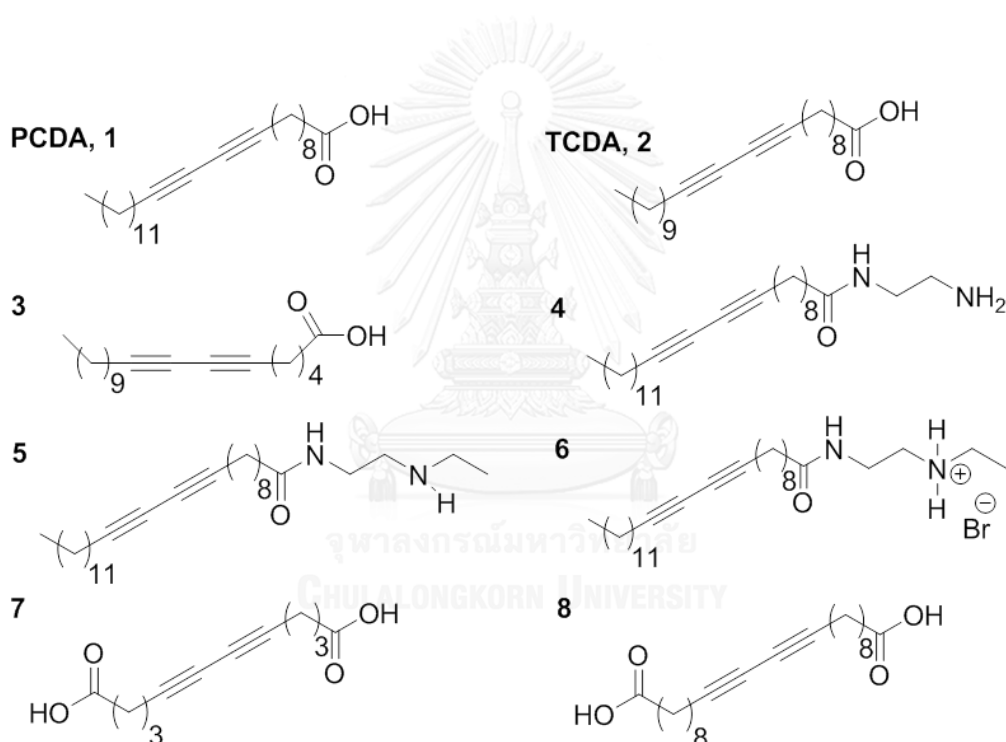


Figure 2.2. Structures of diacetylene monomers 1-8

2.2.1 Colorimetric vesicle-based sensor array

2.2.1.1 Colorimetric detection of surfactants

Diacetylene monomer (DA) **1**, **4**, **5** and **6** were transformed into aqueous sols by sonication then followed by UV-irradiation to afford Polydiacetylene (PDA) vesicle sols. Three anionic surfactants (SDC, SDS, SDBS), three cationic surfactants (TTAB,

DTAB, HTAB), and three nonionic surfactants (TWEEN-20[®], Brij[®] 58, TRITON[®] X-100) were used as target analytes. Solutions of surfactant (0, 10, 20, 30, 40 and 50 μM) were added into each PDA solution at final concentration of 0.1 mM. The resulting solutions were stirred for 5 minutes before monitoring by UV-Vis spectrometry. The spectra were recorded from 800 to 400 nm with the zero absorbance set at 800 nm. The λ_{max} of the blue and the red phase of each sample were determined at blank (0 μM) and 50 μM of surfactant concentration.

2.2.1.2 Data preparation for vesicle-based colorimetric response

After collections of the electronic absorption spectra by the UV-visible spectrophotometer, the percentage of colorimetric response (%CR) was determined based on the difference of blue color fraction according to the following equation:-

$$\%CR = 100 \times (FB_0 - FB)/FB_0$$

where FB_0 is the blue color fraction before UV irradiation and FB is the blue color fraction after UV irradiation.

FB is the blue fraction calculated from $A_{\text{blue}} / (A_{\text{blue}} + A_{\text{red}})$ where A_{blue} and A_{red} are the absorbance detected at the λ_{max} of the blue and the red forms of polydiacetylenes respectively.

2.2.2 Colorimetric paper-based sensor array

2.2.2.1 Fabrication of PDAs coated paper

DAs **1-8** were dissolved in THF or chloroform (2% w/v) and dropped on a filter paper (Whatman No.1 chromatography paper). In order to create the equal size of DA dots, 2 μL of the DA solutions were controllably dropped by auto pipette. The desired number of repetitions was pursued according to numbers of the DA dots casted at different locations on the filter paper. The filter paper was allowed to dry at an ambient temperature in the dark for 60 min. Polymerization was carried out under a hand-held UV lamp ($500 \mu\text{W}/\text{cm}^2$) at a wavelength of 254 nm for 1 min. The

UV lamp was hung at a height of 10 cm above the filter paper. The DA dots on the filter paper turned into blue PDA dots as a result of the UV irradiation.

2.2.2.2 Preparation of saturated analytes vapor

Eighteen organic solvents, i.e. pentane, hexane, cyclohexane, toluene, o-xylene, benzene, diethyl ether, dichloromethane (DCM), 2-propanol, tetrahydrofuran (THF), chloroform (CHCl_3), ethanol (EtOH), ethyl acetate (EtOAc), acetone, methanol (MeOH), acetonitrile (MeCN), dimethylformamide (DMF) and dimethylsulfoxide (DMSO), were selected as target analytes in this work. 3 mL of each solvent was poured into a chamber and held for 60 minutes in close system to achieve saturated the analyte vapor. The filter paper containing blue PDA dots was attached on the inner surface of the chamber lid. The lid was then used to close the chamber tightly at 30 °C for 60 min.

2.2.2.3 Image processing methods for screening

The photographic images of the PDAs dots were recorded to monitor the color changes via scanner and digital camera. For the digital camera, the images were captured at the distance of 30 cm. Each image of PDA dots was saved in TIF format and cropped into a size of $0.5 \times 0.5 \text{ cm}^2$.

2.2.2.4 Data preparation

Raw color images (TIF format) - the mean intensity in the red, green, blue channel was recorded and converted into RGB numerical values with three numeric values ranging from 0 to 255. RGB values generated from the digital and scanned images were quantified using 8-bit Adobe Photoshop®. 8-bit digital image encodes pixel color values by devoting 8 bits to each of the red (255, 0, 0), green (0, 255, 0) and blue (0, 0, 255) components to create the maximum number of 256 colors (8 bit).

A three-dimensional vector (ΔR , ΔG , ΔB) of each PDA sensor was determined from the RGB values before and after the exposure of VOCs to vapor.

$$\Delta \text{ RGB value} = (\text{RGB value of VOCs}) - (\text{RGB value of blank})$$

whereas; R = red, G = green and B = blue values.

The Δ RGB of 12 replicated measurements were averaged and standard errors of each data set were determined.

2.3 Data analysis using multivariate statistical analysis

The data obtained from the sensor measurement results were arranged as $p \times n$ data matrices; where p corresponds to the repetitions treated as objects (row) and n corresponds to the measurement results (e.g., ΔI , %CR or Δ RGB) treated as variables (column). Multivariate statistical analyses were performed on the $p \times n$ data matrices using Unscrambler® 9.7 or XLStat® 2010.

2.3.1 Descriptive data analysis

Principal component analysis (PCA) was performed on all of the data matrices as the descriptive data analysis to achieve an overview of all the information in the dataset.

2.3.2 Predictive data analysis

Linear discriminant analysis (LDA) was performed on all of the data matrices to classify samples to different classes and to test the discriminating ability of sensor array. Factorial discriminant analysis (FDA) was performed on the selected PCs obtained from the PCA technique to quantify the classification accuracy.

Full cross-validation with a leave-one-out technique was applied to PCA, FDA and LDA models to assess the performance of each model based on the classification validity and accuracy of the samples in the validation set. In the cross-validation, a sample in p was randomly removed from the data set and LDA was used to determine the centroid coordinate for each known class of the rest of the samples ($p-1$). The removed sample was then classified to the group of which

centroid closest to the sample score coordinate. The procedure was repeated until all samples were classified. The ratio of the numbers of the samples correctly predicted by LDA to the total numbers of the samples defined the accuracy percentage.

2.3.3 Feature Selection

Feature selection was performed based on full cross-validation with a leave-one-out technique applied to PCA and LDA models for dimensionality reduction by removing sensing elements with low contribution from the array.



CHAPTER III

RESULTS AND DISCUSSION

3.1 Fluorescence Sensor Array

Fluorescence technique has been a detection method of choice due to its high sensitivity and vast potential for selectivity enhancement. Many biological molecules are able to luminesce or fluoresce from natural intrinsic fluorescent probes, e.g. tryptophan or natural fluorescent proteins, involving delocalized electrons in aromatic molecules. Conjugation of extrinsic fluorophore with the target biomolecules is an alternative approach, the quality of information provided by fluorescence measurements can be clearly improved. [43]. Phenyleneethynylene is an important class of π -conjugated molecules currently applied as fluorescent transducers in various optical sensing systems [44, 45]. Recently, dendritic fluorophores composed of phenyleneethynylene repeating units with various charges are developed as a biosensor array.

3.1.1 Fluorescence sensor array for protein discrimination

A sensor array composed of fluorescent compounds **1-5** having various interaction sites, created by different combinations of cationic trimethylammonium, anionic carboxylate and non-ionic methyl ester on the peripheries, was initially developed to investigate the fluorescence responses toward a set of eight commercially available proteins Bovine Serum Albumin (BSA, pI = 4.8, 66.3 kDa), tetrameric Concanavalin A (ConA, pI = 5.5, 106 kDa), Cytochrome C (CytC, pI = 10.7, 12.3 kDa), Histone (His, pI = 10.8, 21.5 kDa), Human Serum Albumin (HSA, pI = 5.2, 69.4 kDa), Lysozyme (Lys, pI = 11.0, 14.4 kDa), Myoglobin (Myo, pI = 7.2, 17.0 kDa), and Papain (Pap, pI = 9.6, 23.0 kDa) [34]. These proteins were selected based on the variation of isoelectric points (pI) and molecular weights.

The addition of these proteins into solutions of the fluorophores was conducted to observe the fluorogenic response changes. As shown in **Figure 3.1**, different fluorescence patterns were visualized.

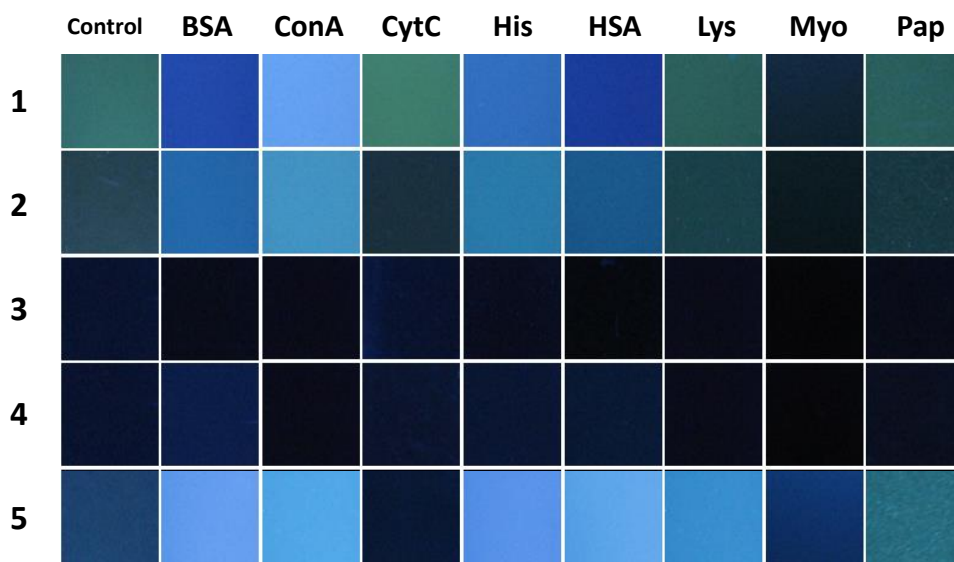


Figure 3.1. Cropped photographic images of the fluorophore solutions ($2.0 \mu\text{M}$) in phosphate buffer saline (10 mM, pH 7.4) mixed with each of protein solution ($A_{280} = 0.1$) under black light.

Fluorescence intensities obtained from nine replicated measurements for each pair of protein and fluorophore were used to construct a histogram plot as shown in **Figure 3.2**.

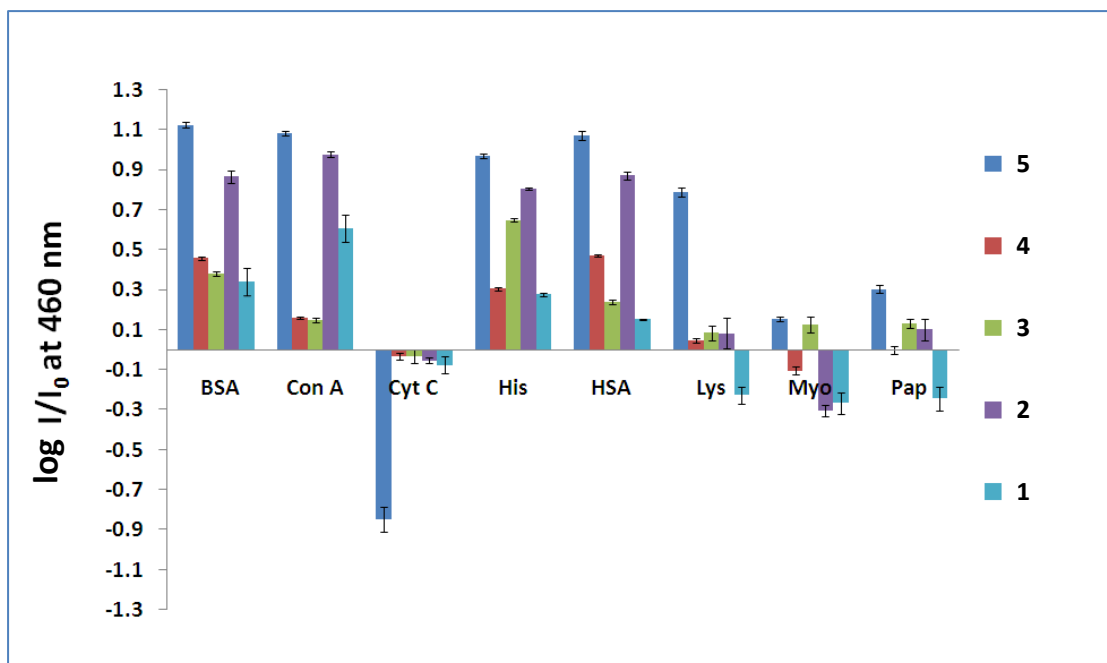


Figure 3.2. Histogram of logarithmic values of relative intensity ($\log I/I_0$) at 460 nm ($\lambda_{\text{ex}} = 375$ nm) of fluorophore solutions ($0.20 \mu\text{M}$) in phosphate buffer saline (10 mM, pH 7.4) mixed with each of protein solution ($A_{280} = 0.01$).

3.1.1.1 Multivariate statistical analysis

Although the histogram in **Figure 3.2** already showed differentiable patterns of the fluorogenic responses toward the eight protein analytes, computational pattern recognition of these multidimensional dataset (5 fluorophores \times 8 proteins \times 9 replicates) can be further realized using multivariate statistical analyses. In this work, an unsupervised PCA method was first applied to the dataset of fluorescent intensity differences (ΔI) at 460 nm without defining classes of protein samples. The PCA transformed the data set into principle component (PC) scores based on their pattern similarity [13, 46]. The same methodology was applied to data matrices of ΔI at various wavelengths ranging from 430 to 520 nm to generate PCA score plots as shown in **Figure 3.3**.

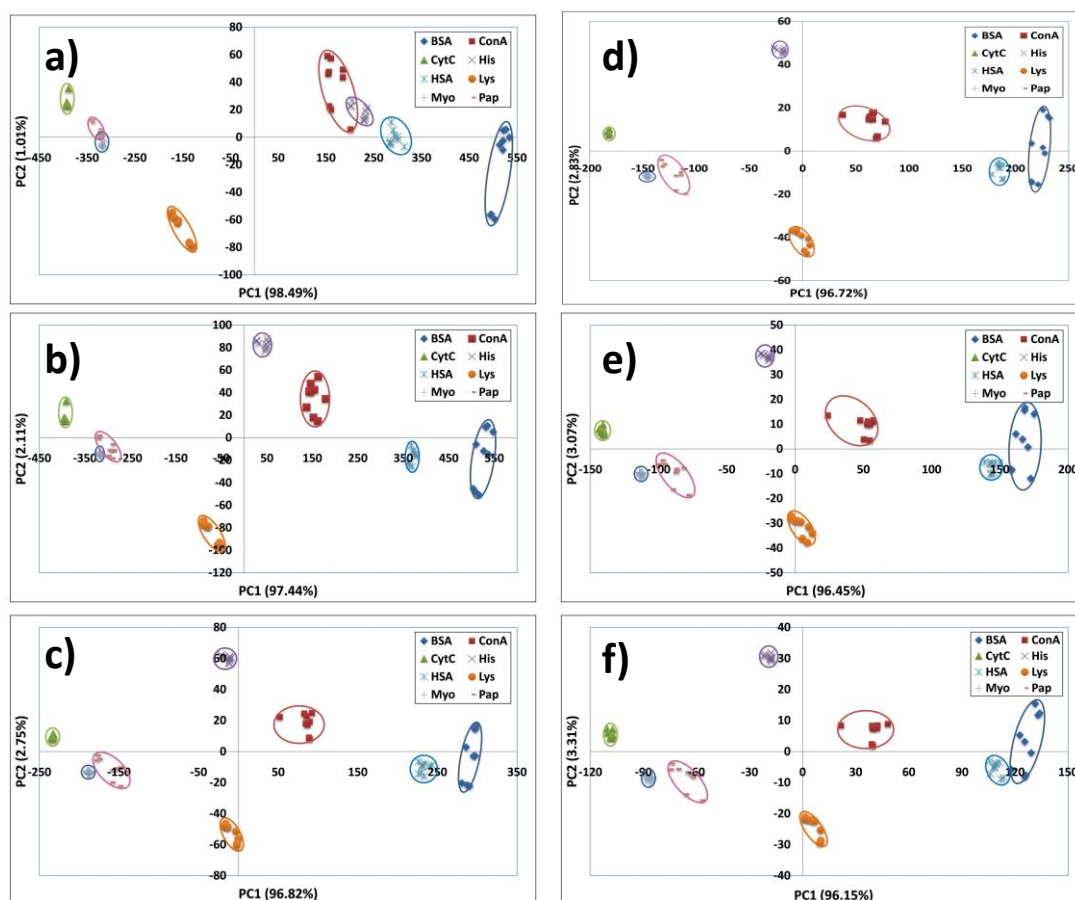


Figure 3.3. PCA score plot of ΔI obtained from the data set of 5 fluorophores \times 8 protein samples \times 9 replicates measured at 430 nm (a), 460 nm (b), 490 nm (c), 500 nm (d), 510 nm (e) and 520 nm (f).

The next step is to test the discriminating ability of the array by factorial discriminant analysis (FDA). FDA is a supervised method [25, 47, 48] which was applied on the PC scores of the selected PCs using a leave-one-out cross-validation to check whether the classification accuracy is acceptable. The cross-validation is a technique applied to the samples whose category is known and are divided into two sets: a training set and a prediction (or evaluation) set. The samples in the training set are used to develop classification/prediction rule; the category of the samples in the prediction set is evaluated by the rules thus obtained and the percentage of the correct classification achieved gives the predictive ability. The subdivision between training and evaluation set will be made by using a leave-one-out routine, which one

observation will be left out of the set at the time and the rest of the data will be used as a training set.

The leave-one-out cross validation cannot be applied directly to PCA due to the fact that PCA itself is not a supervised method. FDA with cross-validation will be applied to test the discriminating ability of the sensor array. Initially, a step-wise discriminant analysis will be performed to select the PCs that are most relevant for discrimination of variables corresponding to the sample groups which are initially defined. FDA will be applied on the selected PC scores to test the discriminating ability by the cross validation technique, i.e. each data point will be reallocated within the various sample groups, distances between the data point and the various centers of gravity of each group is calculated, and the data point will be assigned to the one which the distance is shortest. The comparison of the assignment group with the real group is an indicator of the quality of the discrimination.

Based on PCA scores of the first two PCs at various wavelengths ranging from 430 to 520 nm, the percentage of classification accuracy obtained from FDA revealed that the data at 500 nm provided the highest accuracy at 98.61% (**Table 3.1**). **Figure 3.3d** showed the PC score plot of ΔI measured at 500 nm having a total variance of 99.55% in which PC1 and PC2 contributed 96.72% and 2.83%, respectively.

Table 3.1. Variance contribution of the first two PCs and classification accuracy obtained from PCA and FDA on the fluorescent intensities (ΔI) measured at various wavelengths.

Wavelength (nm)	PCA			FDA
	PC1 (%)	PC2 (%)	Sum of %PCs	% Classification accuracy
430	98.49	1.01	99.50	95.83
460	97.44	2.11	99.55	95.83
490	96.82	2.75	99.57	95.83
500	96.72	2.83	99.55	98.61
510	96.45	3.07	99.52	97.22
520	96.15	3.31	99.46	98.61

3.1.1.1.1 Feature selection

Since every variable sensing element in the array may not be significant for the discrimination of the analytes, some of the elements may generate noise or redundancy that possibly lead the adversely affect to the analysis. For practical way of sensor array, the lower number of sensing elements that impart satisfactory analysis performance is desirable. In our attempt to reduce the sensing elements, PCA loading plot was analyzed in order to identify the importance of each individual fluorophore toward each PC. The plot obtained from the measurement at 500 nm (**Figure 3.4**) showed that **5** was the main contributor to PC1 while **2** was the main contributor to PC2. Consequently, **2** and **5** as the most important sensing elements were selected for further investigation.

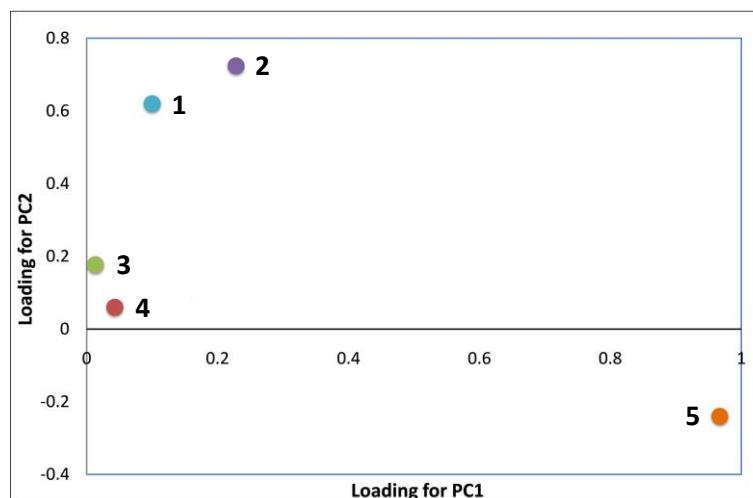


Figure 3.4. PCA loading plot of ΔI measured at 500 nm obtained from the data set of 5 fluorophores \times 8 protein samples \times 9 replicates.

The PCA score plot of the data obtained from these two selected sensing elements (2 and 5) showed in **Figure 3.5a**. The FDA cross-validation results also revealed 100% classification accuracy for these two sensing elements which is higher than the sensor array obtained from five fluorophores. The results ascertained that noise and redundant data were removed from the array. It is also importance to note that this sensor array can even discriminate between the two proteins with amino acid sequence homology such as BSA and HSA.

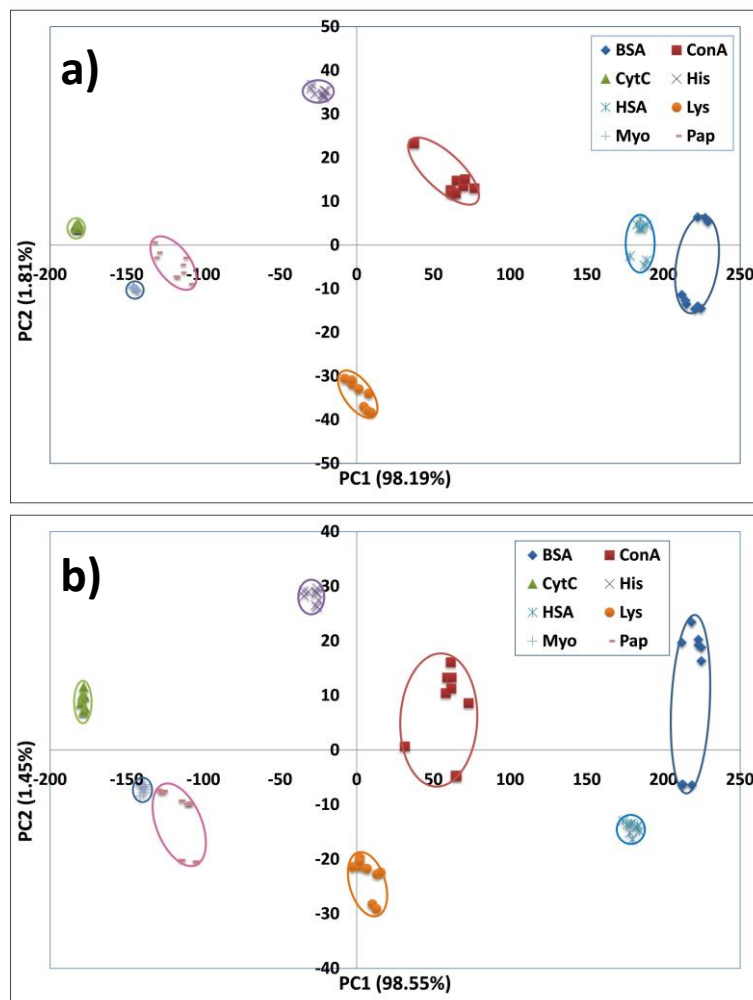


Figure 3.5. PCA score plot of ΔI measured at the excitation wavelength of 500 nm obtained from the dataset of the reduced sensing elements composing of (a) **2** and **5**, and (b) **1** and **5**.

To prove our proposition on the benefits of charge variation leading to the selection of the most suitable fluorophores **2** and **5**, the PCA was again performed on the data obtained from a reduced array comprising the fluorophores **1** and **5**, which represent usual fluorophores with all negatively ($3C^-$) or all positively ($3N^+$) charged groups. It can be seen in **Figure 3.5b** that clusters of two out of eight proteins, Myo and Pap, were located awfully close to each other. The discrimination ability of this array was confirmed by the FDA with leave-one-out cross-validation on the first two PC scores which gave only 97.22% accuracy. This is lower than 100%

classification accuracy obtained from the **2+5** array. The results highlighted the benefits of the charge variation in the development of high performance protein sensing arrays.

3.1.2 Fluorescence sensor array for milk discrimination

Further, the discrimination ability of a set of fluorophores **1-5** was systematically investigated to identify commercial milk samples according to their thermal treatments. The analytical power of the sensor array was evaluated through a sequence of three classes of samples, in the order of degree of complexity and similarity, starting from pure milk proteins, following by commercial milk obtained from different sources, and finally on commercial cow milk samples undergone different thermal treatments.

3.1.2.1 Discrimination of pure milk proteins

Five proteins i.e. α -casein (α -CN, pI = 4.1, 23-27 kDa), β -casein (β -CN, pI = 4.5, 24 kDa), α -lactalbumin (α -LA, pI = 4.2-4.5, 69 kDa), β -lactoglobulin (β -LG, pI = 5.3, 36 kDa) and bovine serum albumin (BSA, pI = 4.7, 66.3 kDa) were selected as the samples for testing the discrimination power of fluorophores **1-5** sensing array. The fluorescence responses of the fluorophores upon mixing with the proteins were measured in cuvette and were tabulated as the intensity difference (ΔI). Principal component analysis (PCA) was used to analyze the dataset of the fluorescence responses to realize any possible hidden patterns of variables in the samples. As an unsupervised pattern recognition, PCA is the dimensionality reduction method applied to unlabeled data matrix while retaining important information and intrinsic variance presenting in the original data as much as possible [13, 14, 17, 49]. In this case, PCA transformed the dataset of 90,225 ΔI values (5 proteins \times 3 samples \times 3 repetitions \times 5 fluorophores \times 401 wavelengths) into scores of uncorrelated variables called principal components (PC) for all 45 observations (5 proteins \times 3 samples \times 3 repetitions). The first two components, PC1 and PC2, accounted for 99.22% of the total variance where the first PC contains the highest degree of data variance and

other PCs follow in the order of decreasing variance. The PC score plot (**Figure 3.6a**) of distinctly showed five clusters of the data corresponding to 5 types of the milk proteins tested. As an unsupervised method, the clusters of data generated by PCA are based only on their similarities and differences perceived by the sensing elements. Therefore, the correct projection of data clusters by PCA indicates a good level of discriminating ability of the fluorophores on the milk proteins. The PCA loading plot (**Figure 3.6b**) was then used to determine the contribution of the original variables (fluorophores and wavelengths) to PC1 and PC2. The loading plot showed that **1** and **5** contributed the most to PC1 and PC2 respectively. The fluorophore **2** gave only moderate contributions to both PCs while **3** and **4** gave insignificant contributions to the PCs. Thus, two out of five fluorophores, **1** and **5**, were selected for further investigation to identify types of commercial milk samples.

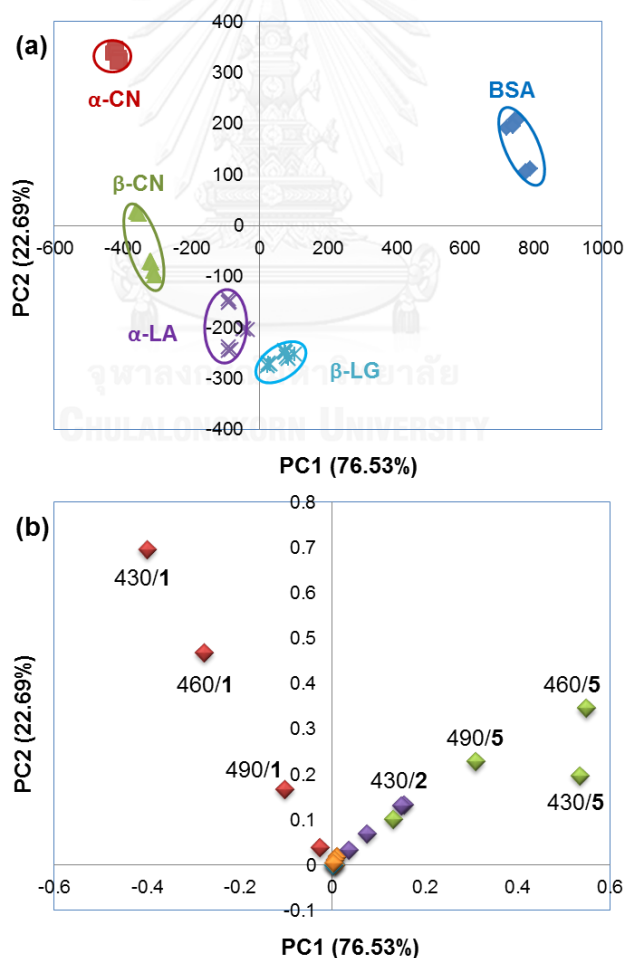


Figure 3.6. (a) PCA score plot and (b) PCA loading plot of ΔI values of fluorophores 1-5 upon mixing with each pure milk protein.

3.1.2.2 Discrimination of commercial milk from different sources

Six types of commercial milk i.e. pasteurized cow milk, UHT cow milk, sterilized cow milk, corn milk, soy milk and fermented cow milk were selected as samples for testing of discriminating ability of our sensor array. It has been reported that each type of milk contains different major proteins i.e. casein in cow milk [50], β -conglycinin and glycinin in soy milk [51], zien in corn milk [52] and casein probiotic bacteria in fermented cow milk [53]. To reduce the effects from turbidity, all samples were diluted 10,000 times before mixing with two fluorophores **1** and **5**. The fluorescence responses were measured in a conventional cuvette. The data matrix containing 14,436 ΔI values (6 milk samples \times 3 repetitions \times 2 fluorophores \times 401 wavelengths) was analyzed by PCA. The two-dimension PC score plot exhibited a cluster map of the milk samples in which six different types of milk can be correctly classified (**Figure 3.7**). The soy milk and corn milk had negative PC1 and PC2 coordinates in the left lower quadrant, the fermented milk had positive PC1 but negative PC2 coordinates in the left upper quadrant whereas all cow milk samples had positive PC1 coordinates, vertically distributed in the upper and lower right quadrant. The results demonstrated that the fluorescence spectra reflected the physico-chemical characteristics of the milk in relation to their sources i.e. plants (soy and corn milks), animals (cow milks) and fermentation (fermented milk). The variation in fluorogenic responses could be attributed to the difference in protein constituents in different milk types that interact with the fluorophores.

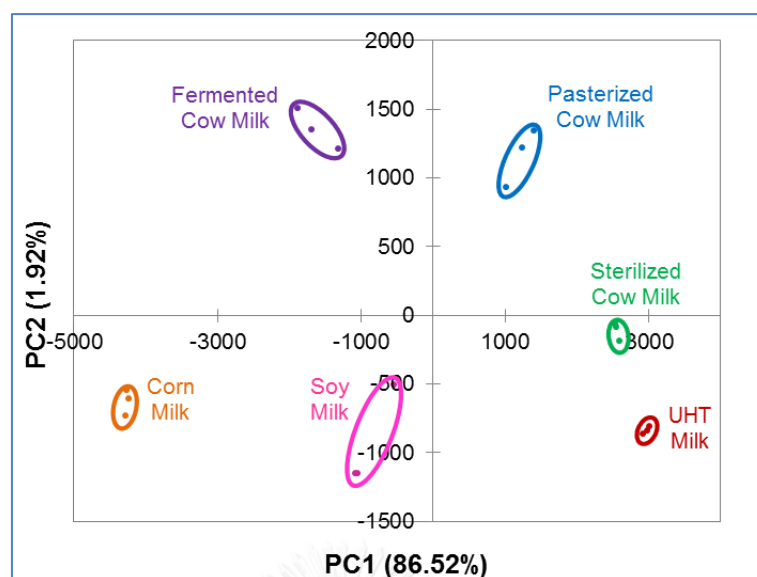


Figure 3.7. PCA score plot of ΔI values in the range of 400-600 nm for 6 types of commercial milk.

3.1.2.3 Discrimination of cow milk according to the thermal processes

Since milk in some types can often be easily distinguished by human sensory including texture, odor or taste. For example, unique beany flavor of soy milk makes its odor or taste distinctive from cow milk, while fermentation causes acidic taste. A more challenge is taken into the identification of different types of cow milk according to their thermal treatments. Thermal treatments induce significant changes in milk constituents especially in the protein structures [54-57] which are hardly perceived by human sensory attributes. Degree of the changes depends on both temperature and time of the thermal treatments which inevitably affect nutritional quality including immunological sensitizing capacity of milk [58, 59],.

One of limitation in conventional right-angle fluorescence spectroscopy is the use of low optical densities of the sample with the absorbance not exceeding 0.1. The dilution of turbid samples is not only tedious for a large number of samples but the results obtained on diluted milk may not represent all the information present in the original milk samples. Many studies overcame these drawbacks by using the

front-face geometric arrangements for fluorescence measurement of which the excitation radiation can be set at various angles, depending on surface of the samples, to avoid or minimize reflected light, scattered radiation and depolarization phenomena [3, 25]. In this study, 96-well plate was used since the light passes through near the vertical axis of the highly absorbent sample solutions and is measured at an angle of less than 90 degrees. The use of microplate fluorescence spectrometry is also more conveniently and effectively applicable for a large numbers of samples.

Four types of thermally treated cow milk i.e. pasteurized, sterilized, UHT and recombined milk were selected for this investigation. The recombined milk used in this study was the UHT processed milk obtained from the combination of fresh milk and dried milk powder. A rather harsh thermal process such as spray- or drum-drying is usually applied in commercial milk powder production [60, 61]. The fluorescence response (ΔI) spectra of **1** and **5** in the presence of milk samples are shown in **Figure 3.8**. Strong fluorescence enhancement of **1** was apparent for pasteurized and UHT milk (denoted as **P** and **U** in **Figure 3.8a**) and moderate enhancement was observed for sterilized milk and recombined milk (denoted as **S** and **R**). **Figure 3.8b** showed strongest fluorescence enhancement of **5** with pasteurized milk and a little lower enhancement with UHT milk. In contrast, sterilized milk and recombined milk induced slight fluorescence quenching (negative values). For statistical analysis, fluorescence spectra were acquired from 9 observations (3 packages of each type of milk \times triplicates for the sample of each package). The PCA applied to the fluorescent intensity differences dataset gave a two-dimensional score plot of the first two PCs as shown in **Figure 3.8c**. As anticipated in accordance to the fluorescence responses described above, only 2 groups of samples was classified by the PCA; one on the left was sterilized and recombined milk samples, and the other on the right was pasteurized and UHT milk samples.

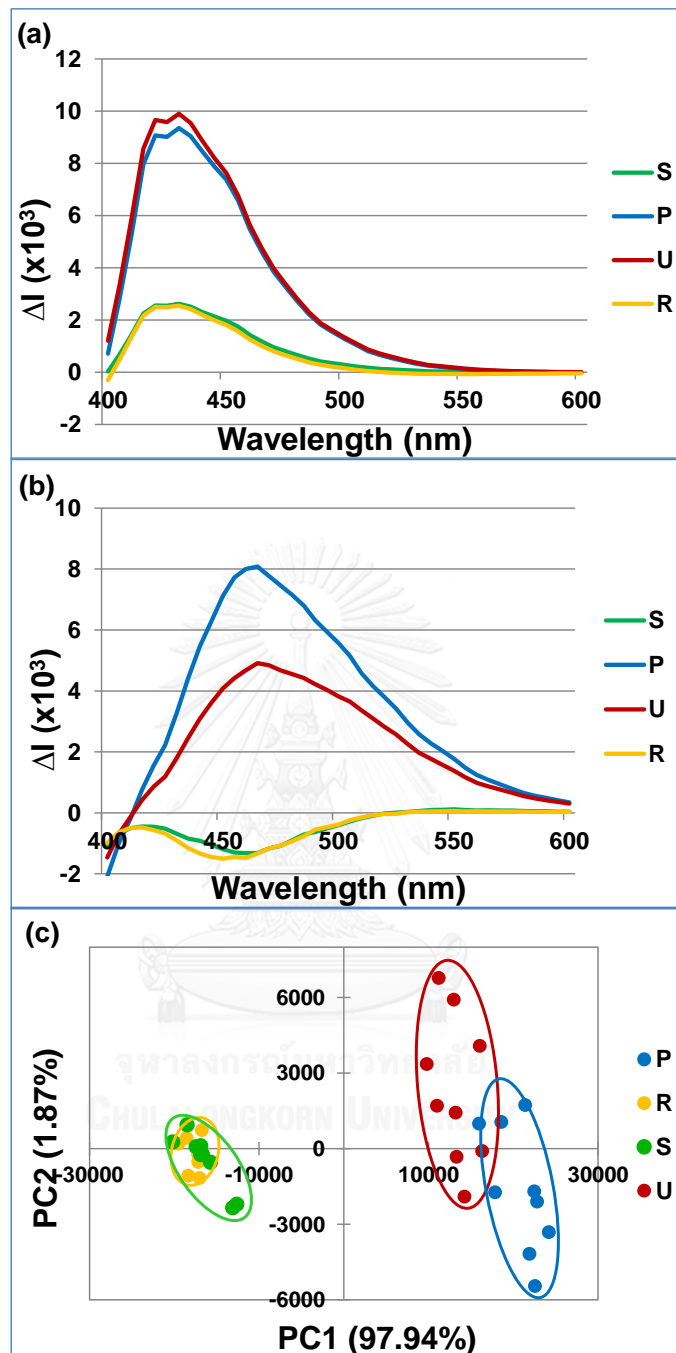


Figure 3.8. Fluorescence responses (ΔI) of fluorophore (a) 1 and (b) 5 upon addition of pasteurized (P), sterilized (S), UHT (U) and recombined (R) milk; (c) PCA score plot of the fluorescence responses of 9 measurements (3 packages \times 3 replicates) of each type of milk. Oval outlines indicate groups of milk samples of the same thermal treatment.

Although the structural changes of milk proteins under thermal treatment has not been completely realized at a molecular level, studies on heat stability of milk proteins shows that a prolong exposure to the temperature over 100 °C can cause coagulation of casein [68]. Therefore, casein, as a major protein in cow milk, is better preserved in its natural form under pasteurization and UHT processes [62-64] because pasteurization (63 °C for 30 min or 72 °C for 15 seconds) and UHT (133 °C for 1 second) applied relatively milder conditions comparing with the sterilization (120 °C for 30 minutes) and powdering (harsher condition depending on the product specification) processes. The degree of casein denaturation is higher under sterilization and powdering processes. The difference in the fluorescence responses of **1** and **5** in this case is thus likely associated with the difference in degree of casein denaturation.

To increase the performance of milk discrimination, two more fluorescence compounds, i.e. **2** and **6** were added into the array. Compound **2** was selected as it was the third contributor to the PC scores previously observed in the protein discrimination (**Figure 3.6b**). **6** which has a smaller molecular size was also added for diversifying fluorescence response patterns. It can be clearly seen from the fluorescence responses of **2** and **6** to pasteurized milk were different from those to UHT milk (**Figure 3.9a** and **3.9b**). The sterilized milk and recombined milk, previously indistinguishable by **1** and **5**, also became differentiable by the responses of **2** and **6**. These additional spectroscopic data could help discriminating the pasteurized milk from UHT milk and the sterilized from recombined milk. The PCA of the fluorescence responses of **1**, **2**, **5** and **6** confirmed that four clusters of the PC scores corresponding to four types of the thermally processed milk were well resolved (**Figure 3.9c**).

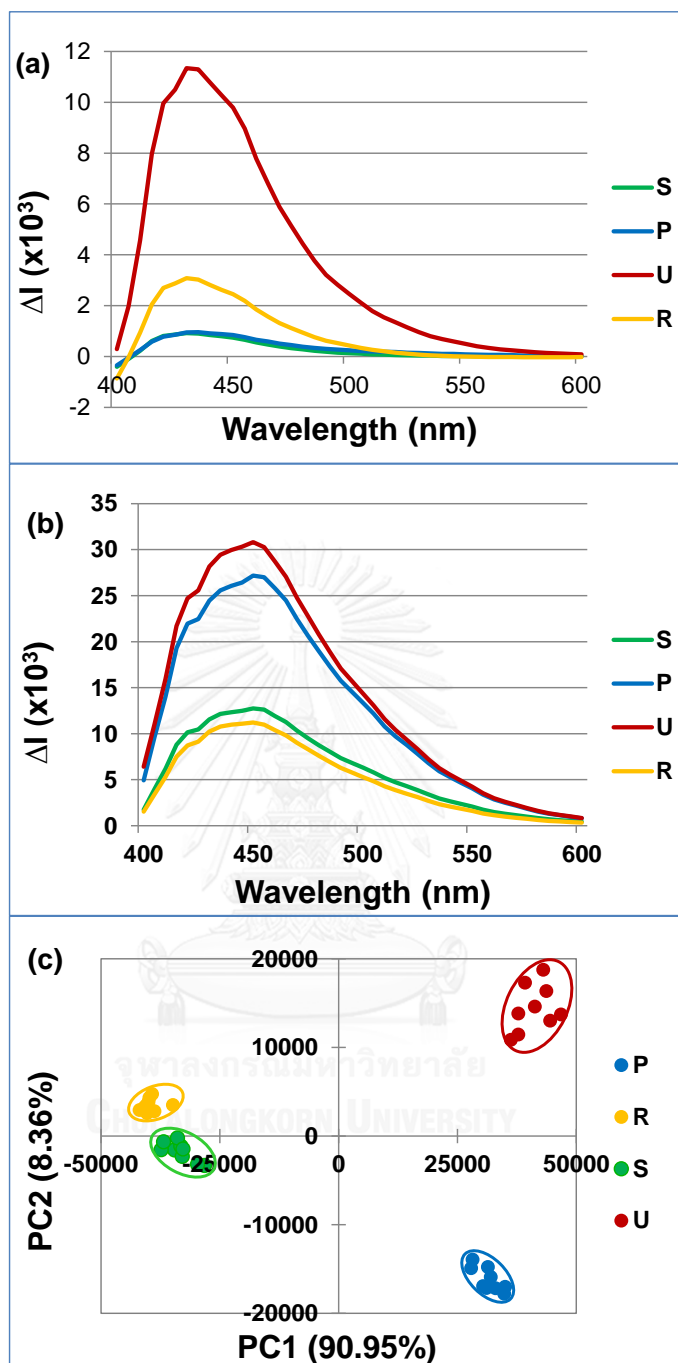


Figure 3.9. Fluorescence responses (ΔI) of fluorophore (a) 2 and (b) 6 upon addition of pasteurized (P), sterilized (S), UHT (U) and recombined (R) milk; (c) PCA score plot of the fluorescence responses of 9 observations (3 packages \times 3 replicates) of each type of milk. Oval outlines indicate groups of the milk samples of the same thermal treatment.

In order to test if the variation in dairy manufacturing process causes any effect on the discrimination ability of the sensor array, milk samples of different brands and production lots of each thermal treatment was analyzed by the fluorophore array constituted of **1**, **2**, **5** and **6**. In this study, 3 brands x 3 lots x 3 packages of each thermally processed milk were collected from local supermarkets. The total of 324 observations (108 samples × 3 replicates) of pasteurized, sterilized, UHT and recombined milks, were analyzed. Upon addition of milk sample to the solutions of **1**, **2**, **5** and **6**, the dataset of 159,408 ΔI values (108 samples x 3 repetitions x 4 fluorophores x 41 wavelengths) were analyzed by PCA to give only two clusters of samples on PC score plot on PC1 and PC2 (**Figure 3.10**).

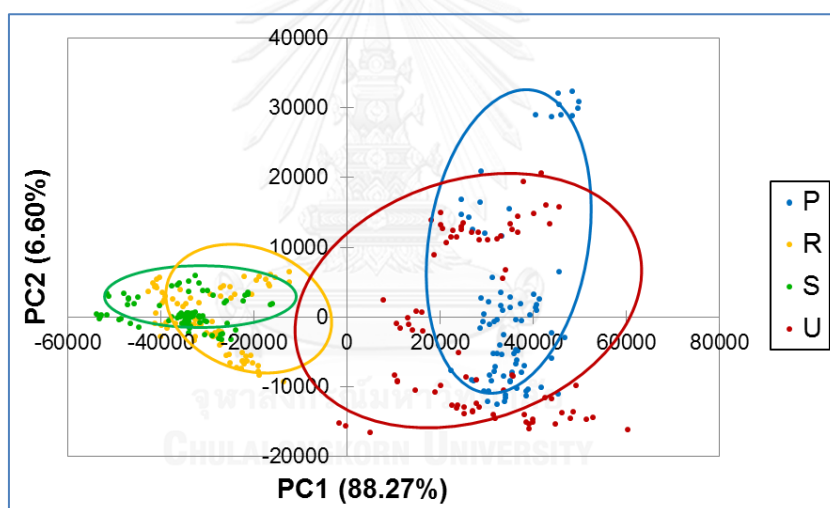


Figure 3.10. PCA score plot of the fluorescence responses of 108 milk samples from pasteurized (P), UHT (U), sterilized (S) and recombined (R) milks.

According to **Figure 3.9**, the cores of pasteurized and UHT milks scattered on the right-hand side, while those of sterilized milks and recombined milks on the left-hand side. The large scattering clusters observed in each type of thermal-processed milk indicated significant variation in manufacturing that reduce the discrimination ability of the PCA method. This is not totally surprising since the PCA is an unsupervised technique which expresses the data in such a way as to highlight similarities without real intention to discriminate the sample types. Therefore, PCA

may not necessarily show the differences between the different types of sample groups. In contrast to PCA, linear discriminant analysis (LDA) is a supervised technique which has its direct aim to classify the pre-identified sample group by using linear combination of data by maximizing the between-classes variance and minimizing the within-classes variance [17, 35, 65]. With the class information given during training, LDA thus normally give better discrimination ability comparing with PCA. LDA was therefore applied to the same spectroscopic data matrix of 108 samples. The first two discriminant variance (LD1 and LD2) accounted for 90.61% and the two-dimension score plot showed 4 well defined clusters in accordance with different thermal-processed milks (**Figure 3.11**). The leave-one-out cross-validation of the LDA scores also revealed a classification accuracy of 100%.

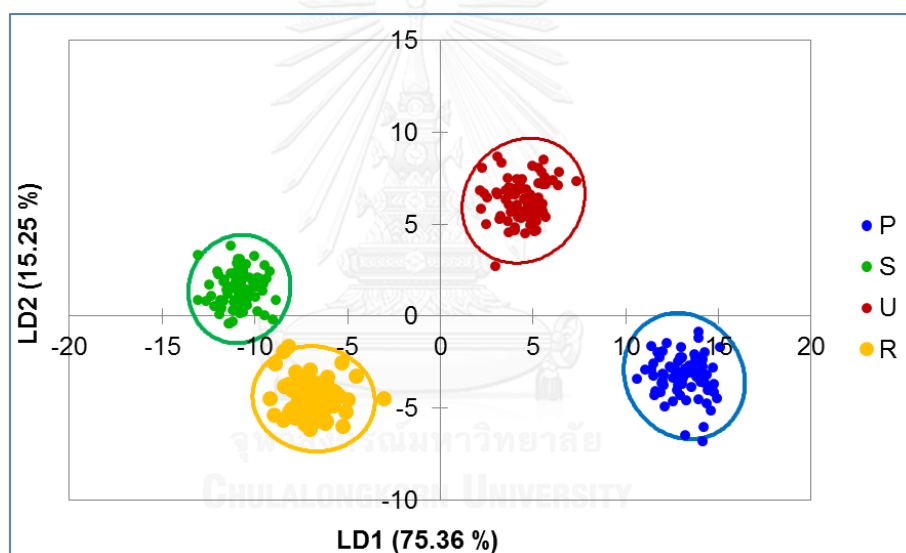


Figure 3.11. LDA score plot of first two discriminant factors (LD1 and LD2) obtained from fluorescence responses data of 108 milk samples from pasteurized (P), UHT (U), sterilized (S) and recombined (R) milk. Oval outlines indicate groups of milk samples of the same thermal treatment at 99% confidence level.

3.1.2.4 Variable selection and reduction

After the successful discrimination of the thermally processed milks via LDA technique, the next step is aimed boost the performance of the sensor array without deteriorating its performance removing the irrelevant or noise variables from the

array [66]. The LDA correlation plot against wavelength (**Figure 3.12**) was used to identify the importance of each individual fluorophore at each wavelength to each linear discriminant factor (LD). The correlation plot showed that **6** yielded the highest discriminant factor for both LD1 and LD2 while **1** and **5** yielded high discriminant factors for only LD1. Since **2** gave the lowest discriminant factor, this sensing element was again removed from the array for further investigation.

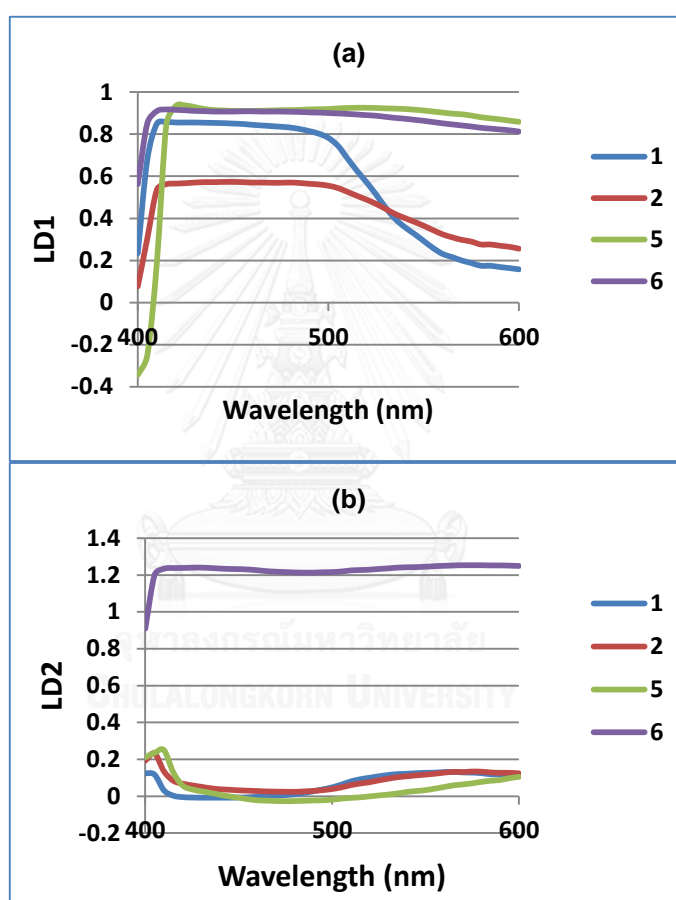


Figure 3.12. LDA correlation plot of (a) first discriminant factor (LD1) and (b) second discriminant factor (LD2) of fluorescence responses of **1**, **2**, **5** and **6** to milk samples at each emission wavelength.

The LDA with cross validation were performed for the arrays of **1+5+6** and **5+6** comparing with the original array of **1+2+5+6**. The classification accuracy of the three-fluorophore array (**1+5+6**) but not the two-fluorophore array (**5+6**) was

comparable to that of the four-fluorophore array (1+2+5+6) throughout the whole wavelength range of 400-600 nm (Figure 3.13). The results confirm that 2 can be removed from the sensor array without losing its discrimination power.

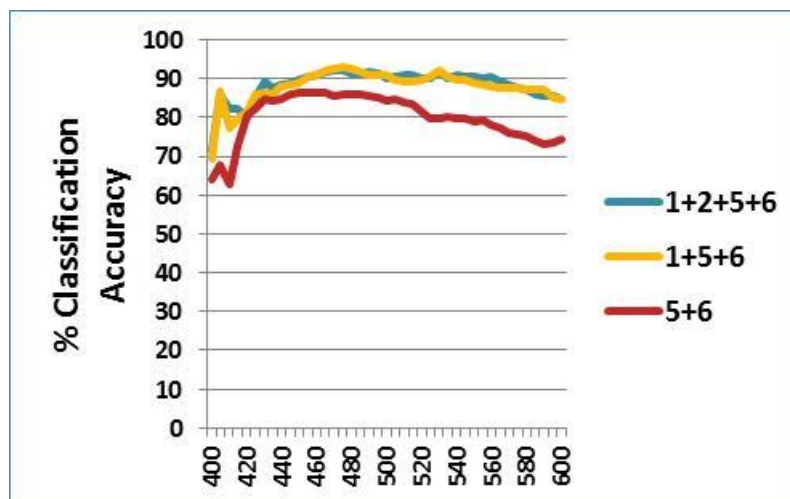


Figure 3.13. Comparison of classification accuracy obtained from leave one out cross-validation method of the LDA of fluorescence responses of the sensor arrays consisting of four (1+2+5+6), three (1+5+6) and two (5+6) fluorophores.

From the **Figure 3.13**, it can be observed that the highest classification accuracy was obtained in the wavelength range of 420-560 nm. The LDA score plot of the data obtained from 1+5+6 in the emission range of 420-560 nm showed that the first two LDs contained 91.10% of the variance and the 108 milk samples could be 100% accurately classified according to their thermal treatment (**Figure 3.14a**).

1-Anilinonaphthalene-8-sulfonic acid (1,8-ANS) is known to bind the hydrophobic pockets of proteins and frequently used in fluorescence spectroscopy. It was previously used along with intrinsic amino acid residues of milk protein such as tryptophan and retinol for classification of milk samples according to their molecular structure of proteins and fats [3, 33]. However, the drawback of this technique is that multiple excitations and emission spectra acquisitions at the corresponding wavelengths are required. This study used only a single excitation wavelength and

acquired the emission spectra in a single range by using the optimized set of fluorophores (**1+5+6**) as the sensing elements. To verify the benefits of the proposed approach in utilizing multiple fluorophores, 1,8-ANS was used as the sensing element and LDA was performed on its ΔI in its emission band from 420 to 560 nm (**Figure 3.14b**). The cross-validation yielded from the LDA gave only 97% classification accuracy which is lower when comparing with 100% obtained from our optimized sensor array. This confirmed a superior result from our sensor array.

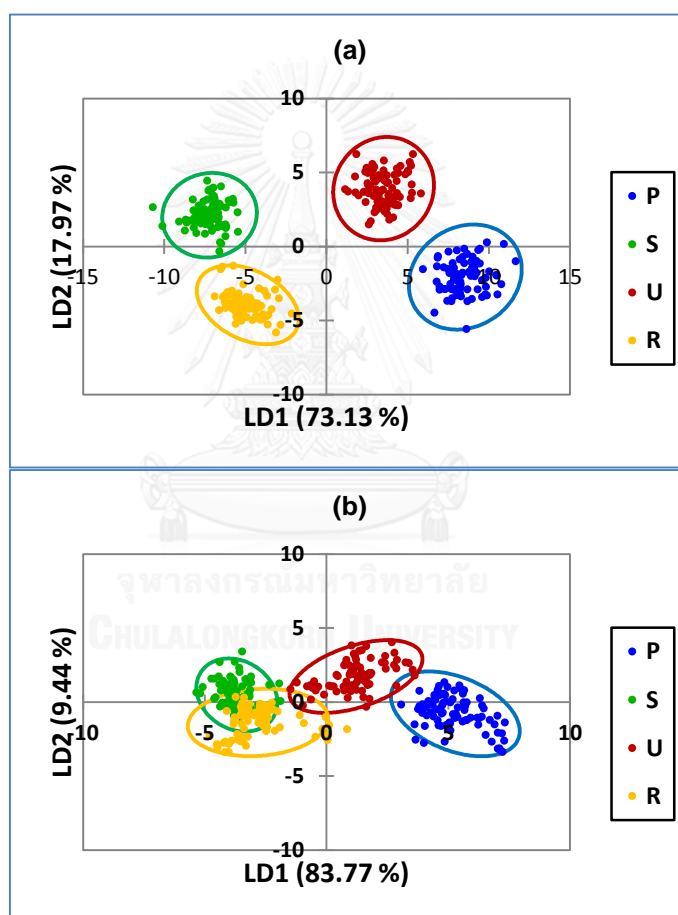


Figure 3.14. LDA score plot obtained from fluorescence responses of (a) **1+5+6** array and (b) 1,8-ANS in emission range of 420-560 nm. Oval outlines indicate groups of the milk samples of same thermal treatment at 99% confidence level.

3.1.3 Fluorescence sensor array for pathogen discrimination

Safe and sufficient food is a fundamental human need. Foodborne diseases are a global health concern encompassing millions of people. A wide spectrum of illnesses is the result of ingestion of foodstuffs majorly contaminated by microbial pathogens. In the past decade, serious outbreaks of foodborne disease have been reported on every part of the world which not only adversely affect people's morbidity and mortality, but also have negative economic consequences for society, organizations, public and private, communities and individuals. The term "foodborne disease" has been traditionally defined as illnesses caused by microorganisms, with often acute reactions, such as diarrhea. World Health Organization (WHO) estimates that worldwide foodborne and waterborne diarrheal diseases kill about 2.2 million people annually (WHO, 2013). Conventional and standard bacterial detection methods such as culture and colony counting methods may take up to a few days to yield a result, while immunology-based methods and polymerase chain reaction based methods, require extensive sample preparation [67, 68]. Eight bacteria i.e. *Vibrio cholera* (Gram-negative, curved rod shape), *Shigella flexneri* (Gram-negative, rod shape), *Bacillus cereus* (Gram-positive, rod shape), enterotoxigenic *Escherichia coli* and *Escherichia coli* (Gram-negative, rod shape), *Listeria monocytogenes* (Gram-positive, rod shape), *Salmonella Typhimurium* (Gram-negative, rod shape) and *Staphylococcus aureus* (Gram-positive, round shape) were selected to represent foodborne pathogens used for testing the discrimination power of the fluorophores **1, 2, 5** sensor array.

The fluorescence responses of the fluorophores upon mixing with the solution containing pathogens (10^8 cfu/well) were measured in the range of 400-700 nm using the excitation wavelength at 375 nm by a microplate fluorometer. Data matrix of the intensity differences (ΔI) at all wavelengths were prepared and analyzed by multivariate statistical analyses.

Table 3.2.. Pathogenic microorganisms responsible for foodborne illness [67, 69-71].

Microorganism	Gram	Shape	Food Sources	Predominant Symptoms
<i>Vibrio cholerae</i>	Negative	Curved rod	Marine foods and environmental waters	abdominal discomfort and diarrhea Vomiting also occurs
<i>Shigella flexneri</i>	Negative	Rod	Salads (potato, tuna, shrimp, macaroni, and chicken), milk and dairy products, and poultry	Abdominal pain, cramps, diarrhea, fever, vomiting, blood, pus, or mucus in stools, tenesmus (straining during bowel movements).
<i>Bacillus cereus</i>	Positive	Rod	Meats, milk, vegetables, fish, rice, pasta, and cheese, stews, gravies, vanilla sauce.	Vomiting, abdominal cramps, diarrhea, nausea.
<i>Listeria monocytogenes</i>	Positive	Rod	Soft cheese, raw milk, improperly processed ice cream, raw leafy vegetables; ready-to-eat deli meats, raw meat and poultry	Fever, muscle aches, and nausea or diarrhea. Pregnant women may have mild flu-like illness, and infection can lead to premature delivery. Elderly patients may have bacteremia or meningitis.
<i>Salmonella Typhimurium</i>	Negative	Rod	Contaminated egg; poultry; meat; dairy foods and fruit juice; raw fruits and vegetables	Fever, headache, shivering, loss of appetite, malaise, constipation, and muscular pain
<i>Staphylococcus aureus</i>	Positive	Round	Raw milk, cheese, meat and meat products, salads, cream-filled bakery products and dairy products	Nausea, vomiting, retching, diarrhea, abdominal pain, prostration
Enterotoxigenic <i>Escherichia coli</i> (ETEC)	Negative	Rod	Contaminated water and food in developing countries	Diarrhea among children and travelers in developing countries
<i>Escherichia coli</i>	Negative	Rod	Food or water contaminated with human feces	Harmless to harmful depending on strains. Some cause food poisoning or food contamination

In this study, PCA was applied to convert the fluorescence dataset with 42 original variances (3 fluorophores \times 14 wavelengths) into PC scores of PC1 and PC2 which accounted for 97.7% of the total variance. The PC score plot (**Figure 3.15**) of all 168 samples (8 bacterial samples \times 21 repetitions) on PC1 and PC2 coordinates gave four distinctive clusters of *Vibrio cholera*, *Staphylococcus aureus*, enterotoxigenic *Escherichia Coli* and *Bacillus cereus* distributing on the right half of the plot, while *Shigella flexneri*, *Escherichia coli*, *Listeria monocytogenes* and *Salmonella Typhimurium* cannot be discriminated and gathered as one cluster on the left half. PCA, as a unsupervised statistical method which condense large amounts of data into fewer latent variables, while preserving intrinsic variance of the original data as much as possible, is likely to recognize the similarity of bacteria cell envelope and shape as 3 out of 4 bacteria in the bunched cluster, *Shigella flexneri*, *Escherichia coli* and *Salmonella Typhimurium* are Gram-negative and rod shape. In Gram-negative bacteria, the surface of the outer membrane is composed predominantly of lipopolysaccharides [72], amphiphatic molecules which could be bound with a set of the fluorophores. The only Gram-positive bacterium in the bunch is *Listeria monocytogenes*. Not surprisingly, *Listeria monocytogenes* is unique among other Gram-positive bacteria due to its possession of lipopolysaccharide at its cell surface which makes it resemble structure and function of the Gram-negative bacteria [73, 74]. ETEC is not bunched with *Escherichia coli* cluster on the left side of the PC score plot since PC scores of ETEC are at the lower right quadrant. ETEC is a Gram-negative rod-shaped bacterium that produces more enterotoxin causing secretion of large amounts of fluids and electrolytes which is a cause of diarrhea [72, 75-77]. ETEC typically adhere to host cells via filamentous bacterial surface structures known as colonization factors (CFs) [78, 79] which possesses Coli Surface Antigen (CS), an antigen located in the outer surface coat. CS expressing charged residues at cell surface of ETEC makes its fluorescent pattern distinguished from *Escherichia coli*.

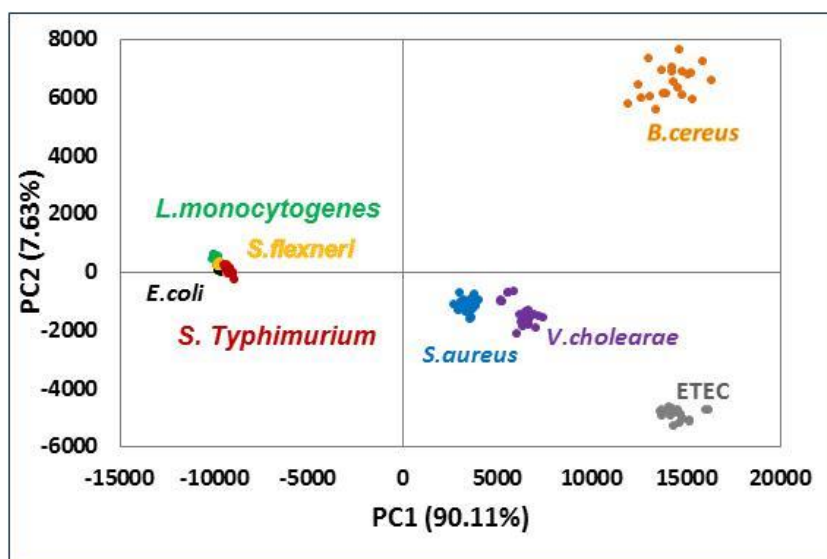


Figure 3.15. PCA score plot of ΔI values of fluorophores 1-3 upon mixing with each bacteria sample

Since PCA is an unsupervised method, the generated clusters of data are based on their similarities and differences perceived by the sensing elements without the consideration of analytes class labels [14]. The results described above showed that PCA did not have enough discrimination performance in some pathogens. The supervised LDA which normally gives better discrimination ability comparing with PCA was then performed for further investigation in this study.

In order to achieve an optimum classification capacity of our sensor array, we applied the LDA analysis onto the same spectroscopic data matrix as done with the PCA. LDA is probably the most frequently used supervised pattern recognition method for the food analysis [17]. LDA is based on the determination of linear discriminant functions, which will find the directions (axes) that maximize the linear separation among the multiple groups of analytes [35, 65].

The fluorescence data matrix of 7,056 ΔI values (3 fluorophores \times 14 wavelengths \times 8 bacterial samples \times 21 repetitions) was converted by LDA to linear discriminant scores. The three-dimension plot showed 8 well defined clusters with no overlap between the groups corresponding to the bacterial pathogens (**Figure**

3.16a). The first three discriminant factors (F1, F2 and F3) contain 58.08, 27.69, and 8.69% respectively, occupying 94.46% of total variation. The leave-one-out cross-validation of the LDA scores also revealed a classification accuracy of 100%.

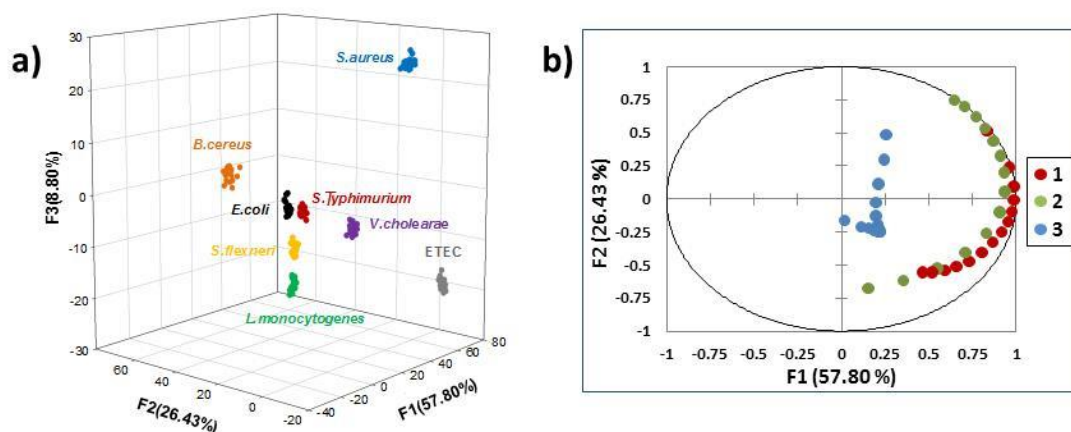


Figure 3.16. LDA score plot of first three discriminant factors (F1, F2 and F3) obtained from fluorescence responses data of bacteria samples (a); LDA loading plot of ΔI values of fluorophores 1, 2, 5 upon mixing with each bacteria sample (b).

3.1.3.1 Variable selection and reduction

Part of the motivation of this work was to apply a set of our fluorescence compounds as a rapid detection method. Reducing sensing elements, by removing either the fluorescence compounds or detection wavelengths, as minimal as possible, is one way to do so. In addition, the removal of irrelevant or noise variables should improve the performance of the sensor array because not every variable would be always significant for the discriminating ability. The LDA correlation plot (**Figure 3.16b**) was used to identify the importance of each individual fluorophore at each wavelength to each of the first two linear discriminant factors (F1 and F2). **Figure 3.16b** shows that 5 gives the lowest contribution to both F1 and F2 while 1 and 2 contribute more to both F1 and F2. Fluorophore 5 was thus removed from the array. The LDA with leave-one-out cross-validation was again applied to the dataset

without the fluorescence responses of **5** to test the discrimination performance of the array after the noise reduction.

Figure 3.17 shows the two-dimension LDA score plot for the sensor arrays of 1 and 2 with the first two discriminant factors (F1 and F2) contained about 86% of the variance. All 8 bacteria strains could be 100% accurately classified. The result confirms that the fluorescence dataset from 1 and 2 contains highly correlated variables which enable effective classification of multivariate data. Fluorophore **5** can be removed from the sensor array for the benefit of the discrimination power by reducing fluorescence measurements as well as the dimension of LDA score plot.

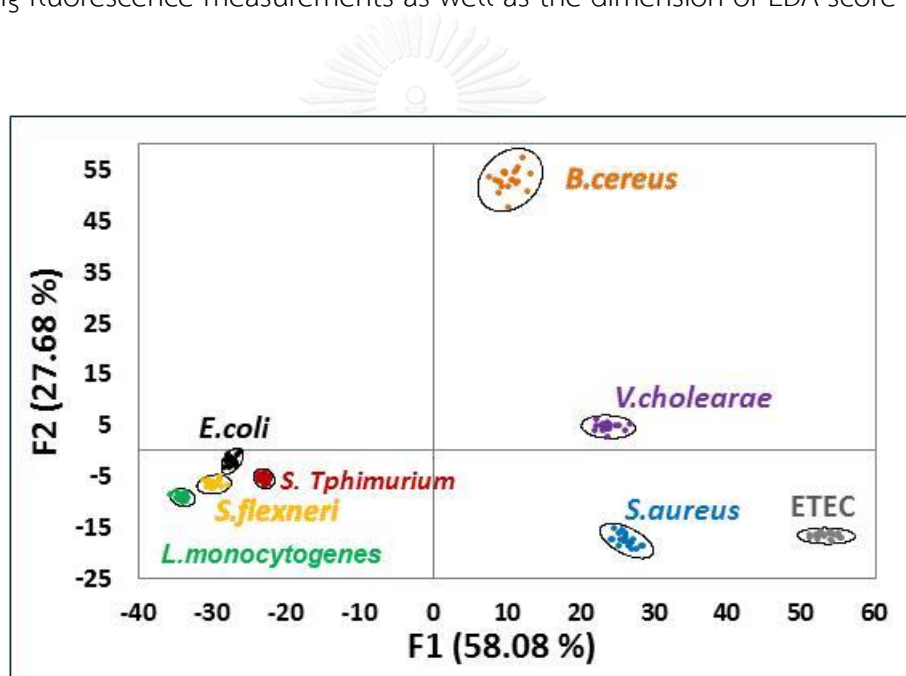


Figure 3.17. LDA score plot of first two discriminant factors (F1 and F2) obtained from ΔI values of fluorophores **2** and **5** upon mixing with each bacteria sample. Oval outlines indicate groups of bacteria sample at 95% confidence level

3.2 Colorimetric Sensor Array

Polydiacetylene (PDA) is conjugated polymer capable of responding to external stimuli by color changes. The dramatic blue to red color changes can be displayed according to its unique chromic properties induced by various external stimuli including light (photochromism) [80], heat (thermochromism) [81], mechanical stress

(mechanochromism) [82], solvents (solvatochromism) [83] and binding of specific chemical or biological agents (affinochromism /biochromism) [84]. The diacetylene lipid is one of the most useful monomers for preparation of PDA sensors due to its efficiency to form nano-vesicles homogenously dispersed in aqueous media which is suitable for biological and environmental sensing applications. Two series of diacetylene monomers, amphiphilic and bolaamphiphilic diacetylene, were selected for this colorimetric sensor study.

3.2.1 Colorimetric vesicle-based sensor array

Anionic Surfactant is a common ingredient in a variety of applications, especially in household products and industrial applications including cleaning agents and emulsifiers in cosmetics, pharmaceutical products and chemical reaction processes. Since anionic surfactants are classified as water pollutants by EU-EPA, rapid and accurate detection method of pollutant contamination in water is one of the key tasks of environmental monitoring. Colorimetric sensor array constitutes of diacetylene monomers **1**, **4**, **5** and **6**.

To test the sensing ability of these PDA with surfactants, the panels each type of surfactants (50 μM) was added to all prepared PDA sols including poly (PCDA) and the results are presented in **Figure 3.18**. The blue PDA sol prepared from PCDA did not exhibit any color change upon mixing with anionic (SDC, SDS, and SDBS) and nonionic (Tween-20®, Brij®58P, and Triton®X-100) surfactants. For the cationic surfactants, DTAB and HTAB induced the sol to form blue precipitation whereas HTAB turns blue color to unusual yellowish green without any precipitation. It is important to note that PDA sols obtained from **4**, **5** and **1+6** mixture exhibited the color transition from blue to red selectively to the anionic surfactant.

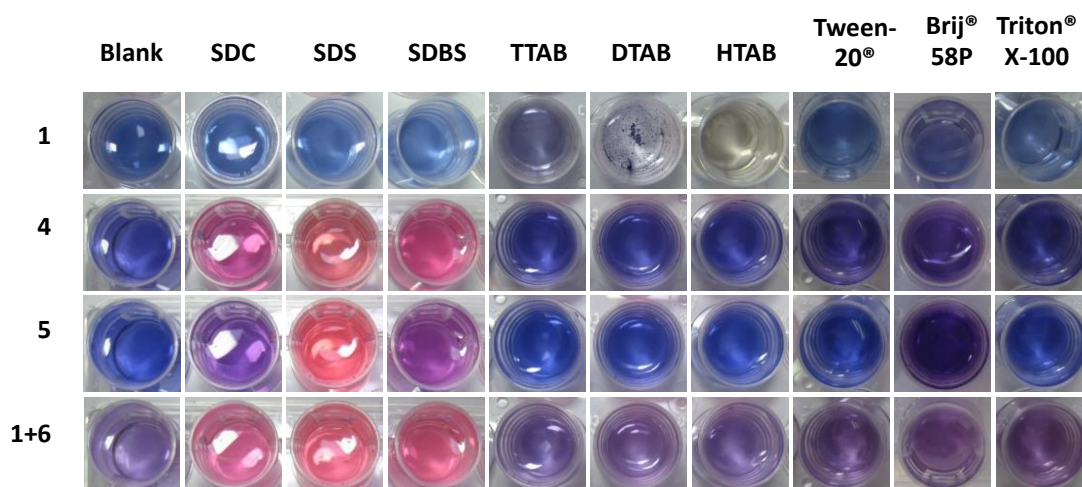


Figure 3.18. Photograph images of PDA sols (0.1 mM) in the absence (Blank) and presence of each surfactant at a concentration of 50 mM.

The colorimetric change of PDA sols in response to surfactants was also monitored by means of UV-visible signals. The addition of anionic surfactants, SDC, SDS, and SDBS to the PDA sols yielded the simultaneous effect of the decrease of the UV absorption at 640 nm and the increase in the absorption at 540 nm as shown in **Figure 3.19**. The vivid color can be observed in the mixing between SDS and PDA sols derived from **4**, **5**, and **1+6** even at the low concentration below 10 μM , while SDC and SDBS caused the obvious color change at the concentration range of 20-40 μM .

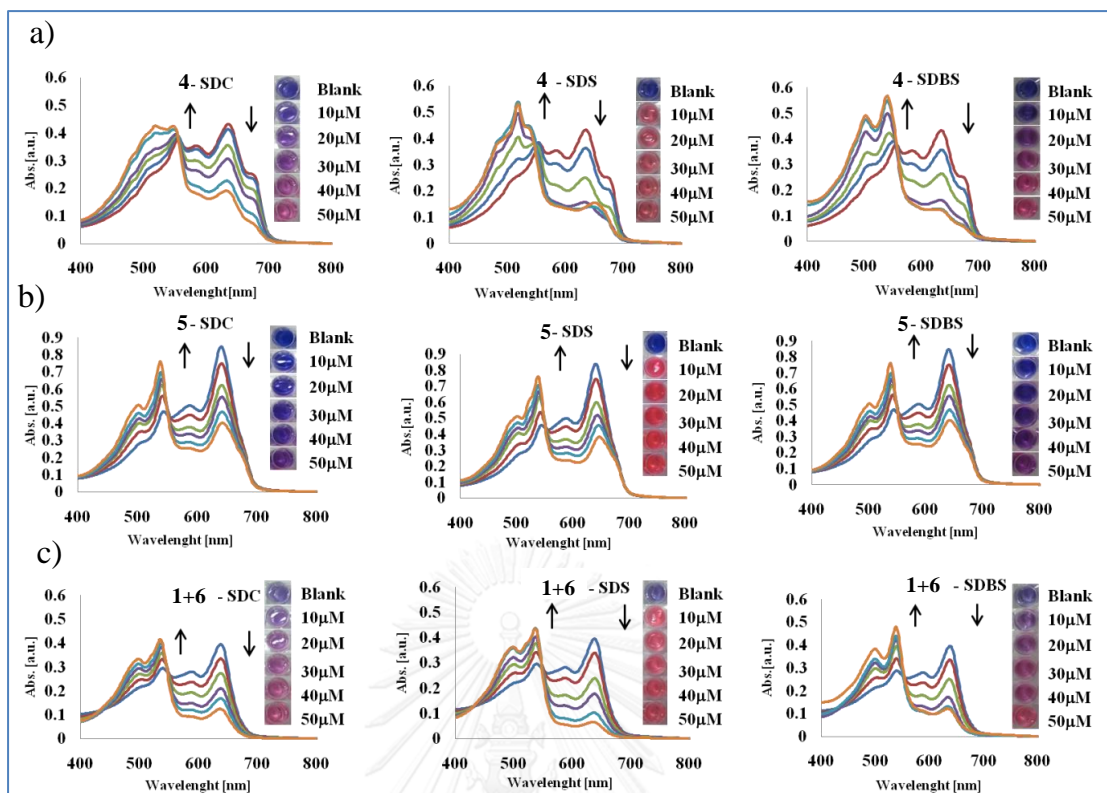


Figure 3.19. UV/Vis spectra and photographs of PDA sols (0.1 mM) derived from a) **4**, b) **5** and c) **1+6** mixture in the presence of various amounts of SDC, SDS, and SDBS at room temperature showing their diverse colorimetric transitions.

The addition of cationic surfactants, TTAB, DTAB, HTAB and Triton®X-100 at a concentration of 50 μM , to PDAs derived from **4**, **5**, and **1+6** were also performed in order to investigate the selectivity of our sensor. **Figure 3.20** showed the no significant color changes monitored by UV-visible spectrophotometer at the absorption wavelength of 640 nm. However, UV-visible spectrum of Tween-20®, and Brij®58P at a concentration of 50 μM showed minor change (**Figure 3.21**).

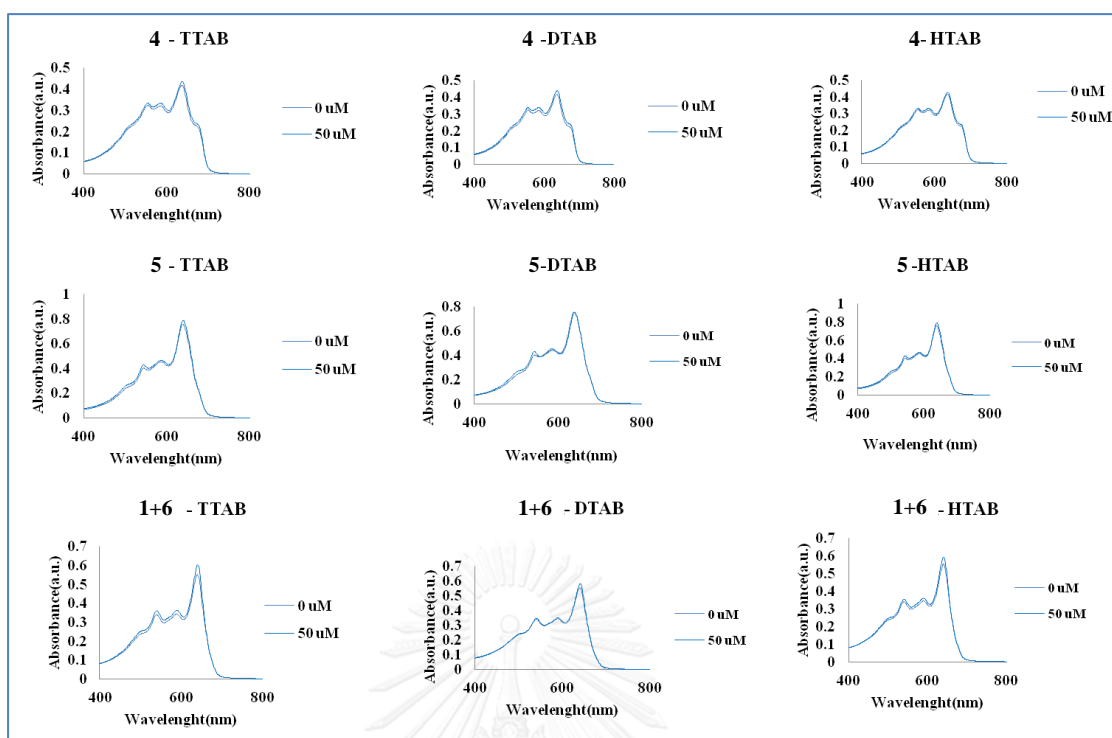


Figure 3.20. UV-visible spectrums of 4, 5, and 1+6 with TTAB, DTAB, and HTAB

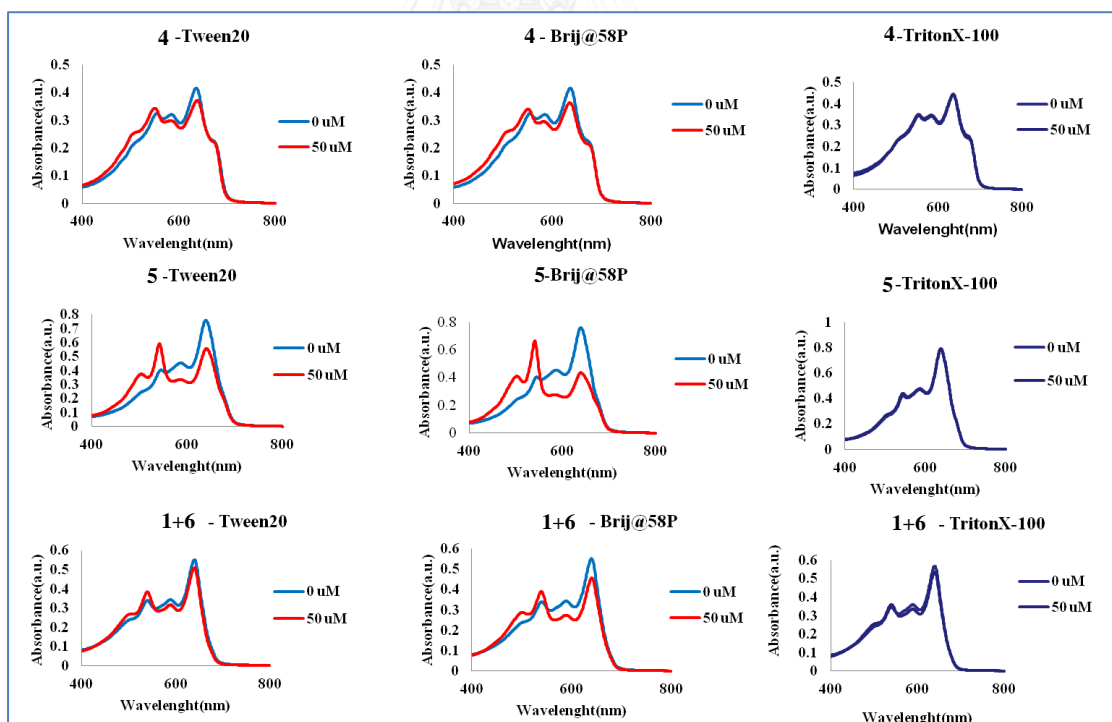


Figure 3.21. UV-visible spectrums of 4, 5, and 1+6 with Tween-20®, Brij®58P, and Triton®X-100

In order to confirm the degree of color change of PDA in the presence of various surfactants, the data from PDAs absorbance spectrums of surfactants at the concentration of 50 μM were converted to colorimetric response (% CR). The conversion percentage of the blue to red phase absorption and the histogram of the average % CR determined from nine replicates (3 samples \times 3 measurements) of the PDA/surfactant mixtures were presented in **Figure 3.22**. The anionic surfactants (SDC, SDS, and SDBS) induced strong color transition of PDAs **4**, **5**, and **1+6** provide more than 50% CR which support the results from naked eye observation. Tween-20® and Brij®58P also provide color change at he %CR in the range between 10-50%. The % CR of PDA/surfactant mixtures showed differentiable patterns of the responses for three anionic surfactants suggesting a possibility of surfactant identification. Since these PDA sols showed different colorimetric sensitivity toward different types of anionic surfactants demonstrated by the %CR, this suggested that they should also be potentially useful for constructing a sensor array for the identification of common anionic surfactants and some nonionic surfactants.

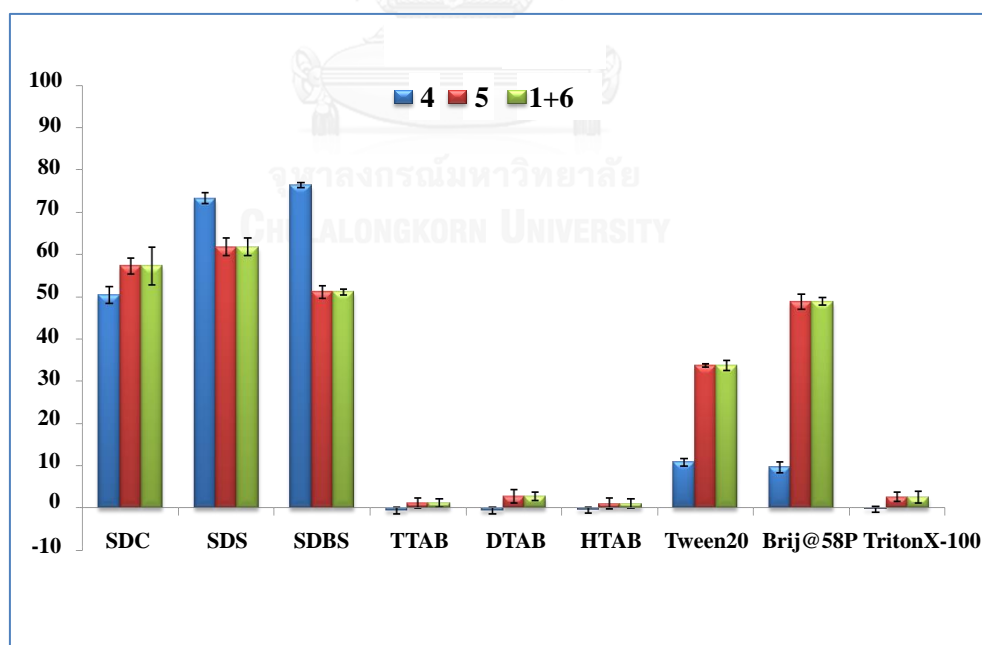


Figure 3.22. Histogram of % CR obtained from nine replicated measurements of nine mixtures of **4**, **5** and **1+6** (0.10 mM) and SDC, SDS, SDBS, TTAB, DTAB, HTAB, Tween-20®, Brij®58P and Triton® X-100 (50 μM).

3.2.1.1 Multivariate statistical analysis

Although, the **Figure 3.22** showed the color pattern for the sensing of anionic surfactants, the multidimensional data matrix (3 PDAs \times 3 surfactants \times 9 replicates) was further statistically analyzed by PCA. PCA transformed the dataset of % CR values into PC scores in which PC1 and PC2 accounted for 94.2 and 5.3% of total variance, respectively, as shown in PCA score plot (**Figure 3.23**). The PCA score plot showed clear clusters of SDC, SDS, SDBS, Tween-20®, and Brij®58P, but overlap for the rest of surfactants. The result from **Figure 3.23** confirms that structurally diverse amino containing PDAs were useful for the detection and identification of the common anionic surfactants and some nonionic surfactants, but not the cationic surfactants.

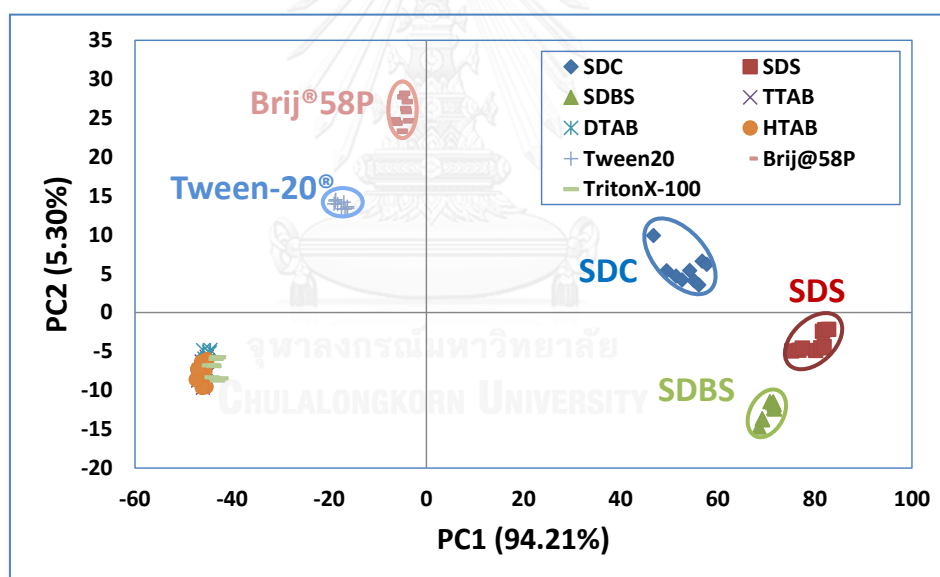


Figure 3.23. PCA score plot of % CR obtained from nine replicated measurements of nine mixtures of **4**, **5** and **1+6** (0.10 mM) and SDC, SDS, SDBS, TTAB, DTAB, HTAB, Tween-20®, Brij®58P and Triton® X-100 (50 μ M).

3.2.2 Colorimetric paper-based sensor array for VOCs discrimination

Volatile organic compounds (VOCs) are large group of carbon-based chemicals having a high vapor pressure that easily evaporate at room temperature. These

compounds affect the environment and health risk concerns. Along with carbon, they contain elements such as hydrogen, oxygen, fluorine, chlorine, bromine, sulfur or nitrogen. Volatile organic compounds are released from various sources such as vehicle exhaust, vapors from industrial facilities or biological sources mainly from plants which may cause outdoor air pollution. They are also released from decorative paints, varnishes and vehicle refinishing which make poor indoor air quality.

Polydiacetylene (PDA) is one among stimulus-responsive colorimetric substances [85] which possesses unique chromic properties that can be developed into practical colorimetric transducers in biosensors. The solvent-dependent color transition of PDA molecules is known to vary according to the PDA structure [83, 86]. Eighteen VOCs were selected as analytes in this study pentane, hexane, cyclohexane, toluene, o-xylene, benzene, ethyl ether, dichloromethane (DCM), 2-propanol, tetrahydrofuran (THF), chloroform (CHCl₃), ethanol (EtOH), ethyl acetate (EtOAc), acetone, methanol (MeOH), acetonitrile (MeCN), dimethylformamide (DMF), dimethylsulfoxide (DMSO). The paper based PDA colorimetric sensor array prepared by diacetylene monomers **1-8** was incubated with each vapor of VOCs. The images of color changes acquired by scanner and digital camera in TIF format were depicted in **Figure 3.24**.

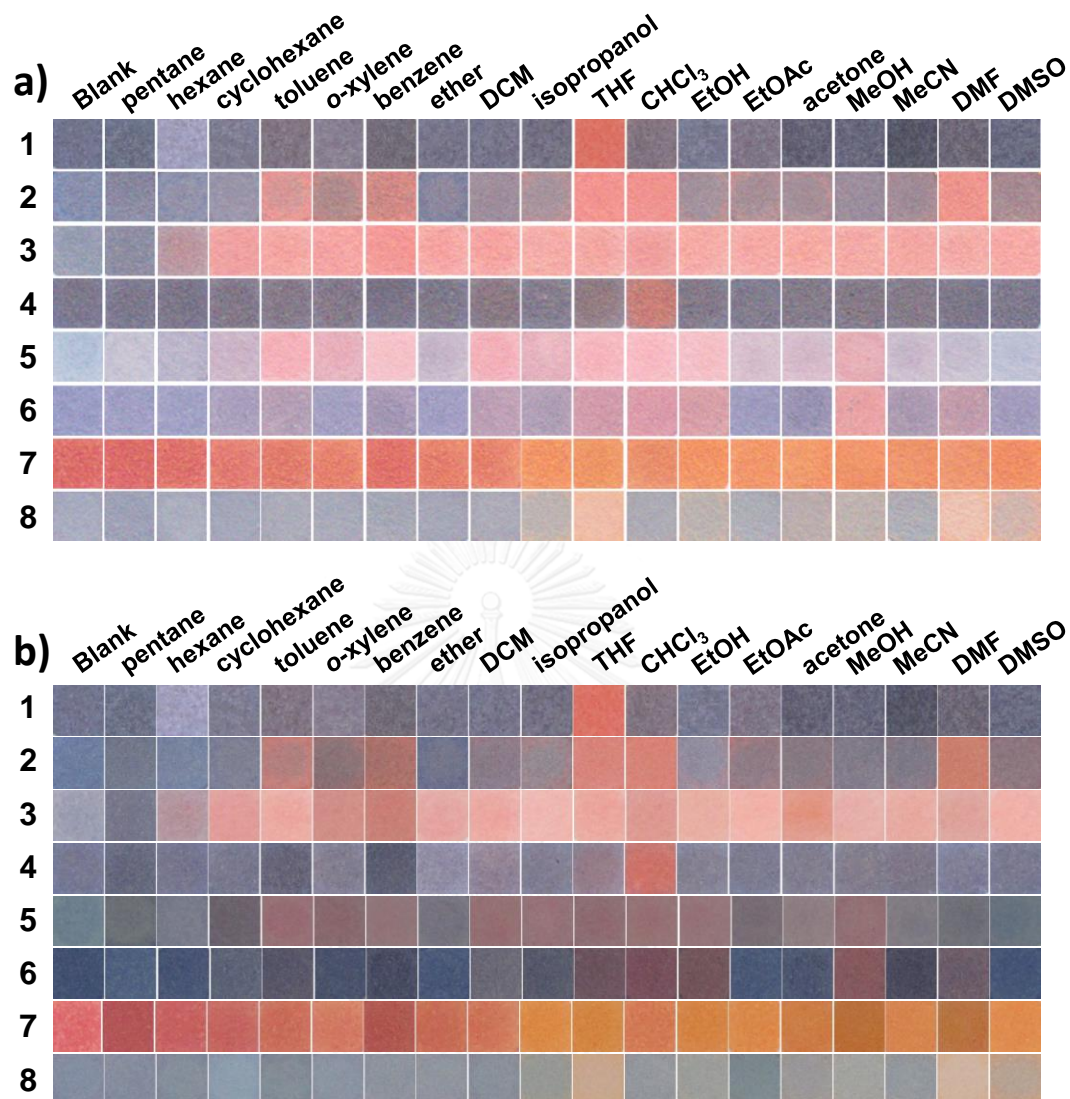


Figure 3.24. (a) Scanned images and (b) photographs of the paper-based PDA sensor array prepared from diacetylene monomers **1-8** before and after 60 min exposure to saturated vapors of volatile organic solvents at 30 °C. All images were acquired at 150 dpi.

Since the input devices of this study are scanner and digital camera, RGB color model was selected. TIF image files were converted to RGB values by AdobePhotoshop®. The colorimetric response of the PDA sensor can be quantified by a color change profile determined by subtracting the RGB values of the images before and after exposure of each PDA with VOCs vapor. **Figure 3.25** shows the

quantitative color change profile of paper-based PDA sensor array prepared from diacetylene monomers **1-8**.

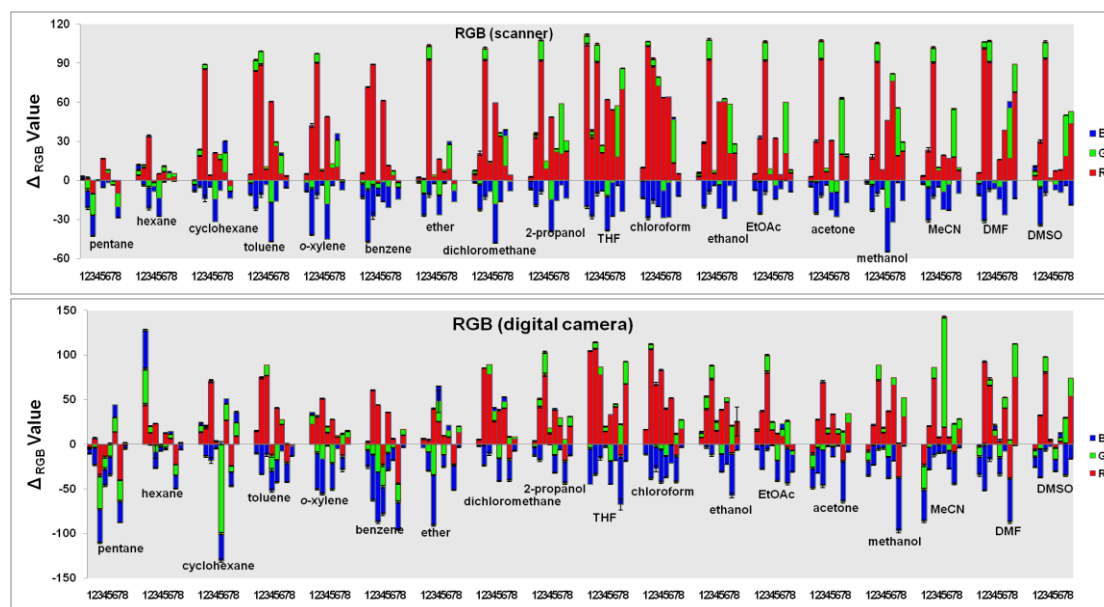


Figure 3.25. Quantitative color change profile of paper-based PDA sensor array prepared from diacetylene monomers **1-8** after 60 min exposure to saturated vapors of volatile organic solvents at 30 °C obtained from (a) scanner and (b) digital camera. Error bars represent standard deviations of the intensity of each polymer.

3.2.2.1 Multivariate Data analysis

Even though the histogram of color change profile as shown in Figure 3.18 demonstrated differentiable pattern of chromic change toward each VOC, computationally based discipline such as PCA is always a good choice to pursue to discover latent information from the data. To carry out this analysis, the data matrix of Δ_{RGB} color values from cropped images was prepared to generate a set of 5,184 colorimetric data (18 VOCs \times 12 replicates \times 8 polymers \times 3 values of each RGB) and applied to PCA. The two-dimension PCA score plots of Δ_{RGB} color values derived from scanned images and photographs of the paper-based PDA sensor array were shown in **Figure 3.25a** and **Figure 3.25b**, respectively.

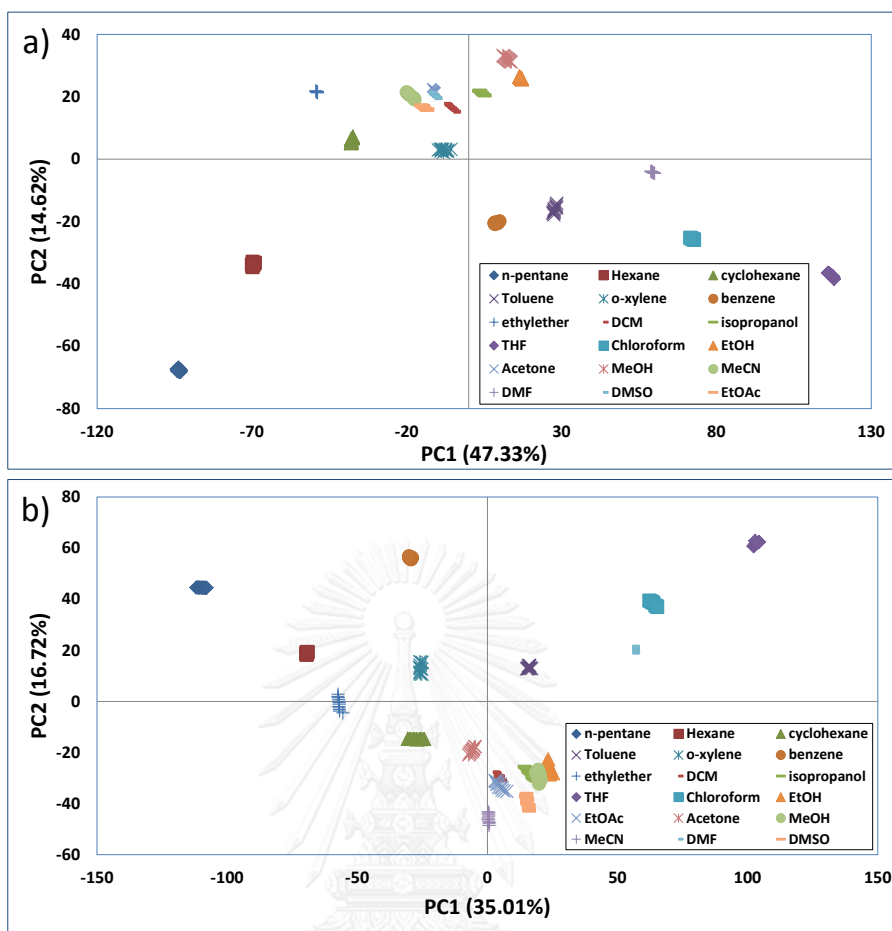


Figure 3.26. PCA score plot of paper-based 1-8 PDAs sensor array derived from (a) scanner and (b) digital camera, upon exposure to 18 solvents. Each data point represents an average of RGB value obtained from 12 replications.

PCA score plot from scanner (**Figure 3.26a**) showed that the first component (PC1) accounts for 47.33% of variance, while the second component (PC2) accounts for 14.62%, providing the sum of these two PCs of 61.95%. FDA is applied on selected principal component (PC) scores to test the discriminating accuracy of the sensor array. An accuracy levels of 100% obtained from FDA is achieved from the first two PC scores indicating an excellent discerned among 18 vapors of VOCs. On the other hand, PCA score plot from digital camera (**Figure 3.26b**) displayed the first component (PC1) accounts for 35.01% of variance, while the second component (PC2) accounts for 16.72%, providing the sum of these two PCs of 51.73%. An accuracy levels of 97.69% obtained from FDA was achieved from the first two PC

scores in **Figure 3.26b**. Not surprisingly, one drawback of the digital cameras is that the intensity of digital images is affected by lighting condition consequently resulting in lower discrimination power of the sensor.

3.2.2.2 Variable selection and reduction

The performance of the array can be improved non-informative sensing elements are reduced. This reduction would make the sensor more practically applicable. The plot of loading factors from PCA performed on RGB values of the scanned images was used to determine which individual PDA sensor provides most contribution to the total variance (**Figure 3.27**).

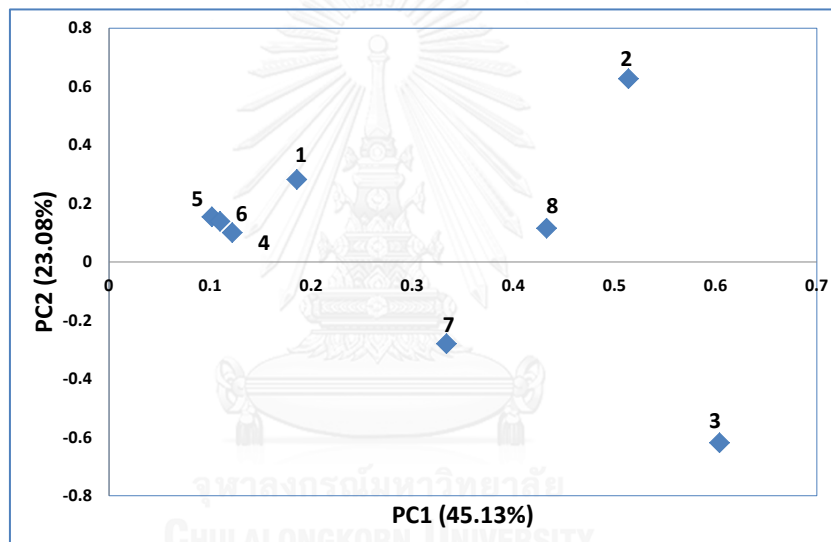


Figure 3.27. PCA loading plots of the eight element sensors in the developed arrays

Figure 3.27 showed that sensors **2**, **3** and **8** had higher influence in PC1 and PC2 than those sensors **1**, **4**, **5**, **6** and **7** of which the loading coordinates appear close to the origin point. Therefore, three PC score plots of the data from each pair of sensing elements selected from sensors **2**, **3** and **8** were generated. Again, FDA with cross validation was applied to test the classification accuracy of the PCA score plots obtained from each pair of sensors. Interestingly, **2+8** (**Figure 3.28a**) and **3+8** (**Figure 3.28b**) sensor pairs provided the classification accuracy of 100% while the **2+3** (**Figure 3.28c**) sensor pair gave a lower accuracy only at 88.43%.

These outcomes imply that the combination of the sensors from amphiphilic (2 or 3) and bolaamphiphilic monomers (8) is preferred for greater sensitivity variance toward the VOCs detection when comparing with the sensor pair containing only amphiphilic monomers (2+3 pair). The high sensitivity variance in turn benefits the discriminating efficiency of the sensor array.

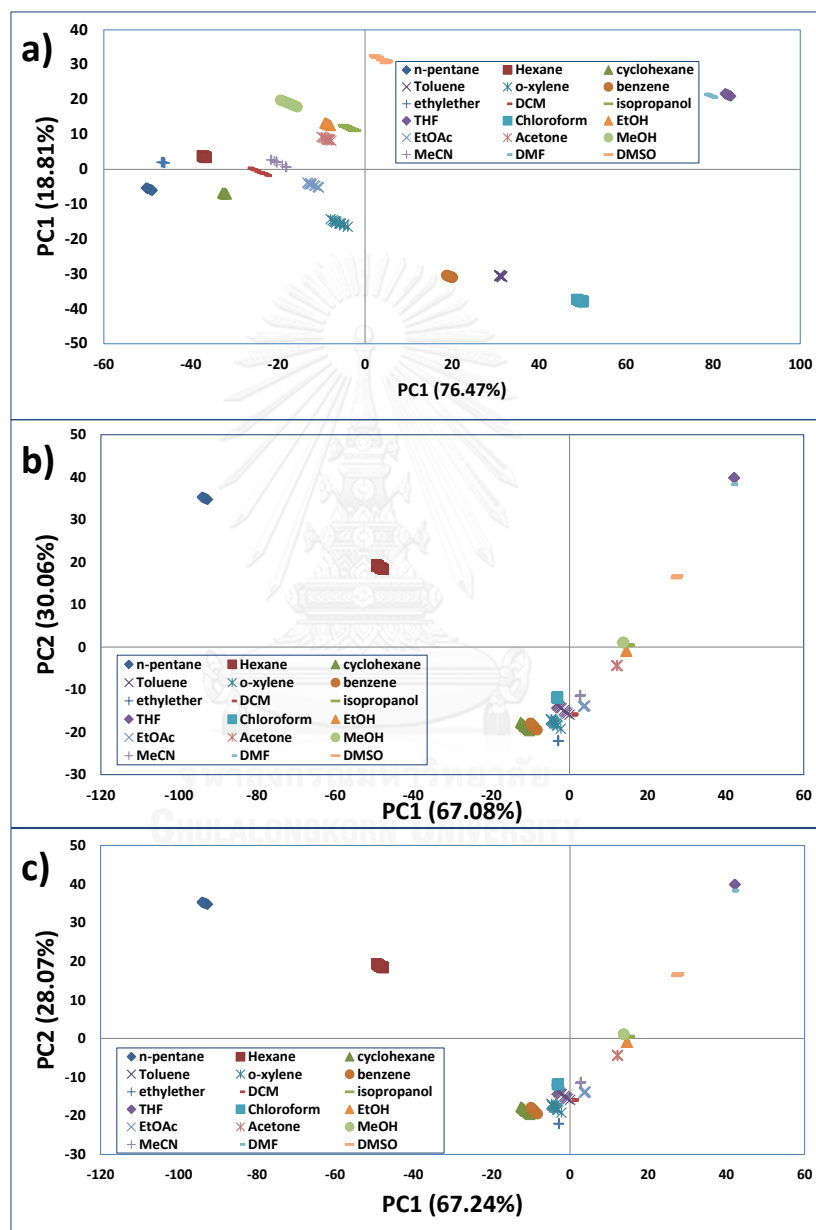


Figure 3.28. PCA score plot of RGB color changes obtained from the paper-based PDA sensor array prepared from (a) 2 and 8; (b) 3 and 8; (c) 2 and 3 after exposure to 18 VOC vapors

3.2.2.3 Colorimetric paper-based sensor array for automotive fuel discrimination

Automotive fuel consists mostly of organic compounds obtained by the fractional distillation of petroleum. However, due to high consumption of the world's gasoline, some alternatives are derived from non-fossil sources such as agricultural crops. In Thailand, for example, many types of automotive fuel have been commercially available such as gasoline octane 91 (B91), 10% ethanol blended gasoline widely known as gasohol octane 91 and 95 (G91 and 95), 20% ethanol blended gasoline (E20), diesel (D) and biodiesel (B5). Authentication of these fuels is of economic, legal and safety-critical importance. Counterfeit fuels are almost impossible to be perceived by consumers. Therefore, the scope of the study in the PDA sensor array was extended to detect commercial automotive fuels.

Subject to the government regulations, fuel dyes must be added to differentiate types of gasoline for prohibiting misuse and piracy (**Figure 3.29**). However, naked-eye color detection cannot investigate chemical components of the gasoline.



Figure 3.29. The commercial gasoline in various grades from Petroleum Authority of Thailand (A) and Esso company (B).

Ten automotive fuels from two petroleum companies, Petroleum Authority of Thailand (A) or Esso Company (B), were selected in this study, Gasohol 95 (G95), Gasohol 91 (G91), Gasoline 91 (B91), Diesel (D), E20 and Biodiesel (B5). Gasohol is gasoline (90%) blended with ethanol (10%). Diesel is a fractional distillate of petroleum fuel oil which is specifically used in diesel engines. E20 is a blend of 20% ethanol and gasoline 80%. Biodiesel refers to a diesel from natural sources, e.g.

vegetable oil or animal fat consisting of long-chain alkyl esters. The term B5 refers to 5% of biodiesel in 95% petroleum diesel.

Ten fuels were tested with the PDA sensors by the same methodology as those performed with the VOC vapors described above. **Figure 3.30** showed the scanned image of PDA paper-based array exposed to 10 gasoline vapors. Strong color change observed for the Gasohol 91, Gasohol 95 and E 20 can be explained by the fact that the content of EtOH played an important role in the color transition of array of PDA paper. The color change profiles of scanned images (before and after exposure with 10 gasoline vapors) were also prepared as the histogram as shown in **Figure 3.31**.

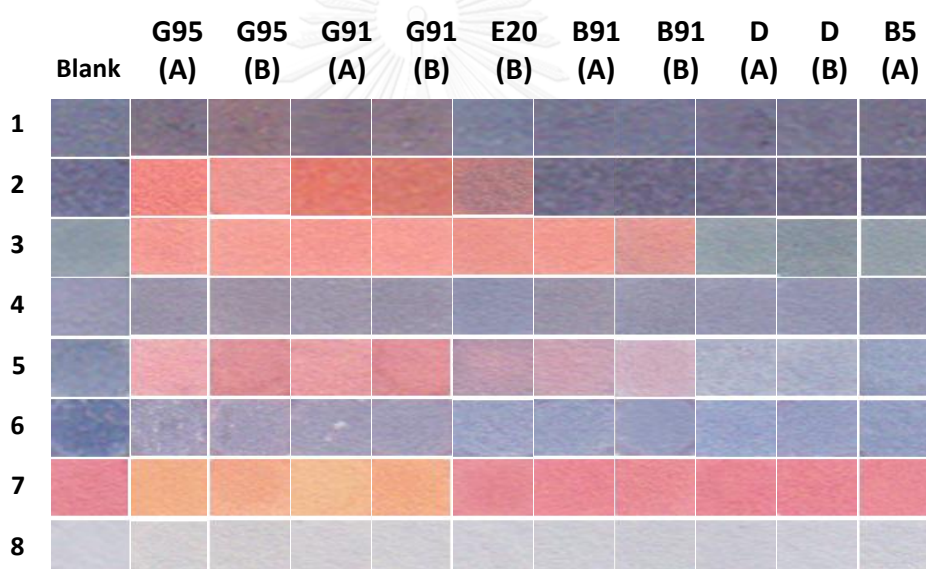


Figure 3.30. Scanned images of the PDA-based paper colorimetric sensor array prepared from diacetylene monomers **1-8** after exposure to vapors of gasoline.

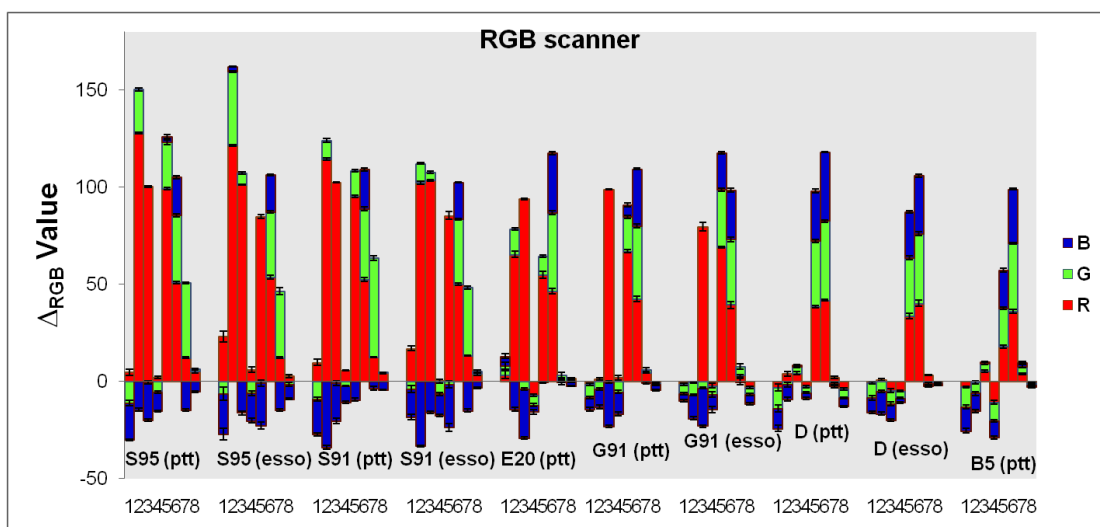


Figure 3.31. RGB color change profile of paper-based PDA sensor array prepared from diacethylene monomers **1-8** after exposure to saturated vapors of automotive fuels

PCA was applied to the dataset containing ΔRGB of 2,160 values (10 gasoline \times 9 repeats \times 8 PDAs \times 3 values of each RGB). **Figure 3.32** showed the PCA score plot having a total variance of 91.14% obtained from the first two PCs in which PC1 and PC2 accounted for 80.20% and 11.74%, respectively. As can be seen in **Figure 3.32**, gasoline of each company was discriminated according to their types. This color pattern induced by different properties of each gasoline imparts special properties to fuel in order to gain optimal engine performance. It should be noted that the cluster of B5 (A), D (A) and D (B) are closely located which can indicate that they have similar chemical and physical properties. 5% of biodiesel in B5 cannot make it distinct from the conventional diesel.

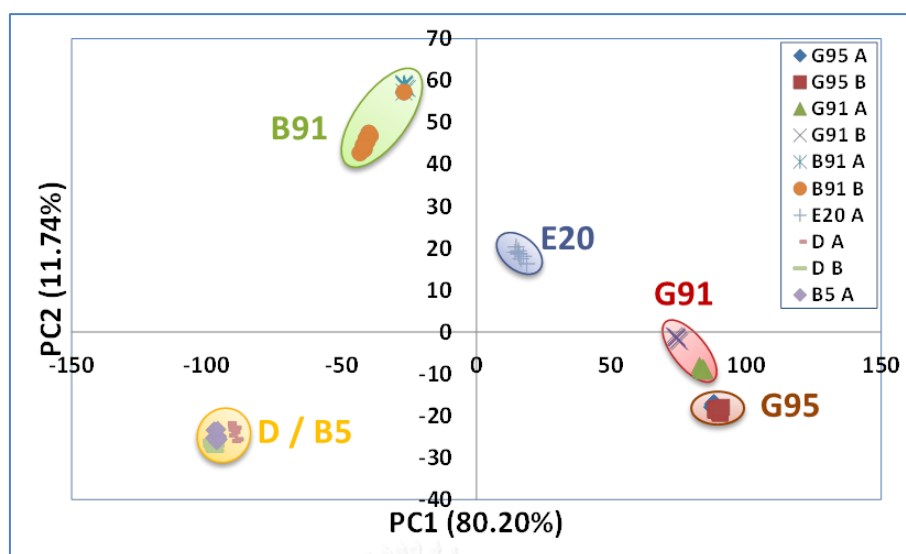


Figure 3.32. PCA score plot of Δ RGB data set obtained from paper-based PDA sensor array derived from diacetylene monomers **1-8** upon exposure to 10 gasoline. Each data point represents an average of RGB value obtained from 9 replications.

Even PDAs **2**, **3** and **8** were formerly identified as the most informative sensing elements for the identification of VOC vapors, photo images in **Figure 3.30** showed that **8** was insensitive to all type of fuels. Therefore, **2** and **3** were selected as the set of reduced variables to test if the discrimination ability can be improved. PCA was again performed on Δ RGB data set obtained from paper-based PDA sensor array derived from diacetylene monomers **2** and **3**. **Figure 3.33** showed the PC score plot yielded from the analysis.

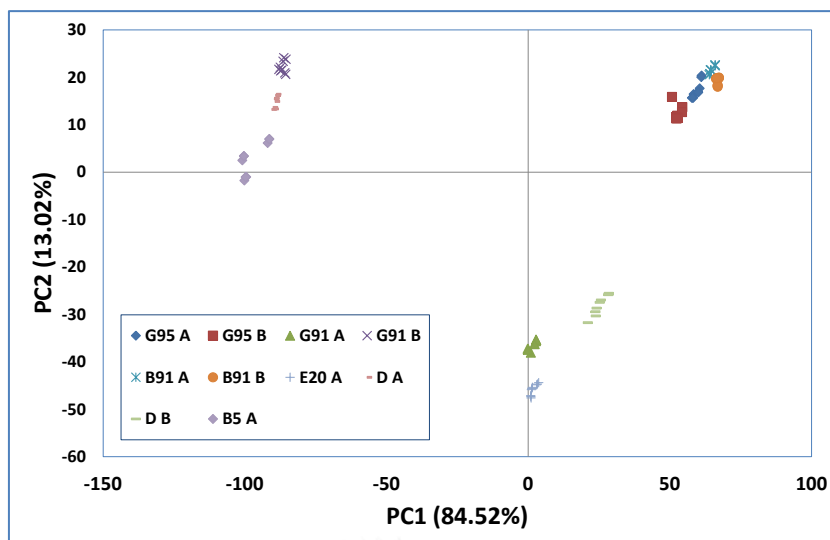


Figure 3.33. PCA score plot of RGB color changes obtained from the paper-based PDA sensor array derived from **2** and **3** upon exposure to automotive fuels (2 PDAs \times 3 replicates \times 10 automotive fuels \times 3 measurements of each RGB).

A total variance of 97.54% in the PCA score plot was obtained from the first two PCs in which PC1 and PC2 accounted for 84.52% and 13.02%, respectively. The plot also showed well-separated clusters of diesel (D), biodiesel (B5), E20, gasohol (H), gasoline (G) confirming the high fuel discrimination ability of sensors **2** and **3**.

CHAPTER IV

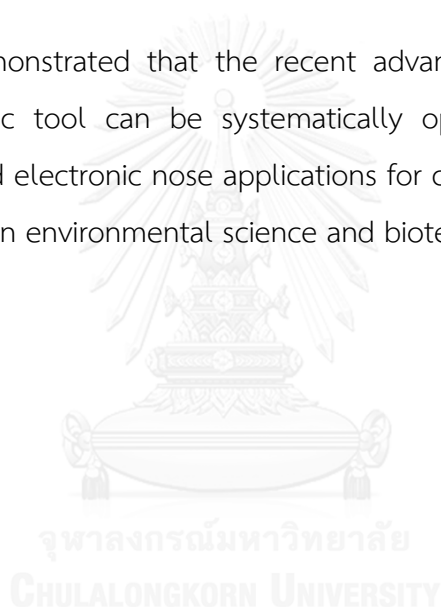
CONCLUSION

In view of the increasing global market demands in chemical biosensor [87], rapid detection and identification in different areas of applications has expanded considerably over the last few years. This work has shown that optical spectroscopic sensors and multivariate statistical analyses can be applied successfully for discriminating chemicals and biological samples. A fluorescence sensor array with synthetic fluorophores was been systematically developed for the identification of commercial milk samples according to their thermal treatment processes. The array and analysis method were sequentially tested against samples in the order of degree of protein complexity and similarity. The sequence of the samples started from randomly selected pure proteins, then pure milk proteins, followed by commercial milk samples obtained from different sources, and finally on commercial cow milk samples undergone different thermal treatments. At the beginning, multivariate statistical analyses, principal component analysis (PCA) and factorial discriminate analysis (FDA), were applied for pattern recognition of the fluorescence responses and capable to discover similarities within proteins with wide range of molecular weights and isoelectric points, milk proteins as well as milks from different sources. However, PCA failed to classify cow milk samples with different thermal treatments while linear discriminant analysis (LDA) provided excellent classification with 100% accuracy, even when the milk samples were derived from different manufacturers. A fluorescence sensor array composed of only the two fluorophores has been also applied for discriminating 8 strains of foodborne pathogens. LDA was also more powerful in recognizing the similarities within the group of bacteria samples induced by fluorophore/cell surface interactions. Notably, the discrimination of ETEC and E.coli can be observed via this technique due to CS expressing charged residues at ETEC cell surface resulting in fluorescent pattern distinguished from E. coli.

Multivariate statistical analyses applied in sensor array based on colorimetric responses of polydiacetylene were also developed for identification of anionic

surfactant and volatile organic compounds (VOCs). The results of PCA performed on colorimetric response (%CR) clearly showed discrimination of anionic surfactants (SDC, SDS and SDBS) from other classes of surfactants at micro molar concentration level (50 μM). In addition, a paper-based polydiacetylene colorimetric sensor array prepared from amphiphilic and bolaamphiphilic diacetylene monomers for identification of VOCs in vapor phase has also been developed. Fingerprint pattern of the irreversible color change pattern as measured by RGB values were analyzed by PCA to distinguish all 18 VOCs tested. This paper-based sensor is also capable to classify various types of automotive fuels including gasoline, gasohol and diesel.

This study demonstrated that the recent advances in optical sensor arrays coupled chemometric tool can be systematically optimized for wider range of electronic tongue and electronic nose applications for classification and identification of complex mixtures in environmental science and biotechnology industries.



REFERENCES

1. Askim, J.R., M. Mahmoudi, and K.S. Suslick, *Optical sensor arrays for chemical sensing: the optoelectronic nose*. Chemical Society Reviews, 2013. **42**(22): p. 8649-8682.
2. Grieshaber, D., et al., *Electrochemical Biosensors - Sensor Principles and Architectures*. Sensors, 2008. **8**(3): p. 1400-1458.
3. Karoui, R., G. Mazerolles, and É. Dufour, *Spectroscopic techniques coupled with chemometric tools for structure and texture determinations in dairy products*. International Dairy Journal, 2003. **13**(8): p. 607-620.
4. Brereton, R.G., *Introduction*, in *Applied Chemometrics for Scientists*. 2007, John Wiley & Sons, Ltd. p. 1-8.
5. Chau, F.-t., et al., *Introduction*, in *Chemometrics*. 2004, John Wiley & Sons, Inc. p. 1-21.
6. Hibbert, D.B., et al., *IUPAC project: A glossary of concepts and terms in chemometrics*. Analytica Chimica Acta, 2009. **642**(1-2): p. 3-5.
7. Photonics4life. *Chemometric Analysis 2010* [cited 2014 12 November]; Available from: <http://www.photonics4life.eu/Consortium/P4L-DB/Browse/Data-Analysis/Chemometric-Analysis>.
8. Kruzlicova, D., M. Danihelova, and M. Veverka, *Quantitative Structure-Antioxidant Activity Relationship of Quercetin And Its New Synthetised Derivatives*. Nova Biotechnologica et Chimica, 2012. **11**(1): p. 37-44.
9. Jain, A.K., R.P.W. Duin, and J. Mao, *Statistical Pattern Recognition: A Review*. IEEE Trans. Pattern Anal. Mach. Intell., 2000. **22**(1): p. 4-37.
10. Isomaa, K., *Chemometric Methods in Plant Metabolomics in Chemistry*. 2013, University of Helsinki. p. 100.
11. Li Vigni, M., C. Durante, and M. Cocchi, *Chapter 3 - Exploratory Data Analysis*, in *Data Handling in Science and Technology*, M. Federico, Editor. 2013, Elsevier. p. 55-126.

12. Trygg, J. and T. Lundstedt, *Chapter 6 - Chemometrics Techniques for Metabonomics*, in *The Handbook of Metabonomics and Metabolomics*, J.C. Lindon, J.K. Nicholson, and E. Holmes, Editors. 2007, Elsevier Science B.V.: Amsterdam. p. 171-199.
13. Dufour, É., et al., *Investigation of variety, typicality and vintage of French and German wines using front-face fluorescence spectroscopy*. *Analytica Chimica Acta*, 2006. **563**(1-2): p. 292-299.
14. Palacios, M.A., et al., *Rational Design of a Minimal Size Sensor Array for Metal Ion Detection*. *Journal of the American Chemical Society*, 2008. **130**(31): p. 10307-10314.
15. Welling, M. *Fisher Linear Discriminant Analysis*. Available from: http://www.ics.uci.edu/~welling/classnotes/papers_class/Fisher-LDA.pdf.
16. Raschka, S. *Linear Discriminant Analysis bit by bit*. 2014 [cited 2014 12 November]; Available from: http://sebastianraschka.com/Articles/2014_python_lda.html.
17. Berrueta, L.A., R.M. Alonso-Salces, and K. Héberger, *Supervised pattern recognition in food analysis*. *Journal of Chromatography A*, 2007. **1158**(1-2): p. 196-214.
18. Patel, U.K.J.K.B.S.A.P.P.J., *A Comparative Study of PCA & LDA Human Face Recognition Methods*. 2011: National Conference on Recent Trends in Engineering & Technology.
19. Zahorian, S.A. and H. Hu. *Nonlinear Dimensionality Reduction Methods for Use with Automatic Speech Recognition*. [cited 2014 12 November]; Available from: <http://www.intechopen.com/books/speech-technologies/nonlinear-dimensionality-reduction-methods-for-use-with-automatic-speech-recognition>.
20. Wold, S., *Chemometrics; what do we mean with it, and what do we want from it?* *Chemometrics and Intelligent Laboratory Systems*, 1995. **30**(1): p. 109-115.
21. Guyon, I. and A.e. Elisseeff, *An Introduction to Variable and Feature Selection*. *Journal of Machine Learning Research*, 2003. **3**: p. 1157-1182.

22. Guyon, I. and A. Elisseeff, *An Introduction to Feature Extraction*, in *Feature Extraction*, I. Guyon, et al., Editors. 2006, Springer Berlin Heidelberg. p. 1-25.
23. *Fluorescence Sensors and Biosensors*. 2005 CRC Press.
24. Mazerolles, G., et al., *Chemometric methods for the coupling of spectroscopic techniques and for the extraction of the relevant information contained in the spectral data tables*. *Chemometrics and Intelligent Laboratory Systems*, 2002. **63**(1): p. 57-68.
25. Sádecká, J.a.T., J., *Fluorescence spectroscopy and chemometrics in the food classification – a review*. *Czech Journal of Food Sciences*, 2007. **25**(4): p. 159–173.
26. Hawe, A., M. Sutter, and W. Jiskoot, *Extrinsic Fluorescent Dyes as Tools for Protein Characterization*. *Pharmaceutical Research*, 2008. **25**(7): p. 1487-1499.
27. *Instrumentation for Fluorescence Spectroscopy*, in *Principles of Fluorescence Spectroscopy*, J. Lakowicz, Editor. 2006, Springer US. p. 27-61.
28. Naomi E. Balaban, J.E.B., *The Handy Anatomy Answer Book*. 2008: Visible Ink Press.
29. (NASA), C.C.I. *Energy: The Driver of Climate* 2013; Available from: <http://www.ces.fau.edu/nasa/module-2/radiation-sun.php>.
30. Bohren, C.F. and E.E. Clothiaux. *Fundamentals of Atmospheric Radiation*. 2008: Wiley-VCH Verlag GmbH. I-XVIII.
31. Boulton, M., *The Colour Wheel*, in *A Practical Guide to Designing for the Web*. 2009, Mark Boulton Design Ltd. p. 135-140.
32. White, P. *New Thoughts on RGB and CMYK*. 2014 [cited 2014 12 November]; Available from: http://www.visionsfirst.com/client-tools/visions-news.aspx?jx=fp_38.
33. Dufour, E. and A. Riaublanc, *Potentiality of spectroscopic methods for the characterisation of dairy products. I. Front-face fluorescence study of raw, heated and homogenised milks*. *Lait*, 1997. **77**(6): p. 657-670.
34. Miranda, O.R., et al., *Array-Based Sensing of Proteins Using Conjugated Polymers*. *Journal of the American Chemical Society*, 2007. **129**(32): p. 9856-9857.

35. You, C.-C., et al., *Detection and identification of proteins using nanoparticle-fluorescent polymer /'chemical nose/' sensors*. *Nat Nano*, 2007. **2**(5): p. 318-323.
36. Wan, Y., et al., *Quaternized magnetic nanoparticles-fluorescent polymer system for detection and identification of bacteria*. *Biosensors and Bioelectronics*, 2014. **55**(0): p. 289-293.
37. Janzen, M.C., et al., *Colorimetric Sensor Arrays for Volatile Organic Compounds*. *Analytical Chemistry*, 2006. **78**(11): p. 3591-3600.
38. Hou, C., et al., *Colorimetric artificial tongue for protein identification*. *Biosensors and Bioelectronics*, 2011. **26**(10): p. 3981-3986.
39. Khumsri, A., *Cyanide Fluorescent Sensors From Diphenylacetylene Derivatives*, in *Chemistry*. 2012, Chulalongkorn University. p. 81.
40. Niamnont, N., et al., *A Polyanionic Dendritic Fluorophore for Selective Detection of Hg²⁺ in Triton X-100 Aqueous Media*. *Organic Letters*, 2009. **11**(13): p. 2768-2771.
41. Niamnont, N., *Synthesis and Fluorescent Properties of Dendritic Polyelectrolyte Fluorophores in Chemistry*. 2010, Chulalongkorn.
42. Phollookin, C., et al., *Tuning Down of Color Transition Temperature of Thermochromically Reversible Bisdiynamide Polydiacetylenes*. *Macromolecules*, 2010. **43**(18): p. 7540-7548.
43. Deshayes, S. and G. Divita, *Chapter Four - Fluorescence Technologies for Monitoring Interactions Between Biological Molecules In Vitro*, in *Progress in Molecular Biology and Translational Science*, C.M. May, Editor. 2013, Academic Press. p. 109-143.
44. Tan, C., M.R. Pinto, and K.S. Schanze, *Photophysics, aggregation and amplified quenching of a water-soluble poly(phenylene ethynylene)*. *Chemical Communications*, 2002(5): p. 446-447.
45. Swager, T. and J. Zheng, *Poly(arylene ethynylene)s in Chemosensing and Biosensing*, in *Poly(arylene ethynylene)s*, C. Weder, Editor. 2005, Springer Berlin Heidelberg. p. 151-179.

46. Poulli, K.I., G.A. Mousdis, and C.A. Georgiou, *Classification of edible and lampante virgin olive oil based on synchronous fluorescence and total luminescence spectroscopy*. *Analytica Chimica Acta*, 2005. **542**(2): p. 151-156.
47. Karoui, R., E. Dufour, and J. De Baerdemaeker, *Monitoring the molecular changes by front face fluorescence spectroscopy throughout ripening of a semi-hard cheese*. *Food Chemistry*, 2007. **104**(1): p. 409-420.
48. Karoui, R., J. De Baerdemaeker, and E. Dufour, *A comparison and joint use of mid infrared and fluorescence spectroscopic methods for differentiating between manufacturing processes and sampling zones of ripened soft cheeses*. *European Food Research and Technology*, 2008. **226**(4): p. 861-870.
49. Mutihac, L. and R. Mutihac, *Mining in chemometrics*. *Analytica Chimica Acta*, 2008. **612**(1): p. 1-18.
50. Fox, P.F., *Chapter 1 - Milk: an overview*, in *Milk Proteins*, A.T.B. Singh, Editor. 2008, Academic Press: San Diego. p. 1-54.
51. Iwabuchi, S. and F. Yamauchi, *Determination of glycinin and .beta.-conglycinin in soybean proteins by immunological methods*. *Journal of Agricultural and Food Chemistry*, 1987. **35**(2): p. 200-205.
52. Shukla, R. and M. Cheryan, *Zein: the industrial protein from corn*. *Industrial Crops and Products*, 2001. **13**(3): p. 171-192.
53. Wouters, J.T.M., et al., *Microbes from raw milk for fermented dairy products*. *International Dairy Journal*, 2002. **12**(2-3): p. 91-109.
54. Parris, N., J.M. Purcell, and S.M. Ptashkin, *Thermal denaturation of whey proteins in skim milk*. *Journal of Agricultural and Food Chemistry*, 1991. **39**(12): p. 2167-2170.
55. Jeanson, S., et al., *Characterization of the Heat Treatment Undergone by Milk Using Two Inhibition ELISAs for Quantification of Native and Heat Denatured α -Lactalbumin*. *Journal of Agricultural and Food Chemistry*, 1999. **47**(6): p. 2249-2254.
56. Dutson, T.R. and M.W. Orcutt, *Chemical changes in proteins produced by thermal processing*. *Journal of Chemical Education*, 1984. **61**(4): p. 303.

57. Dupont, D., O. Rolet-Repecaud, and S. Muller-Renaud, *Determination of the Heat Treatment Undergone by Milk by Following the Denaturation of α -Lactalbumin with a Biosensor*. Journal of Agricultural and Food Chemistry, 2004. **52**(4): p. 677-681.
58. Macej, O., S. Jovanovic, and J. Denin-Djurdjevic, *The influence of high temperatures on milk proteins*. Chemical Industry (Yugoslavia), 2002.
59. Cook, B.B., et al., *The effect of heat treatment on the nutritive value of milk proteins. III. The effect of heat on casein, lactalbumin, and their lactose-induced derivatives, with special reference to digestibility and rate of release of lysine, methionine and tryptophan*. The Journal of nutrition, 1951. **44**(2): p. 217.
60. Dong Chen, X., *Mathematical analysis of powder discharge through longitudinal slits in a slowly rotating drum: Objective measurements of powder flowability*. Journal of food engineering, 1994. **21**(4): p. 421-437.
61. Havea, P., *Protein interactions in milk protein concentrate powders*. International Dairy Journal, 2006. **16**(5): p. 415-422.
62. Anema, S.G., *Chapter 8 - The whey proteins in milk: thermal denaturation, physical interactions and effects on the functional properties of milk*, in *Milk Proteins*, A.T.B. Singh, Editor. 2008, Academic Press: San Diego. p. 239-281.
63. Singh, H., *Heat stability of milk*. International Journal of Dairy Technology, 2004. **57**(2-3): p. 111-119.
64. Kulmyrzaev, A.A., D. Leveux, and É. Dufour, *Front-Face Fluorescence Spectroscopy Allows the Characterization of Mild Heat Treatments Applied to Milk. Relations with the Denaturation of Milk Proteins*. Journal of Agricultural and Food Chemistry, 2005. **53**(3): p. 502-507.
65. Fdez-Ortiz de Vallejuelo, S., et al., *Pattern recognition and classification of sediments according to their metal content using chemometric tools. A case study: The estuary of Nerbioi-Ibaizabal River (Bilbao, Basque Country)*. Chemosphere, 2011. **85**(8): p. 1347-1352.

66. Hixon, J. and Y. Reshetnyak, *Algorithm for the Analysis of Tryptophan Fluorescence Spectra and Their Correlation with Protein Structural Parameters*. Algorithms, 2009. **2**(3): p. 1155-1176.
67. Velusamy, V., et al., *An overview of foodborne pathogen detection: In the perspective of biosensors*. Biotechnology Advances, 2010. **28**(2): p. 232-254.
68. Singh, A., S. Poshtiban, and S. Evoy, *Recent Advances in Bacteriophage Based Biosensors for Food-Borne Pathogen Detection*. Sensors, 2013. **13**(2): p. 1763-1786.
69. Lampel, K.A. *Bad Bug Book 2nd Edition*. 2012 [cited 2014 18 October]; Available from:
<http://www.fda.gov/downloads/Food/FoodSafety/Foodbornellness/FoodbornellnessFoodbornePathogensNaturalToxins/BadBugBook/UCM297627.pdf>.
70. *FOODBORNE ILLNESS Common Agents Causing Foodborne Illness, Grouped by Symptoms*. 2009 [cited 2014 18 October]; Available from:
<http://www.doh.wa.gov/Portals/1/Documents/5100/foodchart.pdf>.
71. *Agents of foodborne illness : a technical series summarising key information on microorganisms associated with foodborne illness / Food Standards Australia New Zealand*, ed. Z. Food Standards Australia New. 2013, Canberra: Food Standards Australia New Zealand.
72. Tortora, G.J., B.R. Funke, and C.L. Case, *Microbiology: an introduction*. 2010: Pearson Benjamin Cummings.
73. Farber, J.M. and P.I. Peterkin, *Listeria monocytogenes, a food-borne pathogen*. Microbiological Reviews, 1991. **55**(3): p. 476-511.
74. Fiedler, F., *Biochemistry of the cell surface of Listeria strains: A locating general view*. Infection, 1988. **16**(2): p. S92-S97.
75. Moon, H.W., R.E. Isaacson, and J. Pohlenz, *Mechanisms of association of enteropathogenic Escherichia coli with intestinal epithelium*. The American Journal of Clinical Nutrition, 1979. **32**(1): p. 119-27.
76. Cravioto, A., S.M. Scotland, and B. Rowe, *Hemagglutination activity and colonization factor antigens I and II in enterotoxigenic and non-*

- enterotoxigenic strains of Escherichia coli isolated from humans*. Infection and Immunity, 1982. **36**(1): p. 189-197.
77. Svennerholm, A.M., et al., *Roles of different coli surface antigens of colonization factor antigen II in colonization by and protective immunogenicity of enterotoxigenic Escherichia coli in rabbits*. Infection and Immunity, 1990. **58**(2): p. 341-346.
78. Alves, C.S., et al., *Escherichia coli Cell Surface Perturbation and Disruption Induced by Antimicrobial Peptides BP100 and pepR*. Journal of Biological Chemistry, 2010. **285**(36): p. 27536-27544.
79. Savarino, S., *Adhesin as immunogen against enterotoxigenic Escherichia coli*. 2006, US Navy: USA.
80. Choi, H., et al., *Synthesis of Poly (ethylene glycol)-Polydiacetylene Conjugates and Their Micellar and Chromic Characteristics*. Journal of nanoscience and nanotechnology, 2008. **8**(10): p. 5104-5108.
81. Ahn, D.J., et al., *Colorimetric Reversibility of Polydiacetylene Supramolecules Having Enhanced Hydrogen-Bonding under Thermal and pH Stimuli*. Journal of the American Chemical Society, 2003. **125**(30): p. 8976-8977.
82. Chance, R.R., *Chromism in Polydiacetylene Solutions and Crystals*. Macromolecules, 1980. **13**(2): p. 396-398.
83. Yoon, J., S.K. Chae, and J.-M. Kim, *Colorimetric Sensors for Volatile Organic Compounds (VOCs) Based on Conjugated Polymer-Embedded Electrospun Fibers*. Journal of the American Chemical Society, 2007. **129**(11): p. 3038-3039.
84. Su, Y.-l., et al., *Biosensor signal amplification of vesicles functionalized with glycolipid for colorimetric detection of Escherichia coli*. Journal of Colloid and Interface Science, 2005. **284**(1): p. 114-119.
85. Lee, J., et al., *Hydrochromic conjugated polymers for human sweat pore mapping*. Nat Commun, 2014. **5**.
86. Jiang, H., et al., *Polydiacetylene-based colorimetric sensor microarray for volatile organic compounds*. Sensors and Actuators B: Chemical, 2010. **143**(2): p. 789-794.

87. Sadana, A. and N. Sadana, *Handbook of Biosensors and Biosensor Kinetics*. 2010: Elsevier Science.





APPENDIX

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Table A1. Proteins structures, molecular weight and isoelectric point

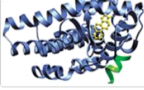
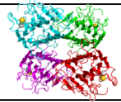

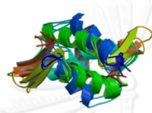
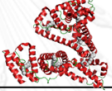


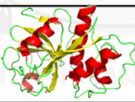
Proteins	Structure	Mw (Dalton)	pI
Bovine Serum Albumin		66,000	4.7
Concanavalin A		26,500/monomer unit	5.1
Cytochrome C		12,000	10
Histone - core histones (H2A, H2B, H3 and H4) - linker histones (H1 and H5)		9, 950-13,200	11
Human Serum Albumin		66,000	5
Lysozyme		14,600	11
Myoglobin		16,700	7.2
Papain		23,400	8-9

Table A2. Milk proteins structures, molecular weight and isoelectric point

Proteins	Mw (kDa)	pI
BSA	66.3	4.8
α -Casein	23-27	4.1
β -Casein	24	4.5
α -Lactalbumin	36	4.2-4.5
β -Lactoglobulin	14.4	5.3

Table A3. Composition of cow milk proteins

	grams/ liter	% of total protein
Total Proteins	33	100
Total Caseins	26	79.5
- α s1 casein	10	30.6
α s2 casein	2.6	8.0
- β casein	9.3	28.4
- κ casein	3.3	10.1
Total Whey Proteins	6.3	19.3
α lactalbumin	1.2	3.7
β lactoglobulin	3.2	9.8
BSA	0.4	1.2
Immunoglobulins	0.7	2.1
Protease peptone	0.8	2.4

Figure A1 LDA score plot obtained from fluorescence responses in emission range of 420-560 nm of (a) **6** at 99.07 % correct classification and (b) 1,8-ANS at 97.84 % correct classification. Oval outlines indicate groups of the milk samples of same thermal treatment at 99% confidence level

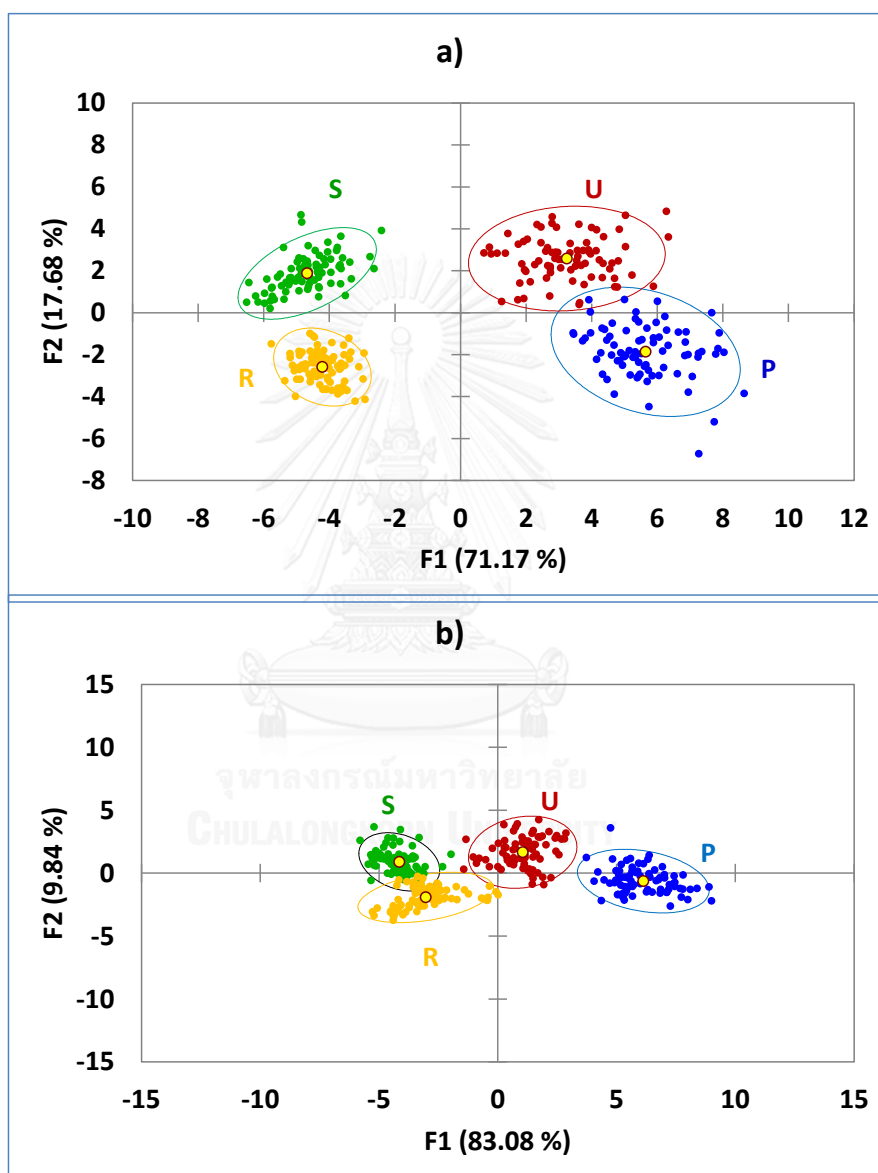


Figure A2 LDA score plot of first two discriminant factors (F1 and F2) obtained from ΔI values of fluorophores 1, 2 and 5 upon mixing with each bacteria sample

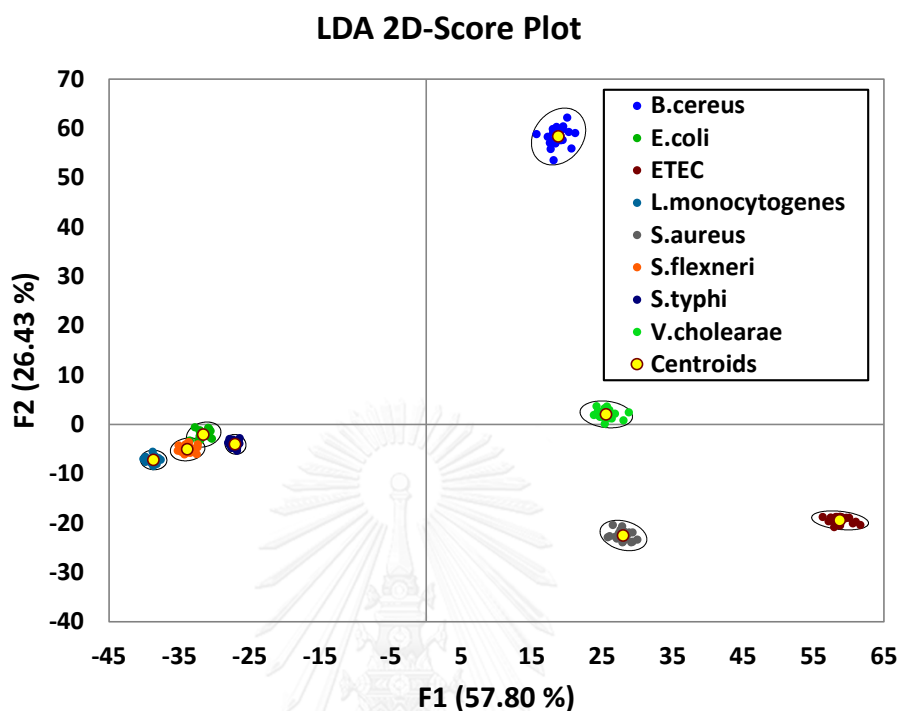


Figure A3. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 10 μM

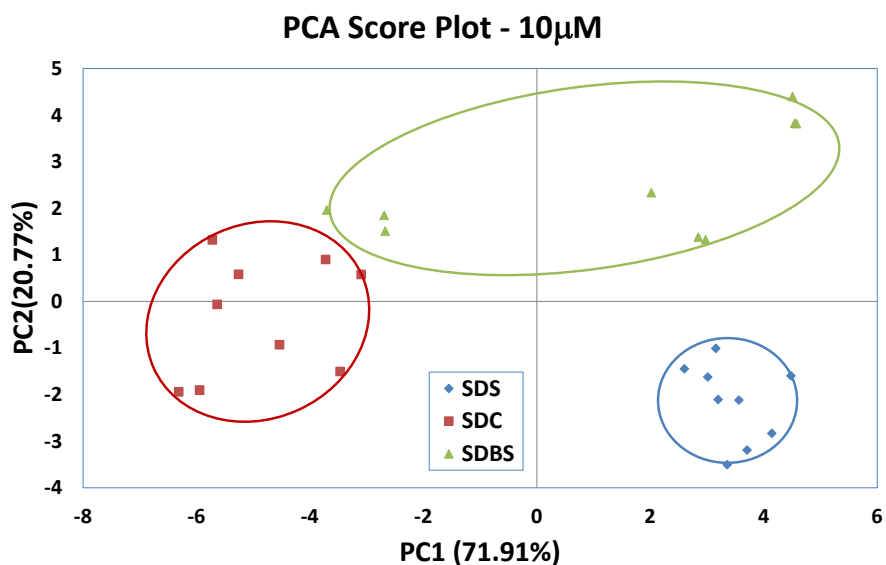


Figure A4. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 20 μM

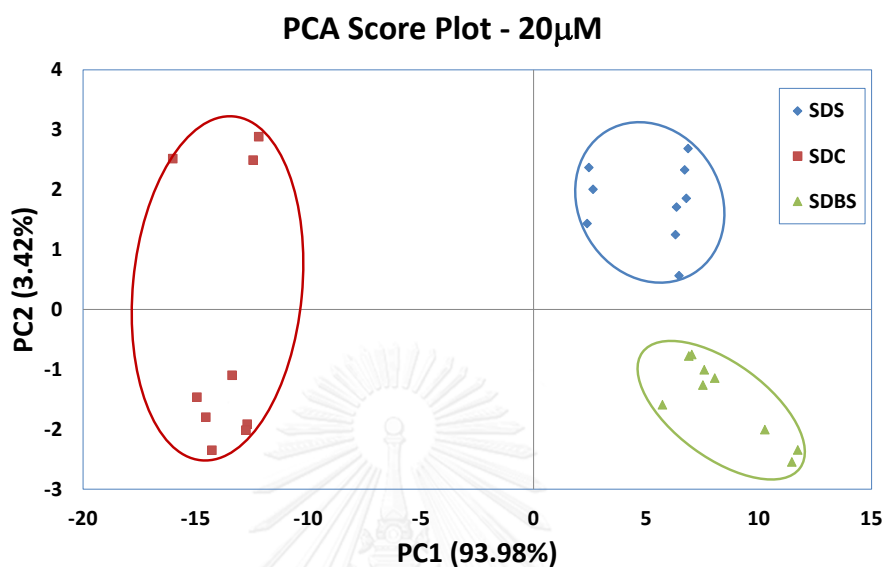


Figure A6. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 40 μM

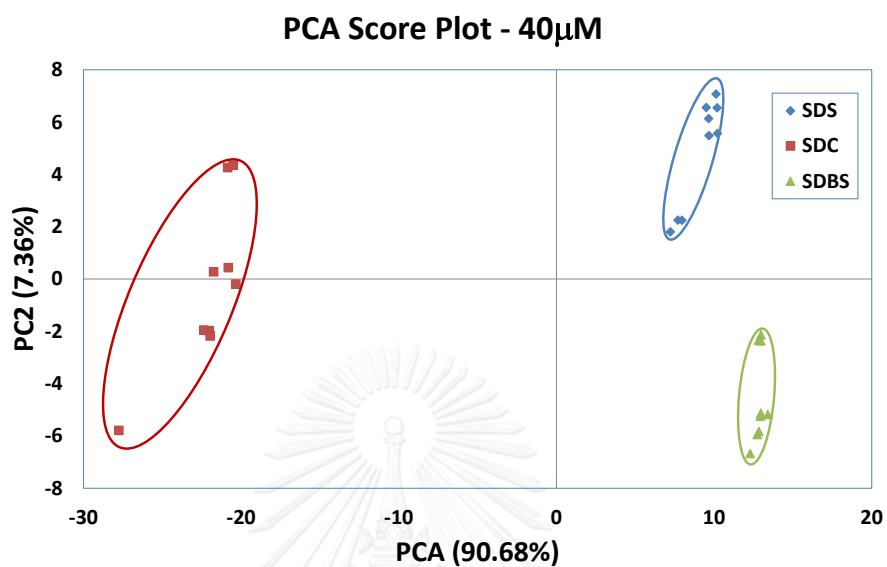


Figure A7. PCA score plot obtained from %CR of 3 anionic surfactants at a concentration of 50 μM

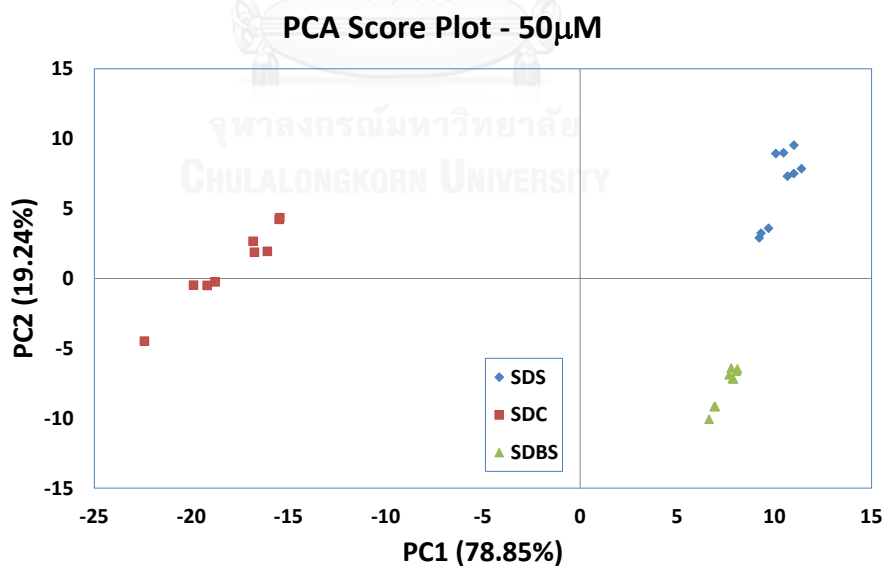


Figure A8. PCA score plot of paper-based 1-8 PDAs sensor array derived from (a) RGB value and (b) CMYK value from scanner upon exposure to 18 solvents. Each data point represents an average of color values obtained from 12 replications.

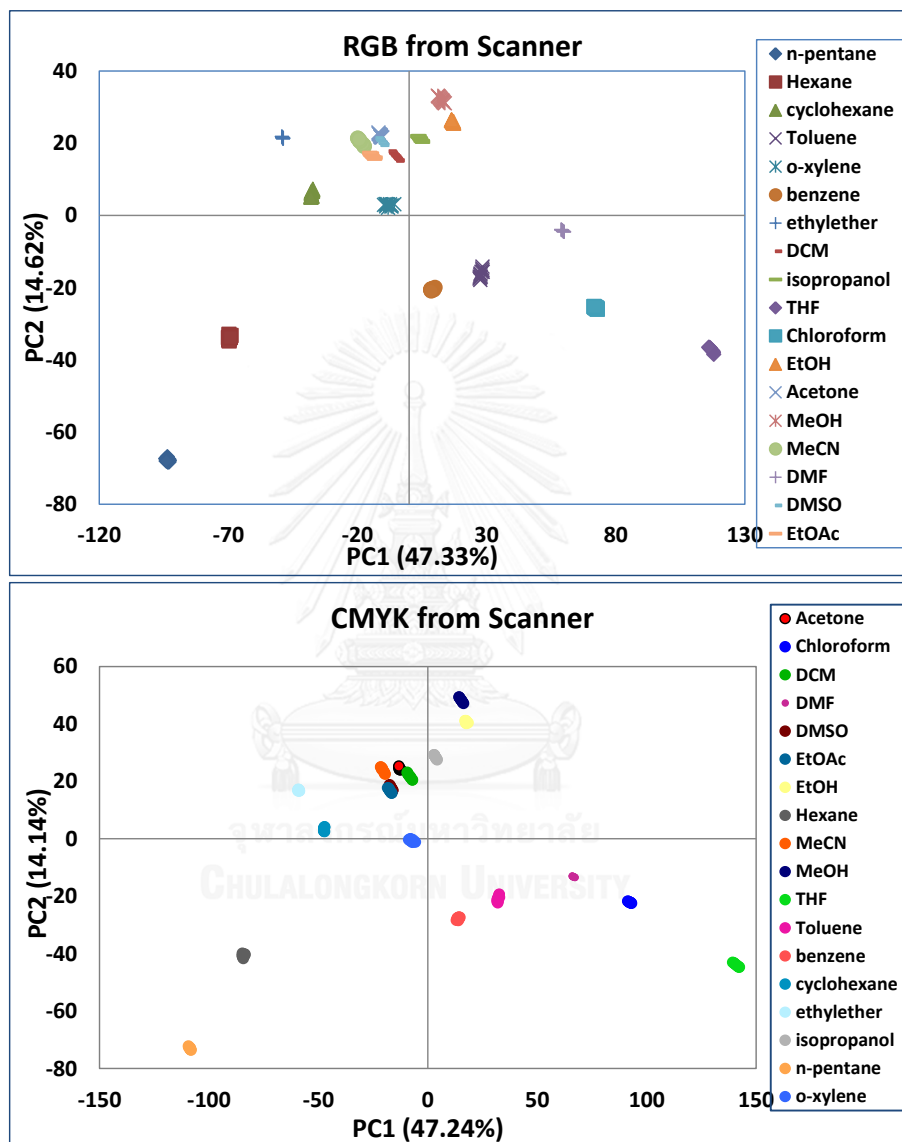
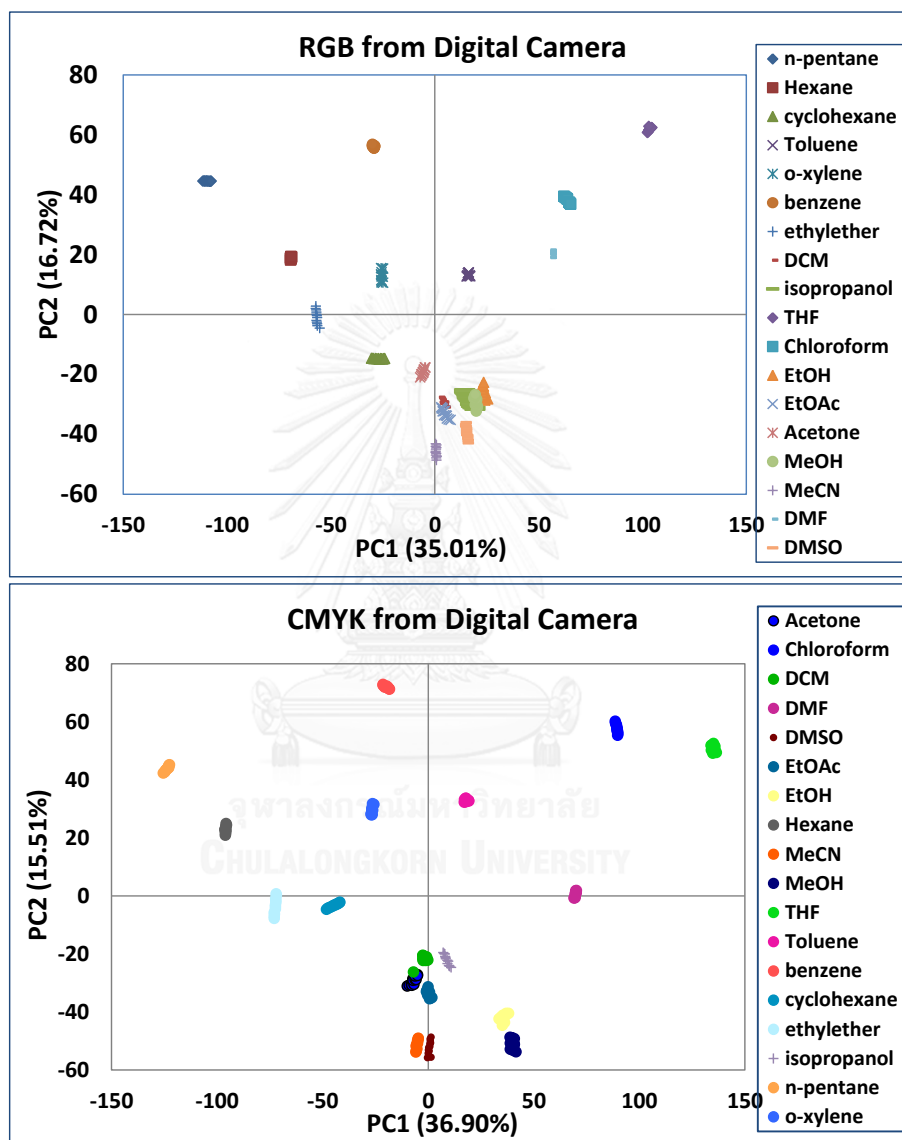


Figure A9. PCA score plot of paper-based 1-8 PDAs sensor array derived from (a) RGB value and (b) CMYK value from digital camera upon exposure to 18 solvents. Each data point represents an average of color values obtained from 12 replications.



VITA

Ms. Radeemada Mungkarndee was born in Bangkok, Thailand. She graduated with high school degree from Satriwittaya School, Bangkok. She received a Bachelor of Science in Chemistry, Department of Chemistry, Faculty of Science, Chulalongkorn University in 1998, and a Master of Science in Petrochemistry and Polymer Science, Faculty of Science, Chulalongkorn University in 2001. She has been working in a field of Intellectual Property after graduation from the master's degree and now working for PTT Global Chemical PCL. In 2009, she also received a Bachelor of Law, Faculty of Law, Chulalongkorn University. She has been a graduate student in biotechnology and become a member of research unit under supervision of Assoc. Prof. Dr. Mongkol Sukwattanasinitt.

